PTO-1390 (01-18)
Approved for use through 8/31/2019. OMB 0651-0021
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

TRANSMITTAL LETTER TO THE UNITED STATES DESIGNATED/ELECTED OFFICE (DO/EO/US) CONCERNING A SUBMISSION UNDER 35 U.S.C. 371

Attorney Docket No.	U.S. Application No. (if known, see 37 CFR 1.5)
SS118-1026US	

International Application No.	International Filing Date	Priority Date Claimed
PCT/US2018/41579	2018-07-11	2018-07-11

Title of Invention
BIOLOGICAL METHODS FOR PREPARING TERPENES

First Named Inventor
Kimberly Ann Aeling

Applicant herewith submits to the United States Designated/Elected Office (DO/EO/US) the following items and other information.

1. ■ This is an express request to begin national examination procedures (35 U.S.C. 371(f)). NOTE: The express request under 35 U.S.C. 371(f) will not be effective unless the requirements under 35 U.S.C. 371(c)(1), (2), and (4) for payment of the basic national fee, copy of the International Application and English translation thereof (if required), and the oath or declaration of the inventor(s) have been received.
2. ■ A copy of the International Application (35 U.S.C. 371(c)(2)) is attached hereto (not required if the International Application was previously communicated by the International Bureau or was filed in the United States Receiving Office (RO/US)).
3. An English language translation of the International Application (35 U.S.C. 371(c)(2))
   a. ☐ is attached hereto.
   b. ☐ has been previously submitted under 35 U.S.C. 154(d)(4).
4. An oath or declaration of the inventor(s) (35 U.S.C. 371(c)(4))
   a. ☐ is attached.
   b. ☐ was previously filed in the international phase under PCT Rule 4.17(iv).

Items 5 to 8 below concern amendments made in the international phase.

PCT Article 19 and 34 amendments

5. ☐ Amendments to the claims under PCT Article 19 are attached (not required if communicated by the International Bureau) (35 U.S.C. 371(c)(3)).
6. ☐ English translation of the PCT Article 19 amendment is attached (35 U.S.C. 371(c)(3)).
7. ☐ English translation of annexes (Article 19 and/or 34 amendments only) of the International Preliminary Examination Report is attached (35 U.S.C. 371(c)(5)).

Cancellation of amendments made in the international phase

8a. ☐ Do not enter the amendment made in the international phase under PCT Article 19.

8b. ■ Do not enter the amendment made in the international phase under PCT Article 34.

NOTE: A proper amendment made in English under Article 19 or 34 will be entered in the U.S. national phase application absent a clear instruction from applicant not to enter the amendment(s).

The following items 9 to 17 concern a document(s) or information included.

9. ■ An Information Disclosure Statement under 37 CFR 1.97 and 1.98.
10. ■ A preliminary amendment.
11. ■ An Application Data Sheet under 37 CFR 1.76.
12. ☐ A substitute specification. NOTE: A substitute specification cannot include claims. See 37 CFR 1.125(b).
13. ☐ A power of attorney and/or change of address letter.
14. ■ A computer-readable form of the sequence listing in accordance with PCT Rule 13*ter*.3 and 37 CFR 1.821-1.825 (not required if sequence listing in text format was indicated on the PCT Request as part of the International Application and the sequence listing was published as part of the international application).
15. ☐ Assignment papers *(cover sheet and document(s))*. Name of Assignee: _____
16. ☐ 37 CFR 3.73(c) Statement *(when there is an Assignee)*.

This collection of information is required by 37 CFR 1.414 and 1.491-1.492. The information is required to obtain or retain a benefit by the public, which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.11 and 1.14. This collection is estimated to take 15 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Mail Stop PCT, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

U.S. APPLN. No. (if known – see 37 CFR 1.5)	INTERNATIONAL APPLICATION No.	ATTORNEY DOCKET No.
	PCT/US2018/41579	SS118-1026US

17. [X] Other items or information:

Search Report and Written Opinion of PCT/US2018/41579 of 13 September 2018;
Second Written Opinion of PCT/US2018/41579 of 13 September 2019

The following fees have been submitted.	CALCULATIONS
18. [■] Basic national fee (37 CFR 1.492(a)) **$300**	$ 300
19. [■] Examination fee (37 CFR 1.492(c)) • If the written opinion prepared by ISA/US or the international preliminary examination report prepared by IPEA/US indicates all claims satisfy provisions of PCT Article 33(1)-(4)..... **$0** • All other situations **$760**	$ 760
20. [■] Search fee (37 CFR 1.492(b)) • If the written opinion prepared by ISA/US or the international preliminary examination report prepared by IPEA/US indicates all claims satisfy provisions of PCT Article 33(1)-(4)..... **$0** • Search fee (37 CFR 1.445(a)(2)) has been paid on the international application to the USPTO as an International Searching Authority ..... **$140** • International Search Report prepared by an ISA other than the US and provided to the Office or previously communicated to the US by the IB..... **$520** • All other situations **$660**	140 $
TOTAL OF 18, 19, and 20 =	$ 1200
[■] Additional fee for specification and drawings filed in paper over 100 sheets (excluding sequence listing in compliance with 37 CFR 1.821(c) or (e) in an electronic medium or computer program listing in an electronic medium) (37 CFR 1.492(j)). Fee for each additional 50 sheets of paper or fraction thereof **$400**	4400

Total Sheets	Extra Sheets	Number of each additional 50 or fraction thereof (round **up** to a whole number)	RATE
634 - 100 =	534 / 50 =	11	x **$400**

$

Surcharge for furnishing any of the search fee, examination fee, or the oath or declaration after the date of commencement of the national stage (37 CFR 1.492(h))..... **$140**	$ 140

CLAIMS	NUMBER FILED	NUMBER EXTRA	RATE	
Total claims	21 - 20 =	1	x **$100**	$ 100
Independent claims	9 - 3 =	6	x **$460**	$ 2760
MULTIPLE DEPENDENT CLAIM(S) (if applicable)			**+ $820**	$

Fee for submission of Sequence Listing text file of 300 MB to 800 MB (37 CFR 1.21(o)(1))..... **$1,000**	$
Fee for submission of Sequence Listing text file of more than 800 MB (37 CFR 1.21(o)(2))..... **$10,000**	$
Processing fee for furnishing the English translation later than 30 months from the earliest claimed priority date (37 CFR 1.492(i))..... **$140 +**	$
TOTAL OF ABOVE CALCULATIONS =	$ 8600
[ ] **Applicant asserts small entity status.** See 37 CFR 1.27. Fees above are reduced by ½.	
[ ] **Applicant certifies micro entity status.** See 37 CFR 1.29. Fees above are reduced by ¾. Applicant must attach form PTO/SB/15A or B or equivalent.	
TOTAL NATIONAL FEE =	$ 8600
Fee for recording the enclosed assignment (37 CFR 1.21(h)). The assignment must be accompanied by an appropriate cover sheet (37 CFR 3.28, 3.31)..... **$50.00** per property +	$
TOTAL FEES ENCLOSED =	$ 8600

a. ☐ A check in the amount of $_____ to cover the above fees is enclosed.

b. ☐ Please charge my Deposit Account No. _____ in the amount of $_____ to cover the above fees.

c. ■ The Director is hereby authorized to charge additional fees which may be required, or credit any overpayment, to Deposit Account No. 506423 as follows:

i. ■ any required fee.

ii. ☐ any required fee except for excess claims fees required under 37 CFR 1.492(d) and (e) and multiple dependent claim fee required under 37 CFR 1.492(f).

d. ■ Fees are to be charged to a credit card. **WARNING:** Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038. The PTO-2038 should only be mailed or faxed to the USPTO. However, when paying the basic national fee, the PTO-2038 may NOT be faxed to the USPTO.

ADVISORY: If filing by EFS-Web, do **NOT** attach the PTO-2038 form as a PDF along with your EFS-Web submission. Please be advised that this is **not** recommended and by doing so your **credit card information may be displayed via PAIR**. To protect your information, it is recommended to pay fees online by using the electronic payment method.

NOTE: Where an appropriate time limit under 37 CFR 1.495 has not been met, a petition to revive (37 CFR 1.137(a) or (b)) must be filed and granted to restore the International Application to pending status.

Statement under 37 CFR 1.55 or 1.78 for AIA (First Inventor to File) Transition Applications

☐ This application (1) claims priority to or the benefit of an application filed before March 16, 2013, and (2) also contains, or contained at any time, a claim to a claimed invention that has an effective filing date on or after March 16, 2013.

NOTE 1: By providing this statement under 37 CFR 1.55 or 1.78, **this application, with a filing date on or after March 16, 2013, will be examined under the first inventor to file provisions of the AIA**.

NOTE 2: A U.S. national stage application may not claim priority to the international application of which it is the national phase. The filing date of a U.S. national stage application is the international filing date. See 35 U.S.C. 363.

Correspondence Address

☒ The address associated with Customer Number: 122515 **OR** ☐ Correspondence address below

Name					
Address					
City		State		Zip Code	
Country		Telephone			
Email					

Signature	/Silvia Salvadori/	Date	2020-01-13
Name (Print/Type)	Silvia Salvadori	Registration No. (Attorney/Agent)	48265

Privacy Act Statement

The **Privacy Act of 1974 (P.L. 93-579)** requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether disclosure of these records is required by the Freedom of Information Act.
2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.
3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.
4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).
5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.
6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).
7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (*i.e.*, GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.
8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspection or an issued patent.
9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 18/4 1579

A. CLASSIFICATION OF SUBJECT MATTER

IPC - C07H 21/04; C07K 14/41 5; C 12N 1/1 6, 1/1 9, 1/21 , 5/1 0, 9/02, 9/88, 9/96, 15/09, 15/52 (201 8.01)
CPC -
C12N 1/16, 5/1 0, 9/88, 9/96, 15/00, 15/09, 15/52, 15/63, 15/70, 15/74, 15/82, 15/8249; C07H 21/04; C07K 14/41 5; C 12P 7/02, 7/26

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
See Search History document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
See Search History document

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
See Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X — Y	WO 2014/100504 A2 (VERDEZYNE, INC.) 26 June 2014; page 9, 1st paragraph; page 10, 1st paragraph; page 27, 3rd paragraph; page 41, 1st paragraph; page 143, 2nd paragraph	1-4, 5/1-4, 83-84 --- 44, 46/44, 47/46/44, 48/47/46/44, 85, 149/146-148
X -- Y	JP 2012-254044 A (KIRIN HOLDING CO LTD) 27 December 2012; page 5, 6th paragraph; page 6, 5th paragraph; page 11, 3rd paragraph; page 12, 1st paragraph; claim 10	43, 46/43, 47/46/43, 48/47/46/43, 146-148 --- 44-45, 46/44-45, 47/46/44-45, 48/47/46/44-45
Y	US 2014/01 13343 A1 (DAVIET, L et al.) 24 April 2014; paragraphs [0038], [0044]	45, 46/45, 47/46/45. 48/47/46/45, 85
Y	US 2003/0044946 A1 (LONGO, VD) 06 March 2003; paragraph [0093]	149/146-148
A	US 2015/0010978 A1 (SAPPHIRE ENERGY, INC.) 08 January 2015; eritire document	1-4, 5/1-4, 43-45, 46/43-45, 47/46/43-45, 48/47/46/43-45, 83-85, 146-148, 149/146-148

☒ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"V" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
4 September 2018 (04.09.2018)	13 SEP 2018
Name and mailing address of the ISA/ Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Authorized officer Shane Thomas PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (January 201 5)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US1 8/41 579

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely.

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☒ Claims Nos.: 6-42, 49-82, 86-145, 150-162
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. Ill Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 6 of 3077

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US1 8/41 579

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2013/0066035 A1 (BURGARD, AP et al.) 14 March 2013; entire document	1-4, 5/1-4, 43-45, 46/43-45, 47/46/43-45, 48/47/46/43-45, 83-85, 146-148, 149/146-148
A	US 2013/0302861 A1 (YISSUM RESEARCH DEVELOPMENT COMPANY OF THE HEBREW UNIVERSITY OF JERUSALEM LTD.) 14 November 2013; entire document	1-4, 5/1-4, 43-45, 46/43-45, 47/46/43-45, 48/47/46/43-45, 83-85, 146-148, 149/146-148
P, X	WO 2018/0941 10 A2 (THE BOARD OF TRUSTEES OF THE LELAND STANFORD JUNIOR UNIVERSITY) 24 May 2018; entire document	1-4, 5/1-4, 43-45, 46/43-45, 47/46/43-45, 48/47/46/43-45, 83-85, 146-148, 149/146-148

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 7 of 3077

ADVANCE E-MAIL

From the INTERNATIONAL BUREAU

PCT

NOTIFICATION OF THE RECORDING OF A CHANGE

(PCT Rule 92*bis*.1 and Administrative Instructions, Section 422)

To:

SALVADORI, Silvia
Silvia Salvadori, P.C.
270 Madison Avenue, 8th Floor
New York, NY 10016
ÉTATS-UNIS D'AMÉRIQUE

Date of mailing *(day/month/year)*
07 January 2020 (07.01.2020)

Applicant's or agent's file reference
SS118-1026

IMPORTANT NOTIFICATION

International application No.
PCT/US2018/041579

International filing date *(day/month/year)*
11 July 2018 (11.07.2018)

1. The following indications appeared on record concerning:

☐ the applicant ☒ the inventor ☐ the agent ☐ the common representative

Name and Address

State of Nationality | State of Residence

Telephone No.

Facsimile No.

E-mail address

2. The International Bureau hereby notifies the applicant that the following change has been recorded concerning:

☒ the person ☐ the name ☐ the address ☐ the nationality ☐ the residence

Name and Address
BEARDSLEE, Thomas
6827 Carnation Drive
Carlsbad, CA 92011
United States of America

State of Nationality | State of Residence

Telephone No.

Facsimile No.

E-mail address
☐ Notifications by e-mail authorized

3. Further observations, if necessary:
The person identified in Box 2 has been added as inventor only.

4. A copy of this notification has been sent to:

☒ the receiving Office
☐ the International Searching Authority
☐ the Authority(ies) specified for supplementary search
☐ the International Preliminary Examining Authority
☐ the designated Offices concerned
☒ the elected Offices concerned
☐ other:

The International Bureau of WIPO
34, chemin des Colombettes
1211 Geneva 20, Switzerland

Authorized officer
Chahine Leila
e-mail pct.team4@wipo.int
Telephone No. +41 22 338 74 04

Form PCT/IB/306 (January 2009) 1/NVM9TUPVGXTNW0

ADVANCE E-MAIL

From the INTERNATIONAL BUREAU

PCT

NOTIFICATION OF THE RECORDING OF A CHANGE

(PCT Rule 92*bis*.1 and Administrative Instructions, Section 422)

To:

SALVADORI, Silvia
Silvia Salvadori, P.C.
270 Madison Avenue, 8th Floor
New York, NY 10016
ÉTATS-UNIS D'AMÉRIQUE

Date of mailing *(day/month/year)*
07 January 2020 (07.01.2020)

Applicant's or agent's file reference
SS118-1026

IMPORTANT NOTIFICATION

International application No.
PCT/US2018/041579

International filing date *(day/month/year)*
11 July 2018 (11.07.2018)

1. The following indications appeared on record concerning:

☐ the applicant ☒ the inventor ☐ the agent ☐ the common representative

Name and Address

State of Nationality | State of Residence

Telephone No.

Facsimile No.

E-mail address

2. The International Bureau hereby notifies the applicant that the following change has been recorded concerning:

☒ the person ☐ the name ☐ the address ☐ the nationality ☐ the residence

Name and Address
DE WALS, Pierre-Yves
925 North Vulcan Ave., Apt. #205
Encinitas, CA 92024
United States of America

State of Nationality | State of Residence

Telephone No.

Facsimile No.

E-mail address
☐ Notifications by e-mail authorized

3. Further observations, if necessary:
The person identified in Box 2 has been added as inventor only.

4. A copy of this notification has been sent to:

☒ the receiving Office
☐ the International Searching Authority
☐ the Authority(ies) specified for supplementary search
☐ the International Preliminary Examining Authority
☐ the designated Offices concerned
☒ the elected Offices concerned
☐ other:

The International Bureau of WIPO
34, chemin des Colombettes
1211 Geneva 20, Switzerland

Authorized officer
Chahine Leila
e-mail pct.team4@wipo.int
Telephone No. +41 22 338 74 04

Form PCT/IB/306 (January 2009) 1/TFG6TKWGKNVWW0

ADVANCE E-MAIL

From the INTERNATIONAL BUREAU

PCT

NOTIFICATION OF THE RECORDING OF A CHANGE

(PCT Rule 92*bis*.1 and Administrative Instructions, Section 422)

To:

SALVADORI, Silvia
Silvia Salvadori, P.C.
270 Madison Avenue, 8th Floor
New York, NY 10016
ÉTATS-UNIS D'AMÉRIQUE

Date of mailing *(day/month/year)*
07 January 2020 (07.01.2020)

Applicant's or agent's file reference
SS118-1026

IMPORTANT NOTIFICATION

International application No.
PCT/US2018/041579

International filing date *(day/month/year)*
11 July 2018 (11.07.2018)

1. The following indications appeared on record concerning:

☐ the applicant ☒ the inventor ☐ the agent ☐ the common representative

Name and Address

State of Nationality | State of Residence

Telephone No.

Facsimile No.

E-mail address

2. The International Bureau hereby notifies the applicant that the following change has been recorded concerning:

☒ the person ☐ the name ☐ the address ☐ the nationality ☐ the residence

Name and Address
KNIGHT, Eric
2744 Loker Ave West, Unit 100
Carlsbad, CA 92010
United States of America

State of Nationality | State of Residence

Telephone No.

Facsimile No.

E-mail address
☐ Notifications by e-mail authorized

3. Further observations, if necessary:
The person identified in Box 2 has been added as inventor only.

4. A copy of this notification has been sent to:

☒ the receiving Office
☐ the International Searching Authority
☐ the Authority(ies) specified for supplementary search
☐ the International Preliminary Examining Authority
☐ the designated Offices concerned
☒ the elected Offices concerned
☐ other:

The International Bureau of WIPO
34, chemin des Colombettes
1211 Geneva 20, Switzerland

Authorized officer
Chahine Leila
e-mail pct.team4@wipo.int
Telephone No. +41 22 338 74 04

Form PCT/IB/306 (January 2009) 1/HEX4Q5ZPN5MSG0

ADVANCE E-MAIL

From the INTERNATIONAL BUREAU

PCT

NOTIFICATION OF THE RECORDING OF A CHANGE

(PCT Rule 92*bis*.1 and Administrative Instructions, Section 422)

To:

SALVADORI, Silvia
Silvia Salvadori, P.C.
270 Madison Avenue, 8th Floor
New York, NY 10016
ÉTATS-UNIS D'AMÉRIQUE

Date of mailing *(day/month/year)*
07 January 2020 (07.01.2020)

Applicant's or agent's file reference
SS118-1026

IMPORTANT NOTIFICATION

International application No.
PCT/US2018/041579

International filing date *(day/month/year)*
11 July 2018 (11.07.2018)

1. The following indications appeared on record concerning:

☐ the applicant ☒ the inventor ☐ the agent ☐ the common representative

Name and Address

State of Nationality | State of Residence

Telephone No.

Facsimile No.

E-mail address

2. The International Bureau hereby notifies the applicant that the following change has been recorded concerning:

☒ the person ☐ the name ☐ the address ☐ the nationality ☐ the residence

Name and Address
SCOTT, Erin
1579 Caudor Street
Encinitas, CA 92024
United States of America

State of Nationality | State of Residence

Telephone No.

Facsimile No.

E-mail address
☐ Notifications by e-mail authorized

3. Further observations, if necessary:
The person identified in Box 2 has been added as inventor only.

4. A copy of this notification has been sent to:

☒ the receiving Office
☐ the International Searching Authority
☐ the Authority(ies) specified for supplementary search
☐ the International Preliminary Examining Authority
☐ the designated Offices concerned
☒ the elected Offices concerned
☐ other:

The International Bureau of WIPO
34, chemin des Colombettes
1211 Geneva 20, Switzerland

Authorized officer
Chahine Leila
e-mail pct.team4@wipo.int
Telephone No. +41 22 338 74 04

Form PCT/IB/306 (January 2009) 1/SZFMR6WT332M50

ADVANCE E-MAIL

From the INTERNATIONAL BUREAU

PCT

NOTIFICATION OF THE RECORDING OF A CHANGE

(PCT Rule 92*bis*.1 and Administrative Instructions, Section 422)

To:

SALVADORI, Silvia
Silvia Salvadori, P.C.
270 Madison Avenue, 8th Floor
New York, NY 10016
ÉTATS-UNIS D'AMÉRIQUE

Date of mailing *(day/month/year)*
07 January 2020 (07.01.2020)

Applicant's or agent's file reference
SS118-1026

IMPORTANT NOTIFICATION

International application No.
PCT/US2018/041579

International filing date *(day/month/year)*
11 July 2018 (11.07.2018)

1. The following indications appeared on record concerning:

☐ the applicant ☒ the inventor ☐ the agent ☐ the common representative

Name and Address

State of Nationality | State of Residence

Telephone No.

Facsimile No.

E-mail address

2. The International Bureau hereby notifies the applicant that the following change has been recorded concerning:

☒ the person ☐ the name ☐ the address ☐ the nationality ☐ the residence

Name and Address
WALBRIDGE, Michael
4 Diane Dr.
Saugus, MA 01906
United States of America

State of Nationality | State of Residence

Telephone No.

Facsimile No.

E-mail address
☐ Notifications by e-mail authorized

3. Further observations, if necessary:
The person identified in Box 2 has been added as inventor only.

4. A copy of this notification has been sent to:

☒ the receiving Office
☐ the International Searching Authority
☐ the Authority(ies) specified for supplementary search
☐ the International Preliminary Examining Authority
☐ the designated Offices concerned
☒ the elected Offices concerned
☐ other:

The International Bureau of WIPO
34, chemin des Colombettes
1211 Geneva 20, Switzerland

Authorized officer
Chahine Leila
e-mail pct.team4@wipo.int
Telephone No. +41 22 338 74 04

Form PCT/IB/306 (January 2009) 1/EEFVDET3PN2770

ADVANCE E-MAIL

From the INTERNATIONAL BUREAU

PCT

NOTIFICATION OF THE RECORDING OF A CHANGE

(PCT Rule 92*bis*.1 and Administrative Instructions, Section 422)

To:

SALVADORI, Silvia
Silvia Salvadori, P.C.
270 Madison Avenue, 8th Floor
New York, NY 10016
ÉTATS-UNIS D'AMÉRIQUE

Date of mailing *(day/month/year)*
07 January 2020 (07.01.2020)

Applicant's or agent's file reference
SS118-1026

IMPORTANT NOTIFICATION

International application No.
PCT/US2018/041579

International filing date *(day/month/year)*
11 July 2018 (11.07.2018)

1. The following indications appeared on record concerning:

☐ the applicant ☒ the inventor ☐ the agent ☐ the common representative

Name and Address

State of Nationality | State of Residence

Telephone No.

Facsimile No.

E-mail address

2. The International Bureau hereby notifies the applicant that the following change has been recorded concerning:

☒ the person ☐ the name ☐ the address ☐ the nationality ☐ the residence

Name and Address
YI, Jian
13313 Thunderhead Street
San Diego, CA 92129
United States of America

State of Nationality | State of Residence

Telephone No.

Facsimile No.

E-mail address
☐ Notifications by e-mail authorized

3. Further observations, if necessary:
The person identified in Box 2 has been added as inventor only.

4. A copy of this notification has been sent to:

☒ the receiving Office
☐ the International Searching Authority
☐ the Authority(ies) specified for supplementary search
☐ the International Preliminary Examining Authority
☐ the designated Offices concerned
☒ the elected Offices concerned
☐ other:

The International Bureau of WIPO
34, chemin des Colombettes
1211 Geneva 20, Switzerland

Authorized officer
Chahine Leila
e-mail pct.team4@wipo.int
Telephone No. +41 22 338 74 04

Form PCT/IB/306 (January 2009) 1/RJD34N3SALYAR0

ADVANCE E-MAIL

From the INTERNATIONAL BUREAU

PCT

COMMUNICATION IN CASES FOR WHICH NO OTHER FORM IS APPLICABLE

To:

SALVADORI, Silvia
Silvia Salvadori, P.C.
270 Madison Avenue, 8th Floor
New York, NY 10016
ÉTATS-UNIS D'AMÉRIQUE

Date of mailing *(day/month/year)* 07 January 2020 (07.01.2020)	
Applicant's or agent's file reference SS118-1026	**REPLY DUE** **see paragraph 1 below**
International application No. PCT/US2018/041579	International filing date *(day/month/year)* 11 July 2018 (11.07.2018)
Applicant RADICI CHIMICA S.P.A.	

1. ☐ REPLY DUE within months/days from the above date of mailing
   ☐ NO REPLY DUE, however, see below
   ☒ IMPORTANT COMMUNICATION
   ☐ INFORMATION ONLY

2. COMMUNICATION:

It has been brought to the attention of the International Bureau that in respect of the above-identified international application, the name of the inventor AELING, Kimberly, Ann contains an error in the international publication no. WO 2019/014310 published on 17 January 2019 (17.01.2019).

The International Bureau will publish a corrected version of the front page reflecting the correct name on 30 January 2020 (30.01.2020).

Please be also informed that the correct bibliographic data will be available within 24 hours on the WIPO' s Internet site, at
http://www.wipo.int/patentscope/search/en/search.jsf
A copy of this Notification is being sent to the receiving Office (RO/US) and to the designated Offices concerned.

The International Bureau of WIPO 34, chemin des Colombettes 1211 Geneva 20, Switzerland	Authorized officer Chahine Leila e-mail pct.team4@wipo.int Telephone No. +41 22 338 74 04

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 14 of 3077

ADVANCE E-MAIL

From the INTERNATIONAL BUREAU

PCT

NOTIFICATION OF THE RECORDING OF A CHANGE

(PCT Rule 92*bis*.1 and Administrative Instructions, Section 422)

To:

SALVADORI, Silvia
Silvia Salvadori, P.C.
270 Madison Avenue, 8th Floor
New York, NY 10016
ÉTATS-UNIS D'AMÉRIQUE

Date of mailing *(day/month/year)*
15 May 2019 (15.05.2019)

Applicant's or agent's file reference
SS118-1026

IMPORTANT NOTIFICATION

International application No.
PCT/US2018/041579

International filing date *(day/month/year)*
11 July 2018 (11.07.2018)

1. The following indications appeared on record concerning:

☒ the applicant ☐ the inventor ☐ the agent ☐ the common representative

Name and Address
VERDEZYNE (ABC), LLC
2735 Sand Hill Road
Menlo Park, CA 94025
United States of America

State of Nationality: US
State of Residence: US
Telephone No.: 925-302-4298
Facsimile No.: 925-365-3818
E-mail address:

2. The International Bureau hereby notifies the applicant that the following change has been recorded concerning:

☒ the person ☐ the name ☐ the address ☐ the nationality ☐ the residence

Name and Address
RADICI CHIMICA S.P.A.
Via G. Verdi, 11
24121 Bergamo
Italy

State of Nationality: IT
State of Residence: IT
Telephone No.:
Facsimile No.:
E-mail address:
☐ Notifications by e-mail authorized

3. Further observations, if necessary:

4. A copy of this notification has been sent to:

☒ the receiving Office
☐ the International Searching Authority
☐ the Authority(ies) specified for supplementary search
☐ the International Preliminary Examining Authority
☒ the designated Offices concerned
☐ the elected Offices concerned
☒ other: VERDEZYNE (ABC), LLC

The International Bureau of WIPO
34, chemin des Colombettes
1211 Geneva 20, Switzerland

Authorized officer
Chahine Leila
e-mail pct.team4@wipo.int
Telephone No. +41 22 338 74 04

Form PCT/IB/306 (January 2009) 1/MT53LNELGHQ30

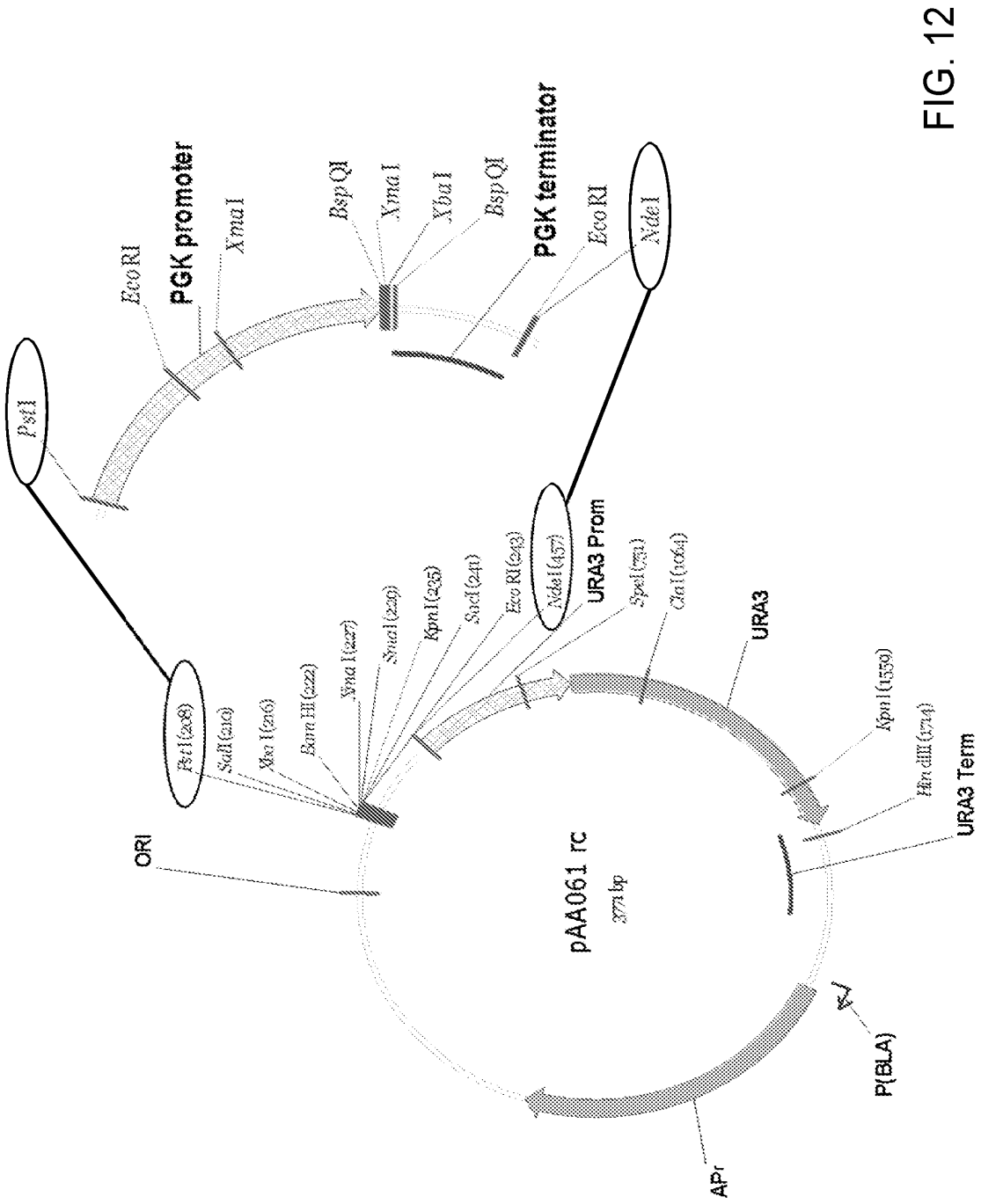


FIG. 12

12/59

REPLACEMENT SHEET

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 16 of 3077

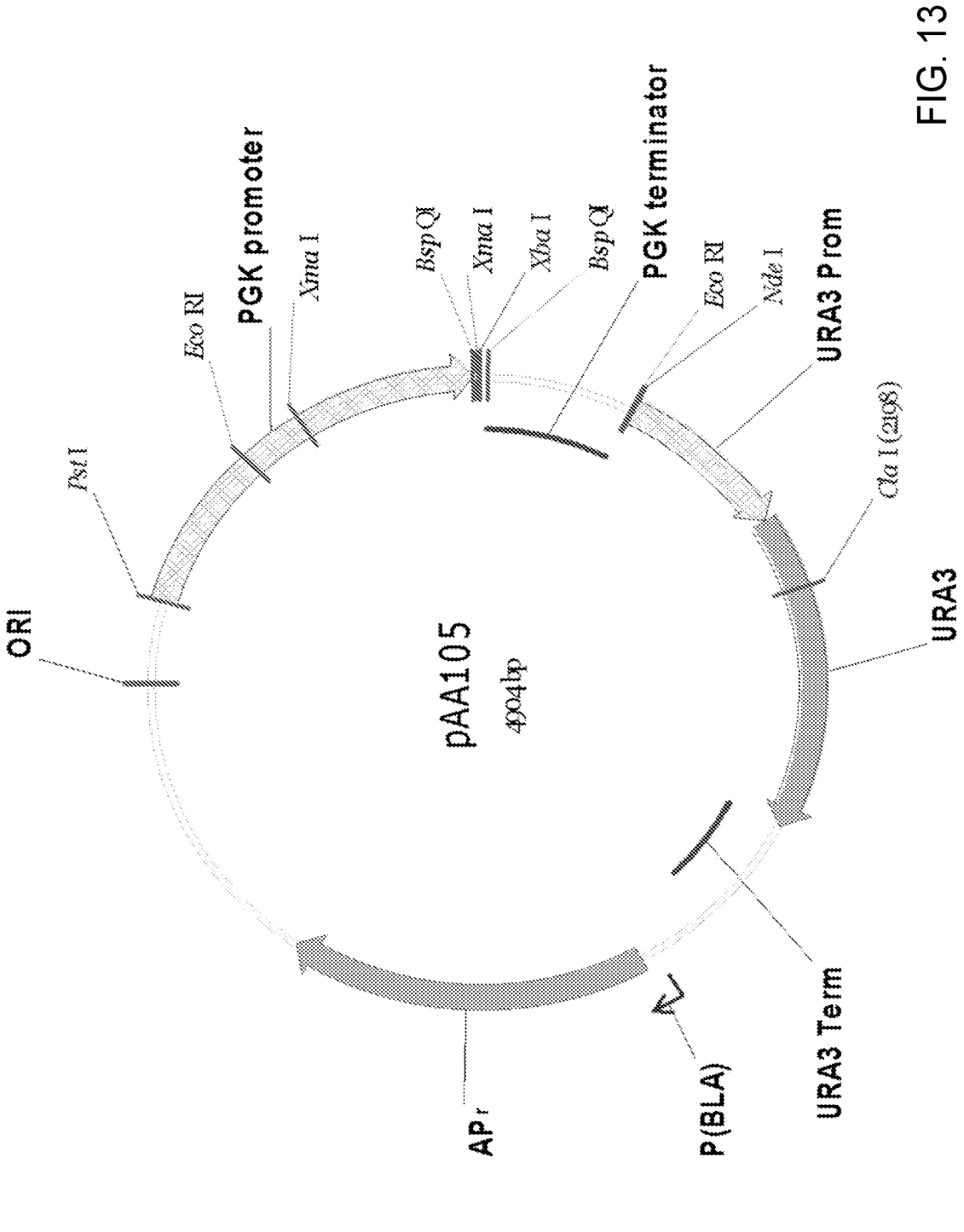


FIG. 13

REPLACEMENT SHEET

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 17 of 3077

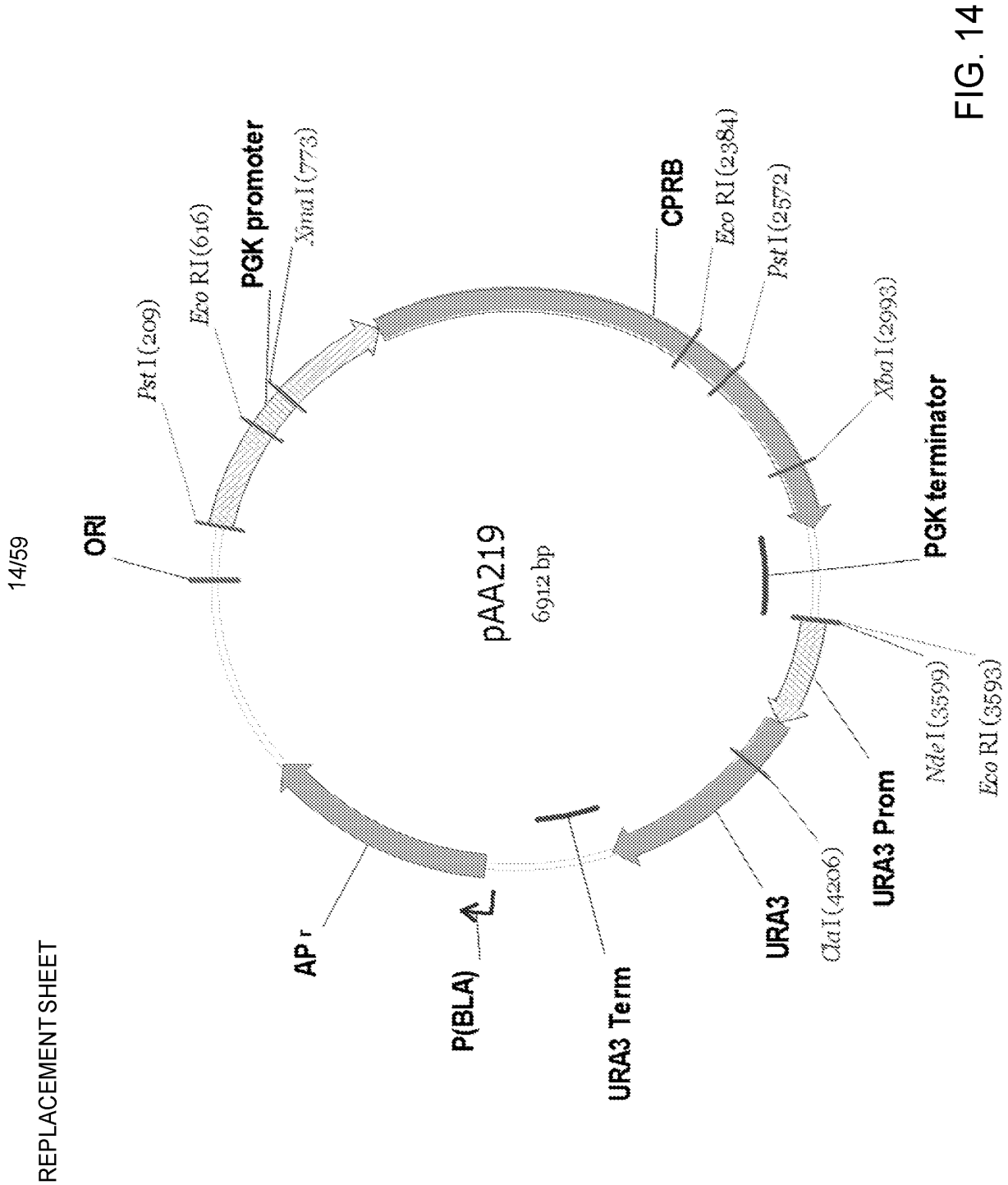


FIG. 14

REPLACEMENT SHEET

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 18 of 3077

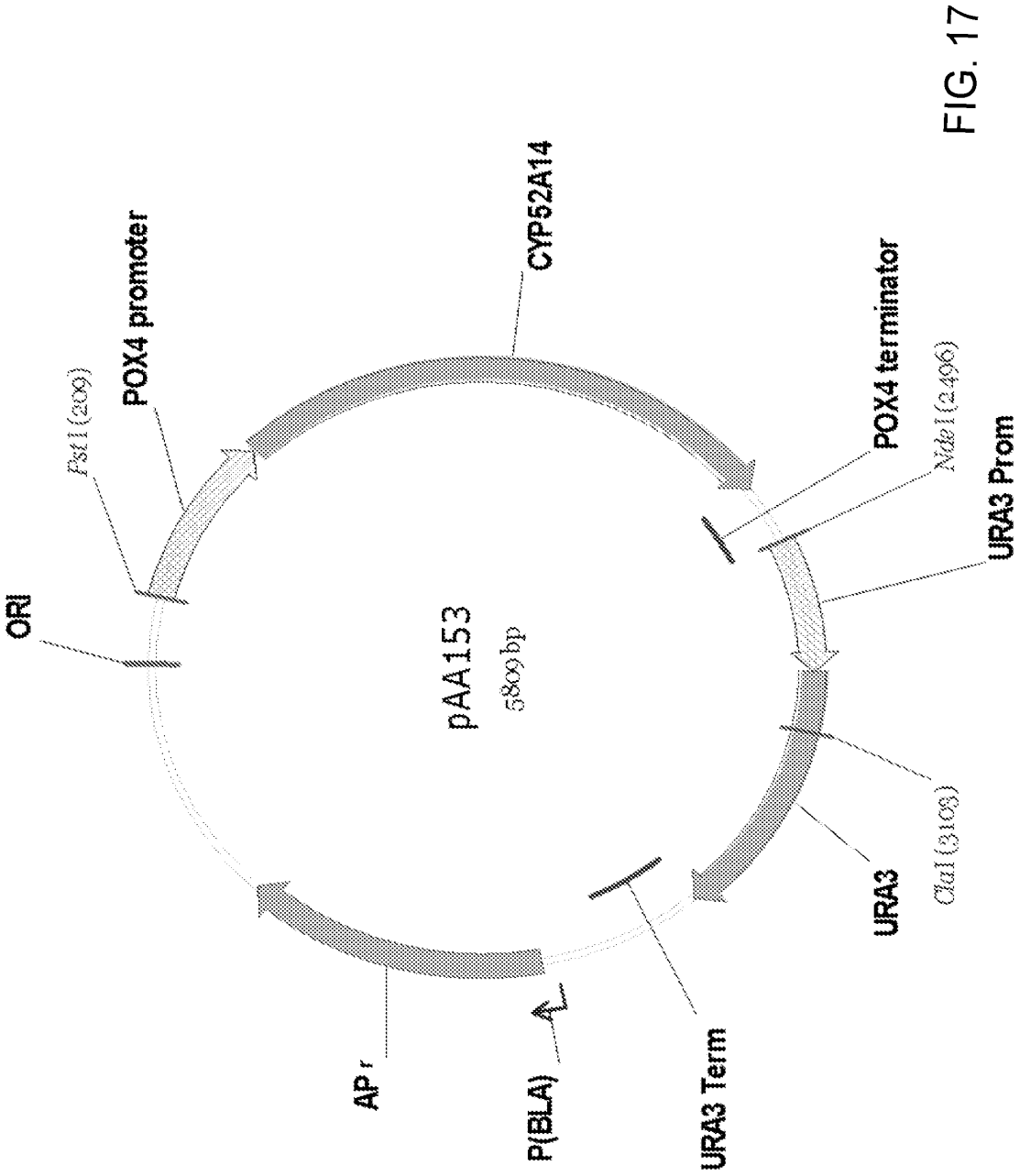


FIG. 17

REPLACEMENT SHEET

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 19 of 3077

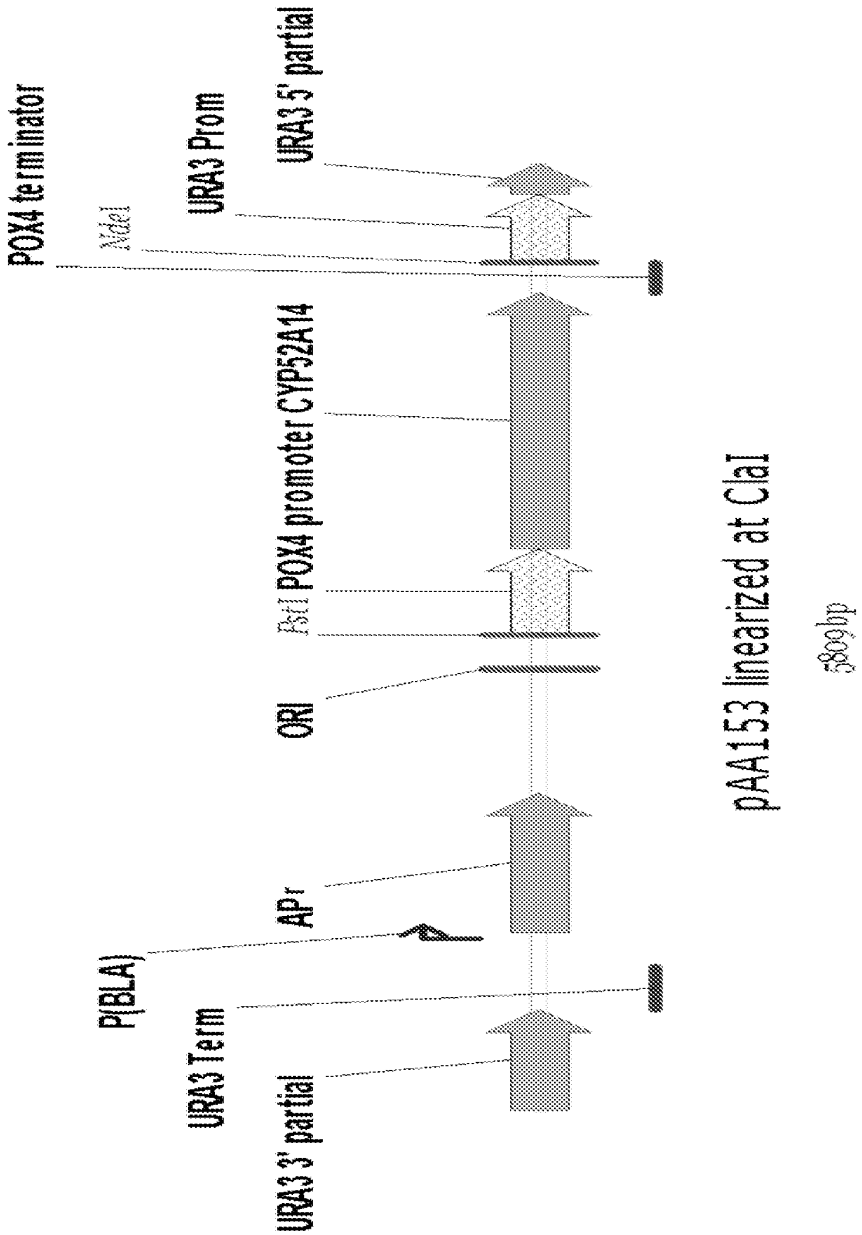


FIG. 18

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 20 of 3077

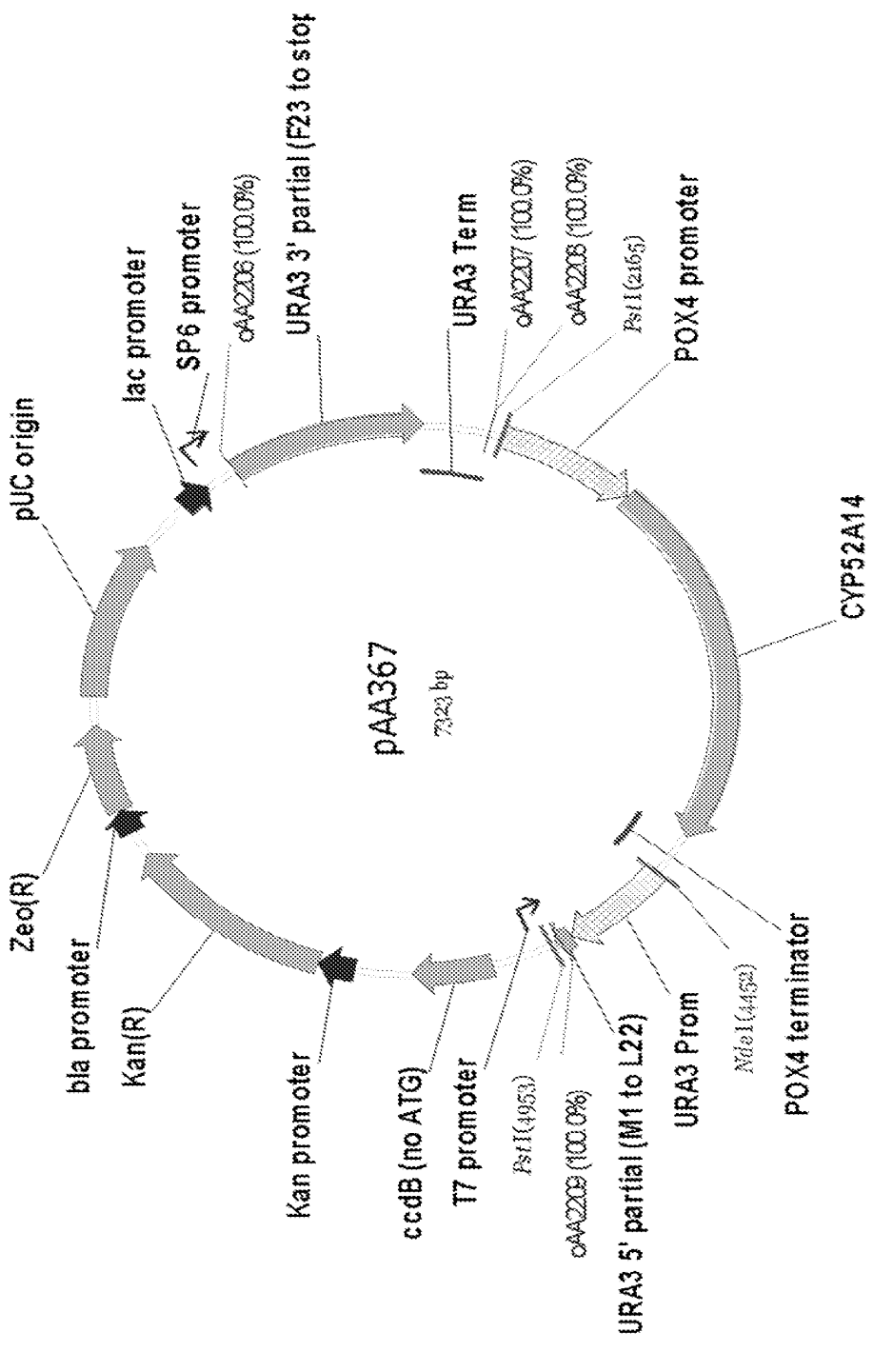


FIG. 19

REPLACEMENT SHEET

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 21 of 3077

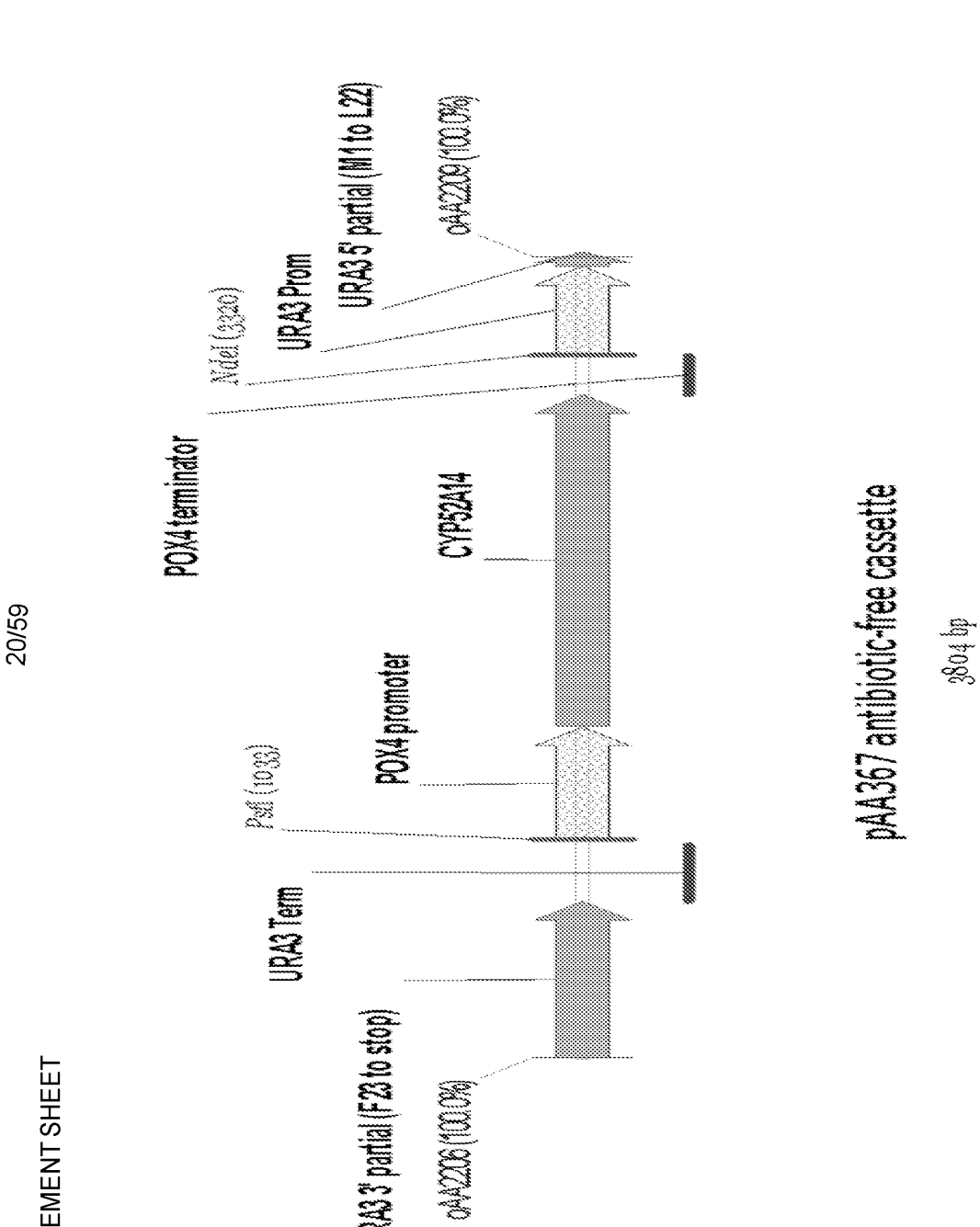


FIG. 20

REPLACEMENT SHEET

REPLACEMENT SHEET

FIG. 24

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 23 of 3077

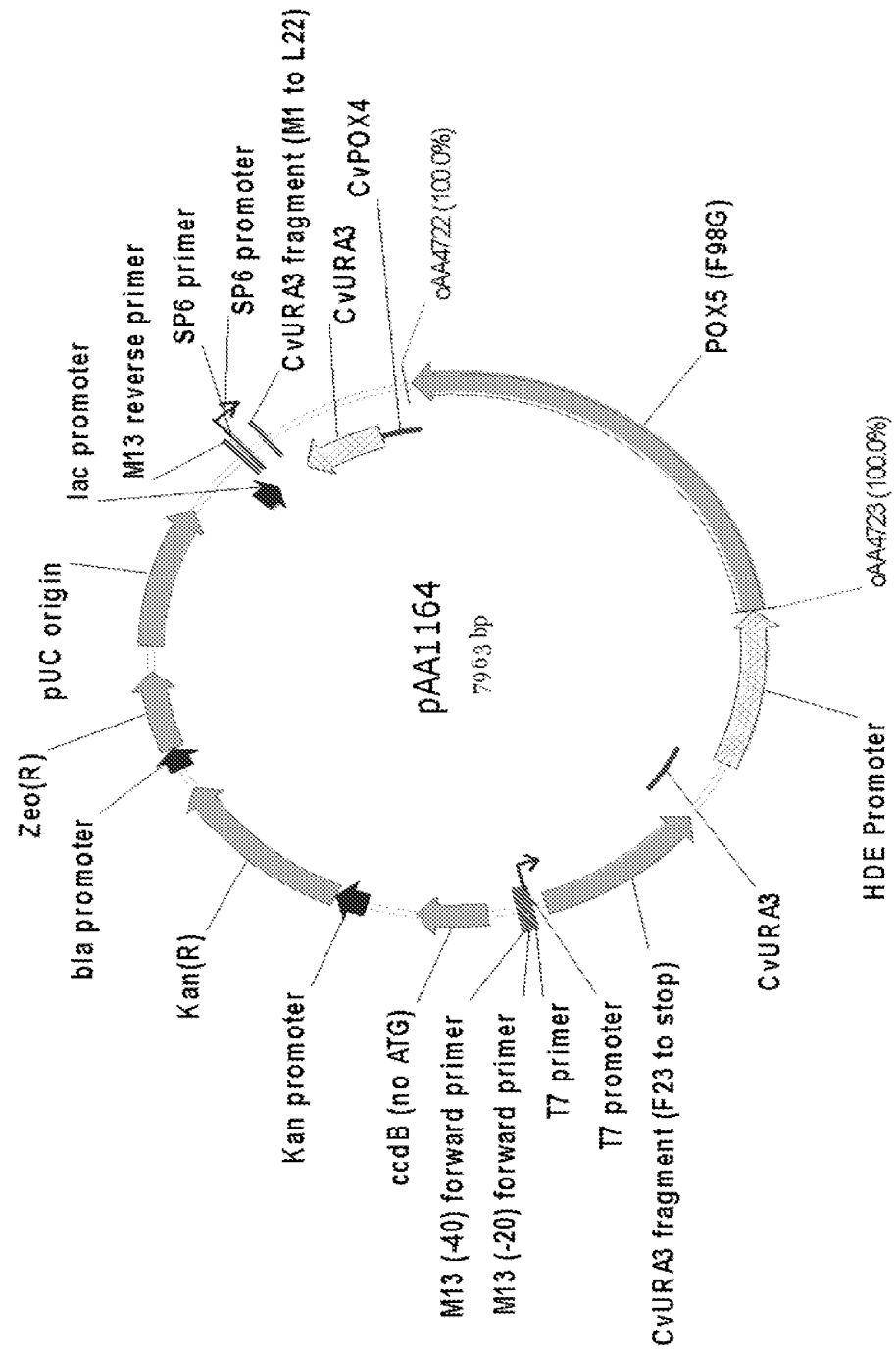


FIG. 26

REPLACEMENT SHEET

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 24 of 3077

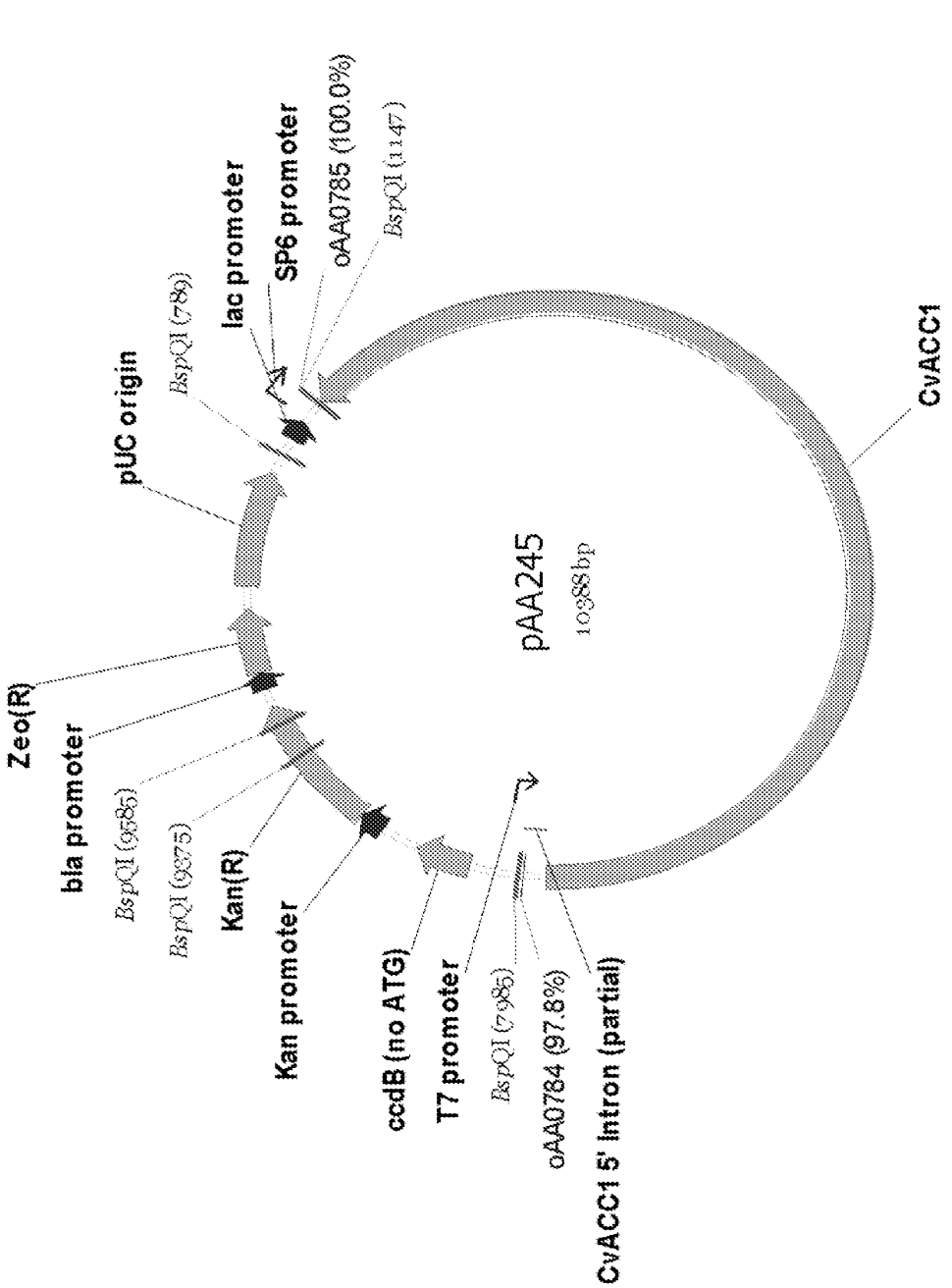


FIG. 28

REPLACEMENT SHEET

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 25 of 3077

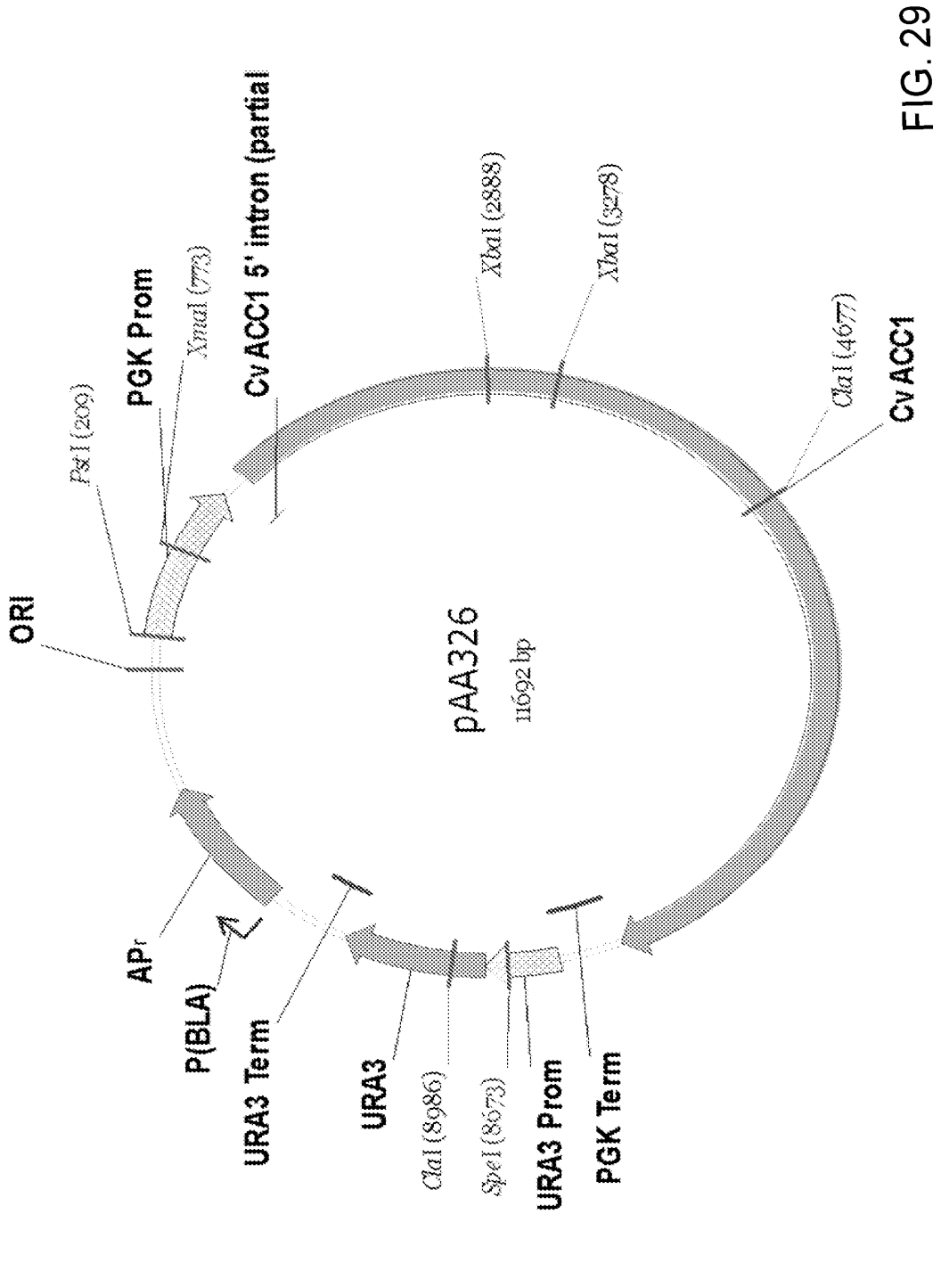


FIG. 29

REPLACEMENT SHEET

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 26 of 3077

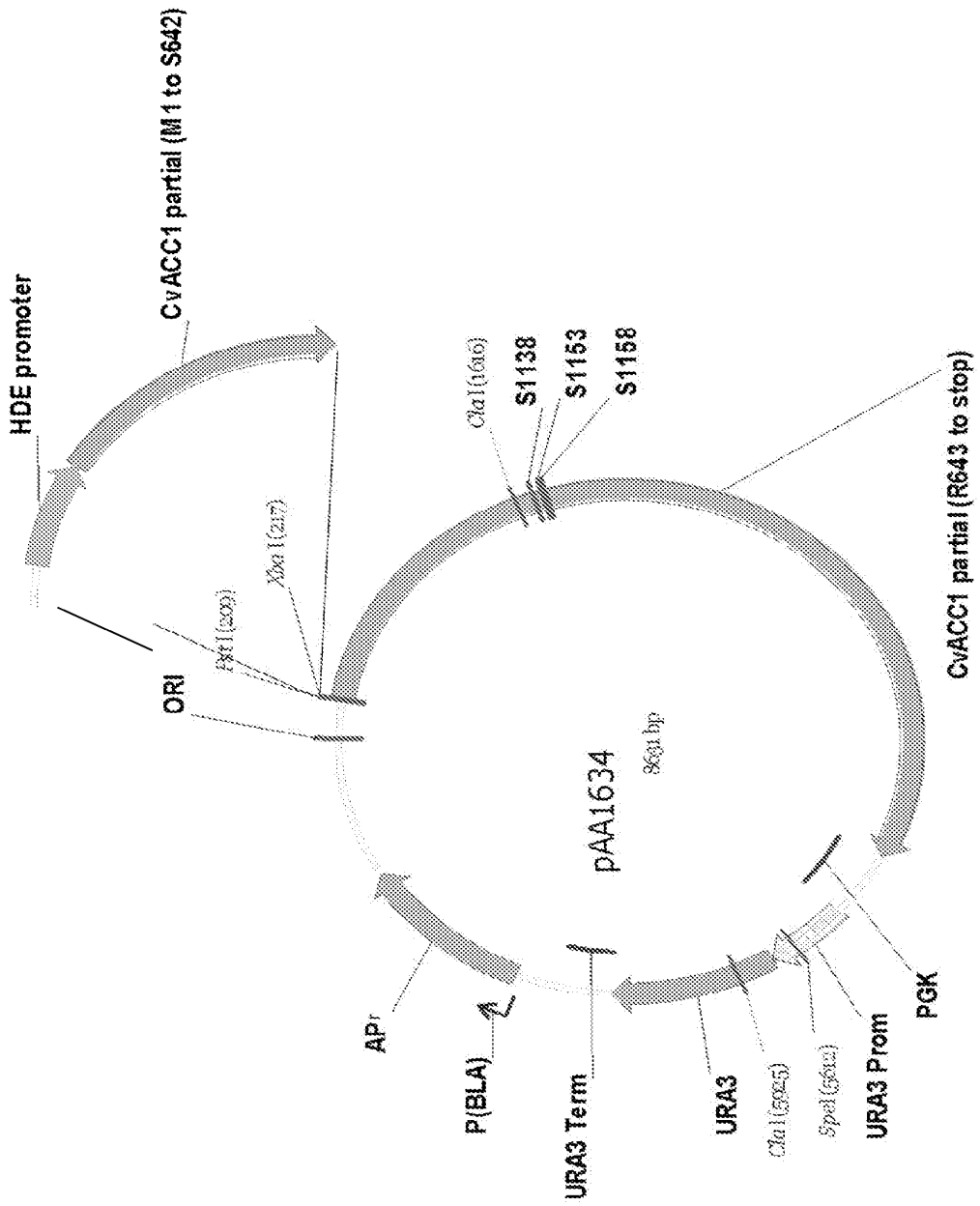


FIG. 30

REPLACEMENT SHEET

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 27 of 3077

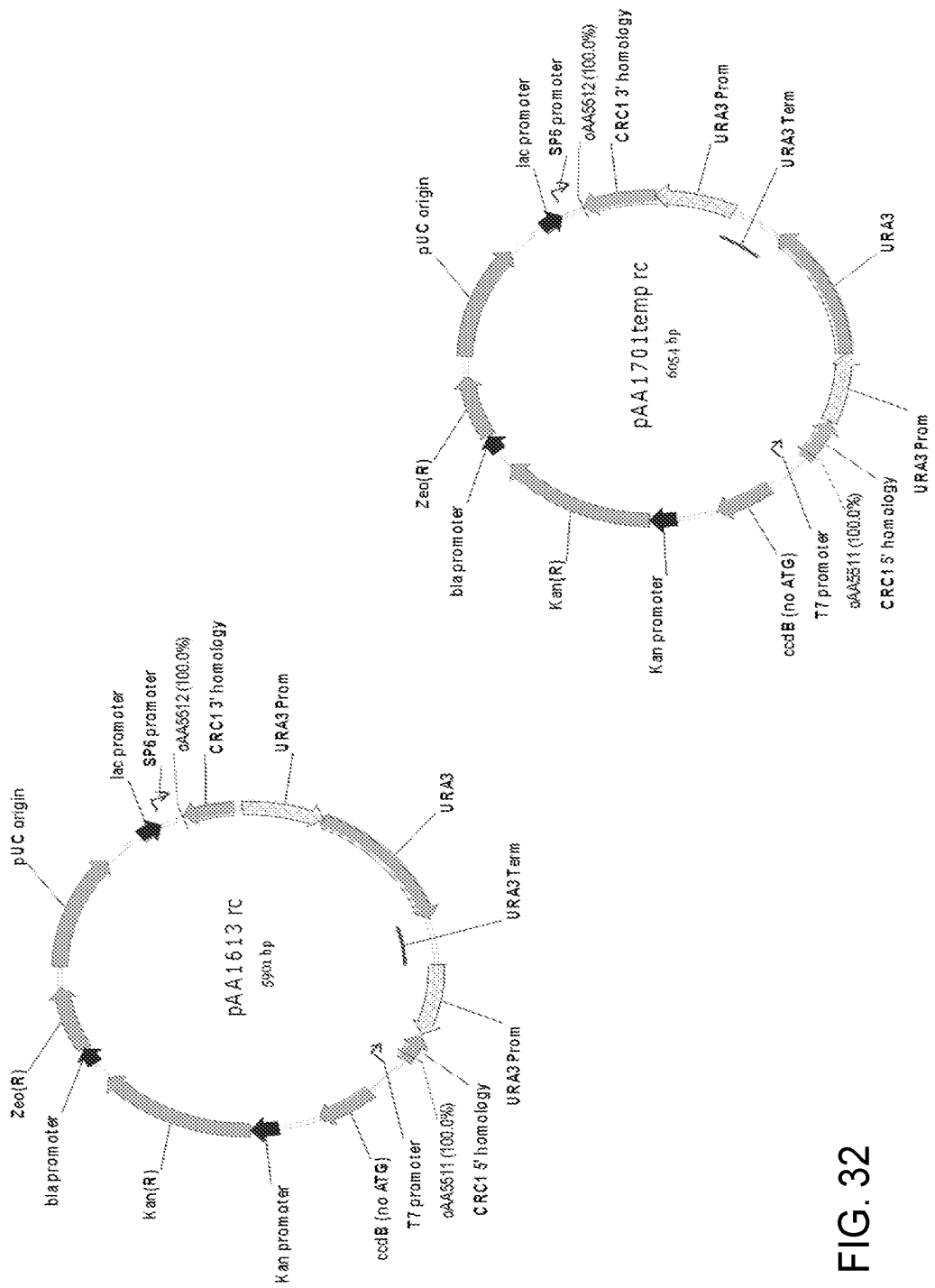

pAA1701temp rc
6064 bp
pUC origin
lac promoter
SP6 promoter
pAA5612 (100.0%)
CRC1 3' homology
URA3 Prom
URA3 Term
URA3
URA3 Prom
CRC1 5' homology
pAA5611 (100.0%)
T7 promoter
ccdB (no ATG)
Kan promoter
Kan(R)
bla promoter
Zeo(R)
pAA1613 rc
5900 bp

FIG. 32

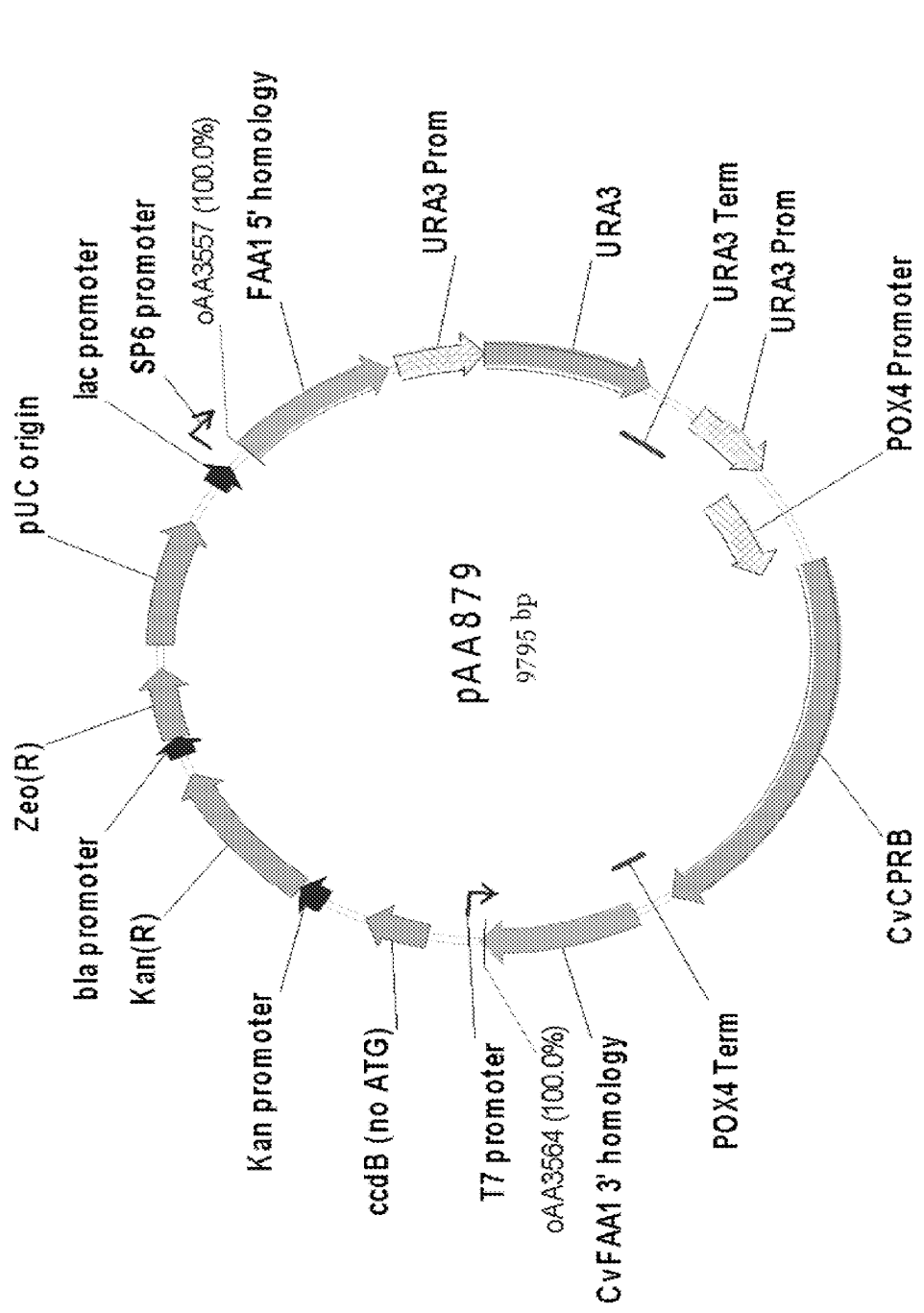


FIG. 35

REPLACEMENT SHEET

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 29 of 3077

Residue highlight	HotSpot score
RESIDUE	9
RESIDUE	8
RESIDUE	7
RESIDUE	6

FIG. 38

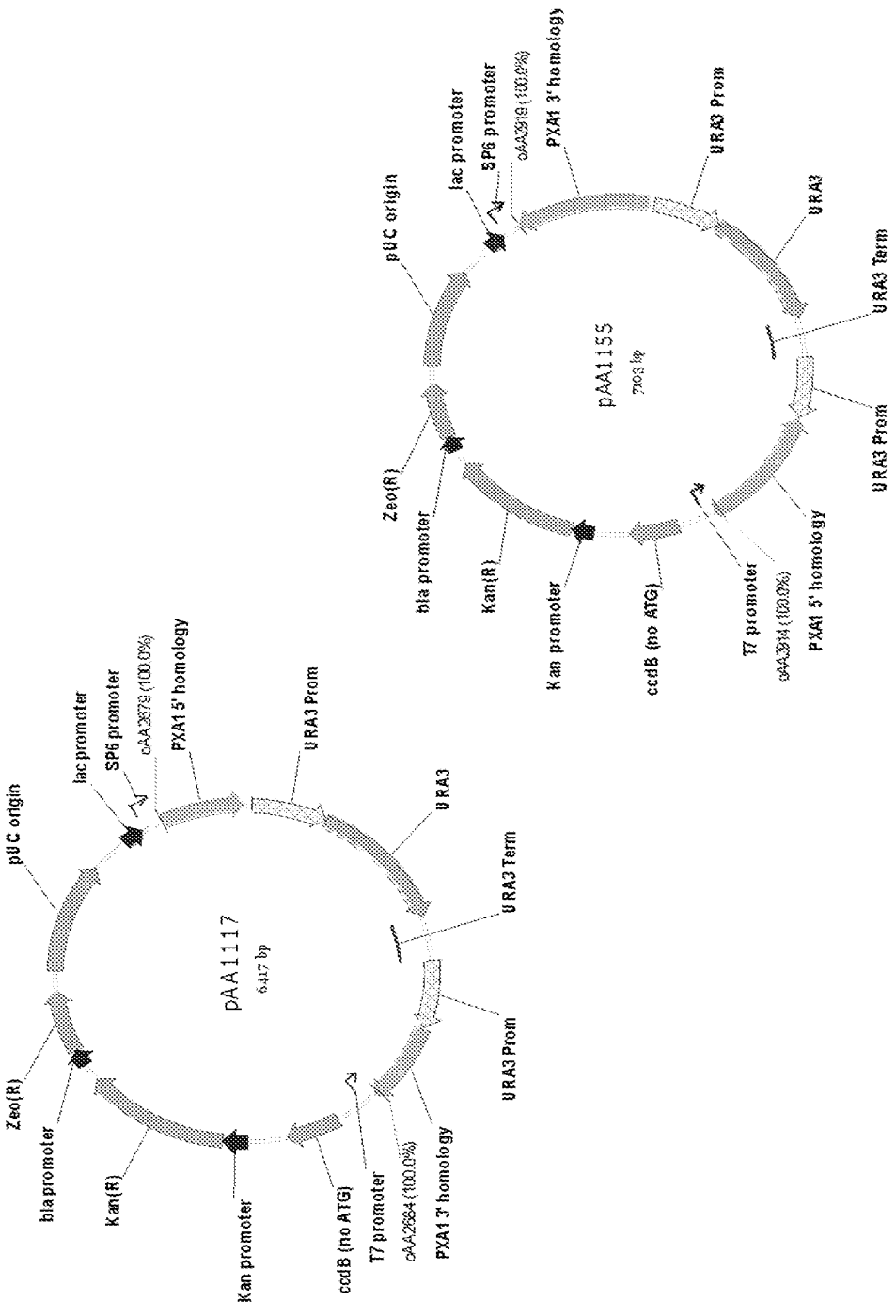

pAA1117
6417 bp
pUC origin
lac promoter
SP6 promoter
cAA28278 (100.0%)
PXA1 5' homology
URA3 Prom
URA3
URA3 Term
URA3 Prom
PXA1 3' homology
cAA28284 (100.0%)
T7 promoter
ccdB (no ATG)
Kan promoter
Kan(R)
bla promoter
Zeo(R)
pAA1155
7703 bp
pUC origin
lac promoter
SP6 promoter
cAA28278 (100.0%)
PXA1 3' homology
URA3 Prom
URA3
URA3 Term
URA3 Prom
PXA1 5' homology
cAA28314 (100.0%)
T7 promoter
ccdB (no ATG)
Kan promoter
Kan(R)
bla promoter
Zeo(R)

FIG. 40

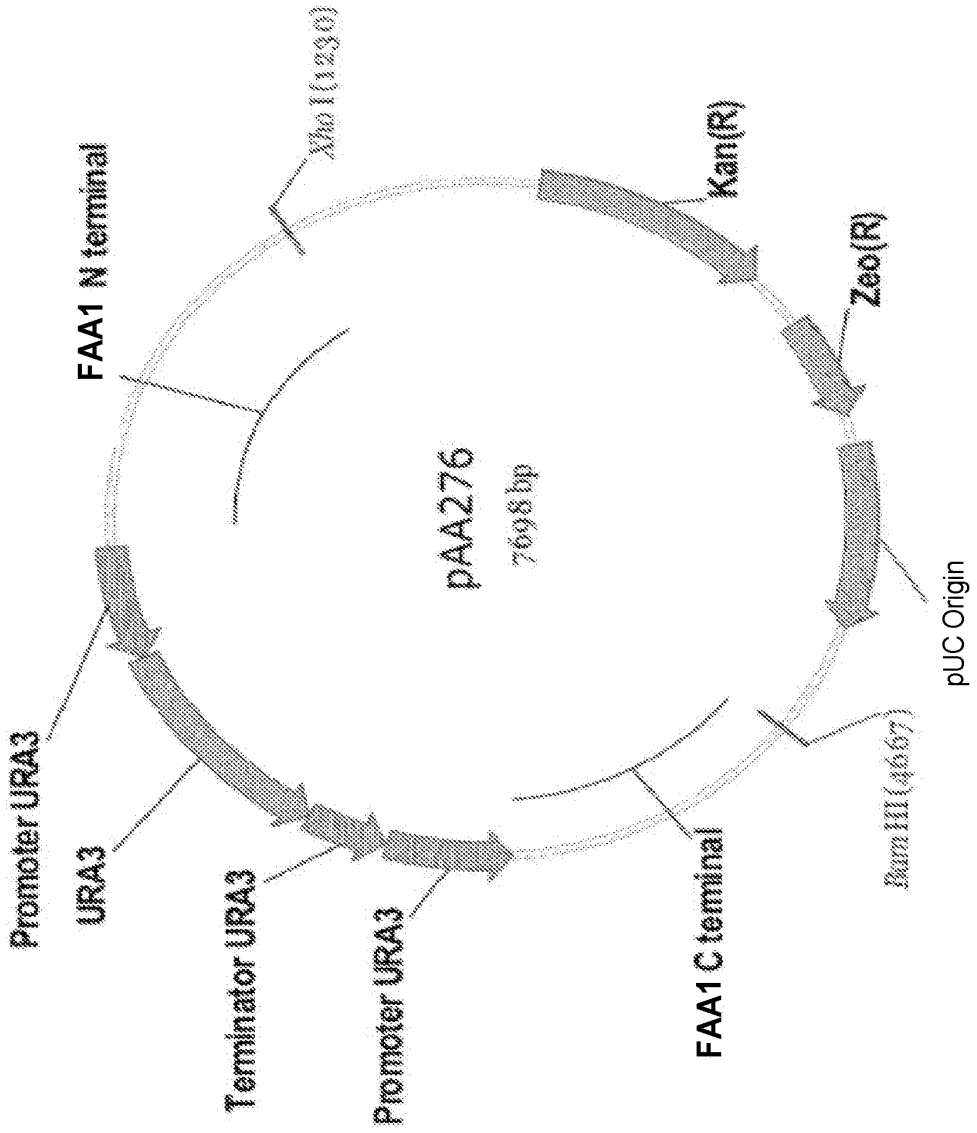


FIG. 42

REPLACEMENT SHEET

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 32 of 3077

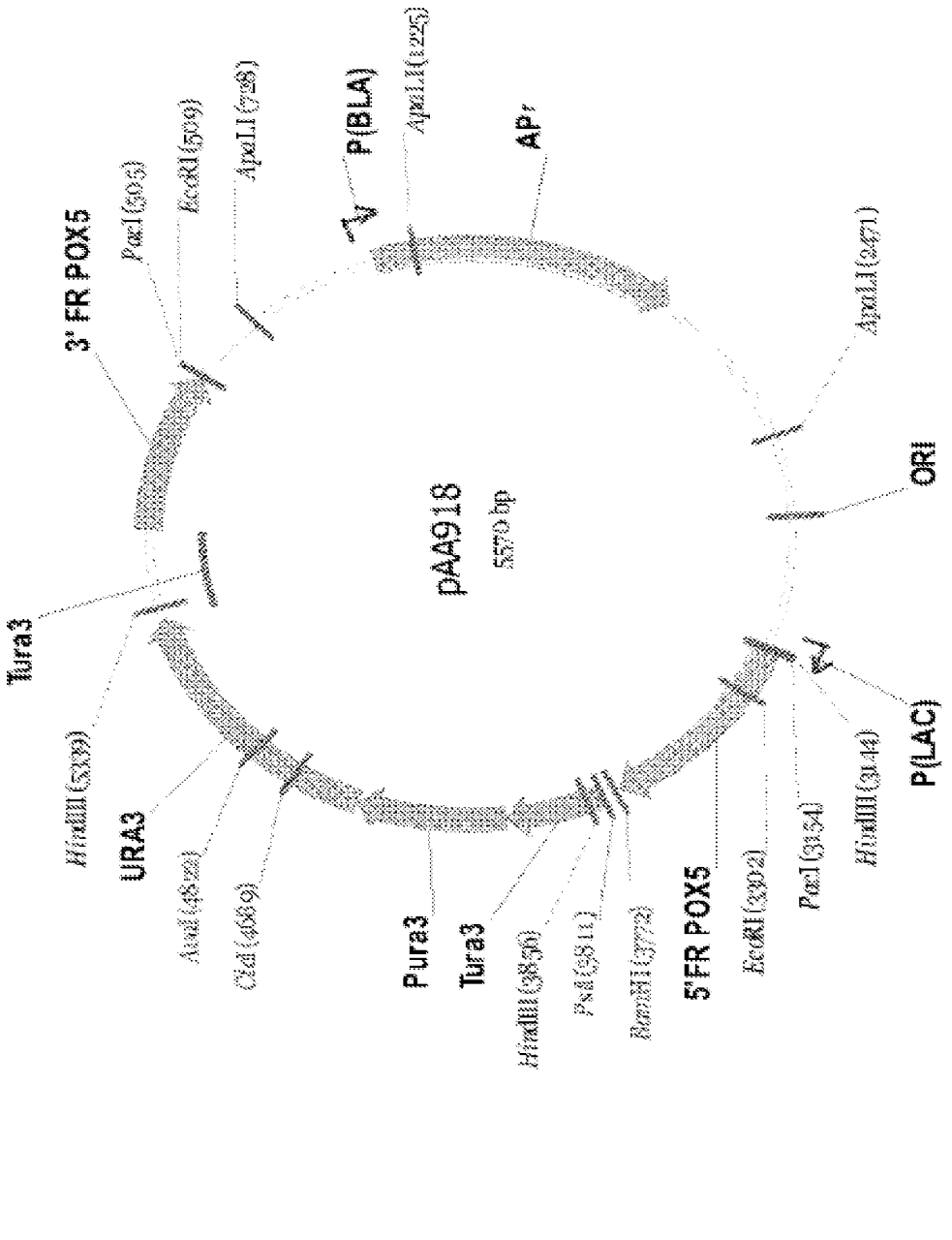


FIG. 43

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 33 of 3077

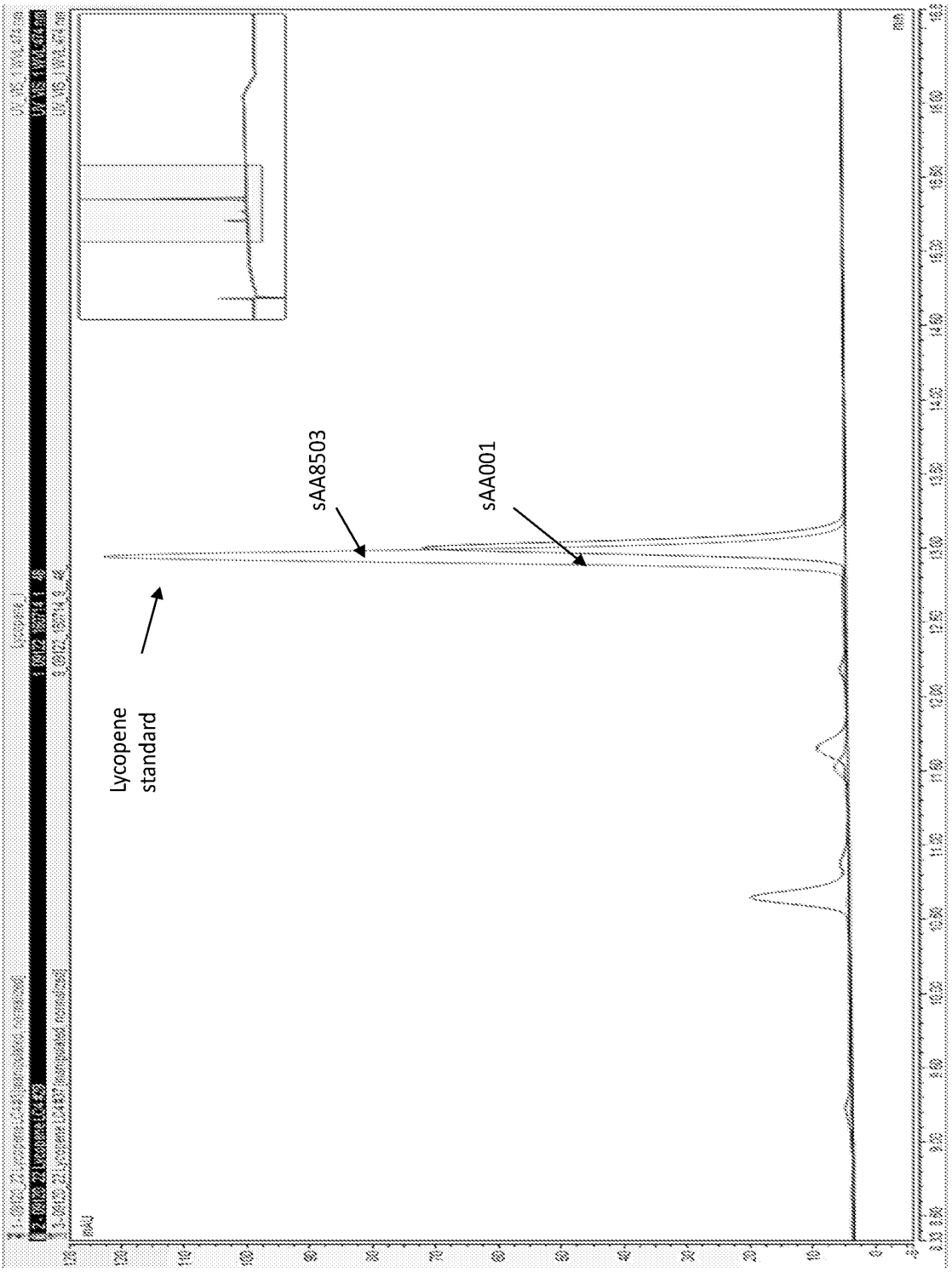


FIG. 57

REMARKS

The specification has been amended herein to add the cross-referencing priority information.

Claims 1-162 are cancelled and claims 163-183 are added. Support for the new claims can be found in the claim amendment filed under PCT Article 34 filed June 10, 2019 in response to the Written Opinion and in the original claims.

It is respectfully submitted that no new matter has been entered and that the present application is in all respects complete and in condition for favorable consideration.

If the Examiner has any questions regarding the claim amendments presented herein, it is requested that the Examiner contact the undersigned at the telephone number shown below.

Respectfully submitted

SILVIA SALVADORI, P.C.

January 13, 2020

By: /Silvia Salvadori/

Silvia Salvadori, Reg. No. 48,265
270 Madison Avenue, 8th Floor
New York, NY 10016
Phone: (212) 897-1938
Fax: (646) 219-6529

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 35 of 3077

AMENDMENTS TO THE CLAIMS

This listing of claims will replace all prior versions, and listings, of the claims in this application.

1-162. (Cancelled).

163. (New) A genetically modified microorganism, comprising:

one or more heterologous nucleic acids encoding one or more terpene biosynthesis polypeptides, wherein expression of at least one of the heterologous nucleic acids is regulated by a nucleic acid that provides for fatty acid or alkane induction of expression of the terpene biosynthesis polypeptide:

- wherein the one or more heterologous nucleic acids that encode one or more terpene biosynthesis polypeptides is chosen from terpene synthase, phytoene synthase, geranylgeranyl diphosphate synthase, phytoene desaturase, lycopene cyclase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase, zeaxanthin glucosyltransferase and valencene synthase, or

- wherein the one or more heterologous nucleic acids is a cytochrome P450 reductase and is regulated by a nucleic acid that provides for alkane induction of expression of the cytochrome P450 reductase.

164. (New) The microorganism of claim 163, wherein the microorganism is a yeast selected from *Candida spp*, *Yarrowia spp*, *Rhodotorula spp*, *Rhodosporidium spp*, *Cryptococcus spp*, *Trichosporon spp*, *Lipomyces spp*, and *Blastobotrys spp*.

165. (New) The microorganism of claims 164, wherein the fatty acid is a saturated fatty acid or an unsaturated fatty acid selected from one or more of oleic acid,

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 36 of 3077

palmitoleic acid, erucic acid, linoleic acid, palmitic acid, caproic acid, enanthic acid, caprylic acid pelargonic acid, capric acid, undecylic acid, lauric acid, myristic acid, pentadecanoic acid, margaric acid, stearic acid arachidic acid, behenic acid, tridecylic acid, and linolenic acid, and wherein the alkane is selected from one or more of hexane, heptane, nonane, decane, undecane, dodecane, tridecane, tetradecane, pentadecane, hexadecane, heptadecane, and octadecane.

166. (New) A genetically modified microorganism, comprising

one or more heterologous nucleic acids encoding one or more terpene biosynthesis polypeptides, and

a genetic modification that alters the expression of a polypeptide providing for transport of acetyl-carnitine in the microorganism,

wherein the expression of at least one of the heterologous nucleic acids is regulated by a nucleic acid that provides for glucose, fatty acid or alkane induction of expression of the terpene biosynthesis polypeptide.

167. (New) The genetically modified microorganism of claim 166, wherein the microorganism is a fungus selected from *Candida* spp, *Yarrowia* spp, *Rhodotorula* spp, *Rhodosporidium* spp, *Cryptococcus* spp, *Trichosporon* spp, *Lipomyces* spp, and *Blastobotrys* spp.

168. (New) The genetically modified microorganism of claim 166, wherein the one or more heterologous nucleic acids encode one or more terpene biosynthesis polypeptides selected from terpene synthase, phytoene synthase, geranylgeranyl diphosphate synthase, phytoene desaturase, lycopene cyclase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 37 of 3077

synthase, zeaxanthin glucosyltransferase, valencene synthase, and cytochrome p450 reductase.

169. (New) The microorganism of claim 166, wherein the microorganism has been genetically modified to increase expression of an endogenous ERG20 gene, wherein expression of ERG20is regulated by a promoter region of the gene encoding oleate-induced peroxisomal protein (POX18), and wherein expression of ERG20is regulated by a terminator region of the gene encoding oleate-induced peroxisomal protein (POX18).

170. (New) A genetically modified *Candida viswanathii* yeast, comprising one or more heterologous nucleic acids encoding one or more terpene biosynthesis polypeptides, wherein expression of at least one of the heterologous nucleic acids is regulated by a nucleic acid that provides for glucose, fatty acid or alkane induction of expression of the terpene biosynthesis polypeptide and

- wherein the one or more heterologous nucleic acids encoding one or more terpene biosynthesis polypeptides is selected from terpene synthase, phytoene synthase, geranylgeranyl diphosphate synthase, phytoene desaturase, lycopene cyclase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase, zeaxanthin glucosyltransferase and valencene synthase, or

- wherein the one or more heterologous nucleic acids is a cytochrome P450 reductase and is regulated by a nucleic acid that provides for glucose or alkane induction of expression of the cytochrome P450 reductase.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 38 of 3077

171. (New) The genetically modified *Candida viswanathii* yeast of claim 170, wherein the one or more heterologous nucleic acids encode phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase.

172. (New) The genetically modified *Candida viswanathii* yeast of claim 171, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a promoter region of the gene encoding a *Candida* hydratase-dehydrogenase-epimerase (HDE), wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a terminator region of the gene encoding a *Candida*acyl co-A oxidase 4 (POX4), wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a promoter region of the gene encoding a *Candida* glyceraldehyde-3-phosphate dehydrogenase (GPD), and wherein the one or more heterologous nucleic acids encode valencene synthase.

173. (New) A method for producing a terpene comprising:

contacting the genetically modified microorganism of claim 163 with a feedstock comprising a carbon source, and

culturing the microorganism under conditions in which the terpenes are produced from the feedstock.

174. (New) A method for producing a terpene comprising:

contacting the genetically modified microorganism of claim 166 with a feedstock comprising a carbon source, and

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 39 of 3077

culturing the microorganism under conditions in which the terpenes are produced from the feedstock.

175. (New) A method for producing a terpene comprising:

contacting the genetically modified *Candida viswanathii* yeast of claim 170 with a feedstock comprising a carbon source, and

culturing the microorganism under conditions in which the terpenes are produced from the feedstock.

176. (New) The method of claim 173, wherein the feedstock comprises one or more saturated fatty acids or one or more unsaturated fatty acids; or one or more saturated fatty acids and one or more unsaturated fatty acids.

177. (New) The method of claim 176, wherein the feedstock comprises one or more fatty acids chosen from oleic acid, palmitoleic acid, erucic acid, linoleic acid, palmitic acid, caproic acid, enanthic acid, caprylic acid pelargonic acid, capric acid, undecylic acid, lauric acid, myristic acid, pentadecanoic acid, margaric acid, stearic acid arachidic acid, behenic acid, tridecylic acid, and linolenic acid.

178. (New) The method of claim 177 wherein the feedstock comprises oleic acid.

179. (New) The method of claim 173, wherein the feedstock comprises one or more vegetable oils, derivatives thereof, or byproducts thereof.

180. (New) The method of claim 179, wherein the feedstock comprises crude palm oil (CPO).

181. (New) A genetically modified yeast comprising one or more heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 40 of 3077

synthase and phytoene desaturase, wherein the amount and/or activity of a Ras2 protein has been decreased.

182. (New) A genetically modified yeast comprising one or more heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase, wherein the amount and/or activity of a Ras2 protein has been decreased.

183. (New) A genetically modified yeast comprising one or more heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase,and cytochrome p450 reductase, wherein the amount and/or activity of a Ras2 protein has been decreased.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 41 of 3077

AMENDMENTS TO THE SPECIFICATION

After the Title, please add the following cross-referencing paragraph:

This application is a U.S. national stage of PCT/US2018/041579 filed on 11 July 2018, which claims priority to and the benefit of U.S. Provisional Application No. 62/532,297, filed on 13 July 2017, the contents of which are incorporated herein by reference in their entireties.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 42 of 3077

Please substitute drawing sheets 12-14, 17-20, 24, 26, 28-30, 32, 35, 38, 40, 42-43 and 57 originally filed with the the replacement drawing sheets attached hereto.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 43 of 3077

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application	:	Aeling K, et al.
International Application No.	:	PCT/US2018/041579
International Filing Date	:	11 July 2018
For	:	**BIOLOGICAL METHODS FOR PREPARING TERPENES**

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

PRELIMINARY AMENDMENT

Sir:

Prior to the issuance of a first Office Action, please amend the above-identified application as follows:

Amendments to the Specification begin on page 2 of this paper.

Amendments to the Claims begin on page 4 of this paper.

Remarks begin on page 10 of this paper.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 44 of 3077

SEQUENCE LISTING

<110> VERDEZYNE (ABC), LLC

<120> BIOLOGICAL METHODS FOR PREPARING TERPENES

<130> VAL-3002-PC

<140> PCT/US2018/041579
<141> 2018-07-11

<150> 62/532,297
<151> 2017-07-13

<160> 527

<170> PatentIn version 3.5

<210> 1
<211> 267
<212> PRT
<213> Candida viswanathii

<400> 1

```
Met Val Ser Thr Lys Thr Tyr Thr Glu Arg Ala Ser Ala His Pro Ser
1               5                  10                  15

Lys Val Ala Gln Arg Leu Phe Arg Leu Met Glu Ser Lys Lys Thr Asn
            20                  25                  30

Leu Cys Ala Ser Ile Asp Val Thr Thr Thr Ala Glu Phe Leu Ser Leu
        35                  40                  45

Ile Asp Lys Leu Gly Pro His Ile Cys Leu Val Lys Thr His Ile Asp
    50                  55                  60

Ile Ile Ser Asp Phe Ser Tyr Glu Gly Thr Ile Glu Pro Leu Leu Val
65                  70                  75                  80

Leu Ala Glu Arg His Gly Phe Leu Ile Phe Glu Asp Arg Lys Phe Ala
                85                  90                  95

Asp Ile Gly Asn Thr Val Met Leu Gln Tyr Thr Ser Gly Val Tyr Arg
            100                 105                 110

Ile Ala Ala Trp Ser Asp Ile Thr Asn Ala His Gly Val Thr Gly Lys
        115                 120                 125

Gly Val Val Glu Gly Leu Lys Arg Gly Ala Glu Gly Val Glu Lys Glu
    130                 135                 140

Arg Gly Val Leu Met Leu Ala Glu Leu Ser Ser Lys Gly Ser Leu Ala
145                 150                 155                 160
```

```
His Gly Glu Tyr Thr Arg Glu Thr Ile Glu Ile Ala Lys Ser Asp Arg
                165                 170                 175

Glu Phe Val Ile Gly Phe Ile Ala Gln Arg Asp Met Gly Gly Arg Glu
            180                 185                 190

Glu Gly Phe Asp Trp Ile Ile Met Thr Pro Gly Val Gly Leu Asp Asp
        195                 200                 205

Lys Gly Asp Ala Leu Gly Gln Gln Tyr Arg Thr Val Asp Glu Val Val
    210                 215                 220

Leu Thr Gly Thr Asp Val Ile Ile Val Gly Arg Gly Leu Phe Gly Lys
225                 230                 235                 240

Gly Arg Asp Pro Glu Val Glu Gly Lys Arg Tyr Arg Asp Ala Gly Trp
                245                 250                 255

Lys Ala Tyr Leu Lys Arg Thr Gly Gln Leu Glu
            260                 265


<210> 2
<211> 627
<212> PRT
<213> Candida viswanathii

<400> 2
Met Phe Asn Phe Lys Leu Ser Gln Gln Val Leu Lys Asn Ser Thr Lys
1               5                   10                  15

Ser Ile Met Pro Ile Leu Lys Lys Pro Phe Ser Thr Ser His Ala Lys
            20                  25                  30

Gly Asp Leu Phe Lys Tyr Gln Ser Gln Leu Pro Lys Leu Pro Val Pro
        35                  40                  45

Thr Leu Glu Glu Thr Ala Ser Lys Tyr Leu Lys Thr Val Glu Pro Phe
    50                  55                  60

Leu Asn Gln Glu Gln Leu Glu Ser Thr Lys Ala Lys Val Ala Glu Phe
65                  70                  75                  80

Val Arg Pro Gly Gly Ala Gly Glu Ala Leu Gln Ala Arg Leu Asn Asn
                85                  90                  95

Phe Ala Ala Asp Lys Asp Asn Trp Leu Ala Glu Phe Trp Asp Asp Tyr
            100                 105                 110
```


LCY Biotechnology Holding, Inc.
Ex. 1002
Page 46 of 3077

```
Ala Tyr Met Ser Tyr Arg Asp Pro Val Val Pro Tyr Val Ser Tyr Phe
        115                 120                 125

Phe Ser His Lys Asp Val Lys Asn Ile Ile Gly Gln Asp Gln Leu Leu
    130                 135                 140

Lys Ala Thr Leu Ile Ala Tyr Tyr Thr Ile Glu Phe Gln Glu Lys Val
145                 150                 155                 160

Leu Asp Glu Ser Leu Asp Pro Glu Val Ile Lys Gly Asn Pro Phe Cys
                165                 170                 175

Met Asn Ala Phe Lys Tyr Met Phe Asn Asn Ser Arg Val Pro Ala Glu
            180                 185                 190

Gly Ser Asp Ile Thr Gln His Tyr Asn Gly Glu Glu Asn Gln Phe Phe
        195                 200                 205

Val Val Ile Tyr Lys Asn Asn Phe Tyr Lys Val Pro Thr His Lys Asn
    210                 215                 220

Gly Gln Arg Leu Thr Lys Gly Glu Ile Tyr Ser Tyr Leu Gln Glu Ile
225                 230                 235                 240

Lys Asn Asp Ala Thr Pro Lys Gly Leu Gly Leu Gly Ala Leu Thr Ser
                245                 250                 255

Leu Asn Arg Asp Glu Trp Leu Ser Ala Tyr Asn Asn Leu Leu Lys Ser
            260                 265                 270

Pro Ile Asn Glu Ala Ser Leu Gly Ser Ile Phe Ala Ser Ser Phe Val
        275                 280                 285

Ile Ala Leu Asp Ser Asn Asn Pro Val Thr Ile Glu Glu Lys Ser Lys
    290                 295                 300

Asn Cys Trp His Gly Asp Gly Gln Asn Arg Phe Phe Asp Lys Pro Leu
305                 310                 315                 320

Glu Phe Phe Val Ser Ala Asn Gly Asn Ser Gly Phe Leu Gly Glu His
                325                 330                 335

Ser Arg Met Asp Ala Thr Pro Thr Val Gln Leu Asn Asn Thr Ile Tyr
            340                 345                 350

Lys Gln Ile Leu Glu Thr Asn Pro Asn Asp Leu Ile Val Glu Ile Gly
        355                 360                 365
```


LCY Biotechnology Holding, Inc.
Ex. 1002
Page 47 of 3077

```
Ser Ser Ala Pro Arg Phe Gly Asn Ala Glu Ile Leu Pro Phe Asp Ile
    370                 375                 380

Asn Pro Thr Thr Arg Ala Asn Ile Lys Asp Ala Ile Ala Lys Phe Asp
385                 390                 395                 400

Ala Thr Ile Ala Ala His Asp Glu Glu Ile Phe Gln His Tyr Gly Tyr
                405                 410                 415

Gly Lys Gly Leu Ile Lys Lys Phe Lys Val Ser Pro Asp Ala Tyr Val
            420                 425                 430

Gln Leu Leu Met Gln Leu Ala Tyr Phe Lys Tyr Thr Gly Lys Ile Arg
        435                 440                 445

Pro Thr Tyr Glu Ser Ala Ala Thr Arg Lys Phe Leu Lys Gly Arg Thr
    450                 455                 460

Glu Thr Gly Arg Thr Val Ser Asn Glu Ser Lys Lys Phe Val Glu Thr
465                 470                 475                 480

Trp Ser Asp Pro Asn Ala Ser Ser Ala Asp Lys Val Ala Thr Phe Gln
                485                 490                 495

Ala Ala Ala Lys Gln His Val Ala Tyr Leu Ser Ala Ala Ala Asp Gly
            500                 505                 510

Lys Gly Val Asp Arg His Leu Phe Gly Leu Lys Gln Met Ile Gln Pro
        515                 520                 525

Gly Glu Pro Ile Pro Glu Ile Phe Thr Asp Pro Ile Phe Ser Tyr Ser
    530                 535                 540

Gln Thr Trp Tyr Ile Ser Ser Ser Gln Val Pro Ser Glu Phe Phe Gln
545                 550                 555                 560

Ser Trp Gly Trp Ser Gln Val Ile Asp Asp Gly Phe Gly Leu Ala Tyr
                565                 570                 575

Leu Ile Asn Asn Asp Trp Ile His Val His Ile Ser Cys Lys Arg Gly
            580                 585                 590

Asn Gly Leu Gln Ser Asp His Leu Lys Trp Tyr Leu Val Asp Ser Ala
        595                 600                 605

Asn Glu Met Lys Asp Val Leu Thr Lys Gly Leu Leu Thr Asp Ala Lys
    610                 615                 620
```


LCY Biotechnology Holding, Inc.
Ex. 1002
Page 48 of 3077

Pro Lys Leu
625

<210> 3
<211> 609
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 3
Met Pro Ile Leu Lys Lys Pro Phe Ser Thr Ser His Ala Lys Gly Asp
1 5 10 15

Leu Phe Lys Tyr Gln Ser Gln Leu Pro Lys Leu Pro Val Pro Thr Leu
20 25 30

Glu Glu Thr Ala Ser Lys Tyr Leu Lys Thr Val Glu Pro Phe Leu Asn
35 40 45

Gln Glu Gln Leu Glu Ser Thr Lys Ala Lys Val Ala Glu Phe Val Arg
50 55 60

Pro Gly Gly Ala Gly Glu Ala Leu Gln Ala Arg Leu Asn Asn Phe Ala
65 70 75 80

Ala Asp Lys Asp Asn Trp Leu Ala Glu Phe Trp Asp Asp Tyr Ala Tyr
85 90 95

Met Ser Tyr Arg Asp Pro Val Val Pro Tyr Val Ser Tyr Phe Phe Ser
100 105 110

His Lys Asp Val Lys Asn Ile Ile Gly Gln Asp Gln Leu Leu Lys Ala
115 120 125

Thr Leu Ile Ala Tyr Tyr Thr Ile Glu Phe Gln Glu Lys Val Leu Asp
130 135 140

Glu Ser Leu Asp Pro Glu Val Ile Lys Gly Asn Pro Phe Cys Met Asn
145 150 155 160

Ala Phe Lys Tyr Met Phe Asn Asn Ser Arg Val Pro Ala Glu Gly Ser
165 170 175

Asp Ile Thr Gln His Tyr Asn Gly Glu Glu Asn Gln Phe Phe Val Val
180 185 190

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 49 of 3077

```
Ile Tyr Lys Asn Asn Phe Tyr Lys Val Pro Thr His Lys Asn Gly Gln
        195                 200                 205

Arg Leu Thr Lys Gly Glu Ile Tyr Ser Tyr Leu Gln Glu Ile Lys Asn
    210                 215                 220

Asp Ala Thr Pro Lys Gly Leu Gly Leu Gly Ala Leu Thr Ser Leu Asn
225                 230                 235                 240

Arg Asp Glu Trp Leu Ser Ala Tyr Asn Asn Leu Leu Lys Ser Pro Ile
                245                 250                 255

Asn Glu Ala Ser Leu Gly Ser Ile Phe Ala Ser Ser Phe Val Ile Ala
            260                 265                 270

Leu Asp Ser Asn Asn Pro Val Thr Ile Glu Glu Lys Ser Lys Asn Cys
        275                 280                 285

Trp His Gly Asp Gly Gln Asn Arg Phe Phe Asp Lys Pro Leu Glu Phe
    290                 295                 300

Phe Val Ser Ala Asn Gly Asn Ser Gly Phe Leu Gly Glu His Ser Arg
305                 310                 315                 320

Met Asp Ala Thr Pro Thr Val Gln Leu Asn Asn Thr Ile Tyr Lys Gln
                325                 330                 335

Ile Leu Glu Thr Asn Pro Asn Asp Leu Ile Val Glu Ile Gly Ser Ser
            340                 345                 350

Ala Pro Arg Phe Gly Asn Ala Glu Ile Leu Pro Phe Asp Ile Asn Pro
        355                 360                 365

Thr Thr Arg Ala Asn Ile Lys Asp Ala Ile Ala Lys Phe Asp Ala Thr
    370                 375                 380

Ile Ala Ala His Asp Glu Glu Ile Phe Gln His Tyr Gly Tyr Gly Lys
385                 390                 395                 400

Gly Leu Ile Lys Lys Phe Lys Val Ser Pro Asp Ala Tyr Val Gln Leu
                405                 410                 415

Leu Met Gln Leu Ala Tyr Phe Lys Tyr Thr Gly Lys Ile Arg Pro Thr
            420                 425                 430

Tyr Glu Ser Ala Ala Thr Arg Lys Phe Leu Lys Gly Arg Thr Glu Thr
        435                 440                 445
```


LCY Biotechnology Holding, Inc.
Ex. 1002
Page 50 of 3077

```
Gly Arg Thr Val Ser Asn Glu Ser Lys Lys Phe Val Glu Thr Trp Ser
    450                 455                 460

Asp Pro Asn Ala Ser Ser Ala Asp Lys Val Ala Thr Phe Gln Ala Ala
465                 470                 475                 480

Ala Lys Gln His Val Ala Tyr Leu Ser Ala Ala Ala Asp Gly Lys Gly
                485                 490                 495

Val Asp Arg His Leu Phe Gly Leu Lys Gln Met Ile Gln Pro Gly Glu
            500                 505                 510

Pro Ile Pro Glu Ile Phe Thr Asp Pro Ile Phe Ser Tyr Ser Gln Thr
        515                 520                 525

Trp Tyr Ile Ser Ser Ser Gln Val Pro Ser Glu Phe Phe Gln Ser Trp
    530                 535                 540

Gly Trp Ser Gln Val Ile Asp Asp Gly Phe Gly Leu Ala Tyr Leu Ile
545                 550                 555                 560

Asn Asn Asp Trp Ile His Val His Ile Ser Cys Lys Arg Gly Asn Gly
                565                 570                 575

Leu Gln Ser Asp His Leu Lys Trp Tyr Leu Val Asp Ser Ala Asn Glu
            580                 585                 590

Met Lys Asp Val Leu Thr Lys Gly Leu Leu Thr Asp Ala Lys Pro Lys
        595                 600                 605

Leu
```

<210> 4
<211> 606
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 4

```
Met Pro Ile Leu Lys Lys Pro Phe Ser Thr Ser His Ala Lys Gly Asp
1               5                   10                  15

Leu Phe Lys Tyr Gln Ser Gln Leu Pro Lys Leu Pro Val Pro Thr Leu
            20                  25                  30
```


LCY Biotechnology Holding, Inc.
Ex. 1002
Page 51 of 3077

```
Glu Glu Thr Ala Ser Lys Tyr Leu Lys Thr Val Glu Pro Phe Leu Asn
        35                  40                  45

Gln Glu Gln Leu Glu Ser Thr Lys Ala Lys Val Ala Glu Phe Val Arg
    50                  55                  60

Pro Gly Gly Ala Gly Glu Ala Leu Gln Ala Arg Leu Asn Asn Phe Ala
65                  70                  75                  80

Ala Asp Lys Asp Asn Trp Leu Ala Glu Phe Trp Asp Asp Tyr Ala Tyr
                85                  90                  95

Met Ser Tyr Arg Asp Pro Val Val Pro Tyr Val Ser Tyr Phe Phe Ser
            100                 105                 110

His Lys Asp Val Lys Asn Ile Ile Gly Gln Asp Gln Leu Leu Lys Ala
        115                 120                 125

Thr Leu Ile Ala Tyr Tyr Thr Ile Glu Phe Gln Glu Lys Val Leu Asp
    130                 135                 140

Glu Ser Leu Asp Pro Glu Val Ile Lys Gly Asn Pro Phe Cys Met Asn
145                 150                 155                 160

Ala Phe Lys Tyr Met Phe Asn Asn Ser Arg Val Pro Ala Glu Gly Ser
                165                 170                 175

Asp Ile Thr Gln His Tyr Asn Gly Glu Glu Asn Gln Phe Phe Val Val
            180                 185                 190

Ile Tyr Lys Asn Asn Phe Tyr Lys Val Pro Thr His Lys Asn Gly Gln
        195                 200                 205

Arg Leu Thr Lys Gly Glu Ile Tyr Ser Tyr Leu Gln Glu Ile Lys Asn
    210                 215                 220

Asp Ala Thr Pro Lys Gly Leu Gly Leu Gly Ala Leu Thr Ser Leu Asn
225                 230                 235                 240

Arg Asp Glu Trp Leu Ser Ala Tyr Asn Asn Leu Leu Lys Ser Pro Ile
                245                 250                 255

Asn Glu Ala Ser Leu Gly Ser Ile Phe Ala Ser Ser Phe Val Ile Ala
            260                 265                 270

Leu Asp Ser Asn Asn Pro Val Thr Ile Glu Glu Lys Ser Lys Asn Cys
        275                 280                 285
```

```
Trp His Gly Asp Gly Gln Asn Arg Phe Phe Asp Lys Pro Leu Glu Phe
    290                 295                 300

Phe Val Ser Ala Asn Gly Asn Ser Gly Phe Leu Gly Glu His Ser Arg
305                 310                 315                 320

Met Asp Ala Thr Pro Thr Val Gln Leu Asn Asn Thr Ile Tyr Lys Gln
                325                 330                 335

Ile Leu Glu Thr Asn Pro Asn Asp Leu Ile Val Glu Ile Gly Ser Ser
            340                 345                 350

Ala Pro Arg Phe Gly Asn Ala Glu Ile Leu Pro Phe Asp Ile Asn Pro
        355                 360                 365

Thr Thr Arg Ala Asn Ile Lys Asp Ala Ile Ala Lys Phe Asp Ala Thr
    370                 375                 380

Ile Ala Ala His Asp Glu Glu Ile Phe Gln His Tyr Gly Tyr Gly Lys
385                 390                 395                 400

Gly Leu Ile Lys Lys Phe Lys Val Ser Pro Asp Ala Tyr Val Gln Leu
                405                 410                 415

Leu Met Gln Leu Ala Tyr Phe Lys Tyr Thr Gly Lys Ile Arg Pro Thr
            420                 425                 430

Tyr Glu Ser Ala Ala Thr Arg Lys Phe Leu Lys Gly Arg Thr Glu Thr
        435                 440                 445

Gly Arg Thr Val Ser Asn Glu Ser Lys Lys Phe Val Glu Thr Trp Ser
    450                 455                 460

Asp Pro Asn Ala Ser Ser Ala Asp Lys Val Ala Thr Phe Gln Ala Ala
465                 470                 475                 480

Ala Lys Gln His Val Ala Tyr Leu Ser Ala Ala Ala Asp Gly Lys Gly
                485                 490                 495

Val Asp Arg His Leu Phe Gly Leu Lys Gln Met Ile Gln Pro Gly Glu
            500                 505                 510

Pro Ile Pro Glu Ile Phe Thr Asp Pro Ile Phe Ser Tyr Ser Gln Thr
        515                 520                 525

Trp Tyr Ile Ser Ser Ser Gln Val Pro Ser Glu Phe Phe Gln Ser Trp
```


LCY Biotechnology Holding, Inc.
Ex. 1002
Page 53 of 3077

530 535 540

Gly Trp Ser Gln Val Ile Asp Asp Gly Phe Gly Leu Ala Tyr Leu Ile
545 550 555 560

Asn Asn Asp Trp Ile His Val His Ile Ser Cys Lys Arg Gly Asn Gly
565 570 575

Leu Gln Ser Asp His Leu Lys Trp Tyr Leu Val Asp Ser Ala Asn Glu
580 585 590

Met Lys Asp Val Leu Thr Lys Gly Leu Leu Thr Asp Ala Lys
595 600 605

<210> 5
<211> 624
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 5
Met Phe Asn Phe Lys Leu Ser Gln Gln Val Leu Lys Asn Ser Thr Lys
1 5 10 15

Ser Ile Met Pro Ile Leu Lys Lys Pro Phe Ser Thr Ser His Ala Lys
20 25 30

Gly Asp Leu Phe Lys Tyr Gln Ser Gln Leu Pro Lys Leu Pro Val Pro
35 40 45

Thr Leu Glu Glu Thr Ala Ser Lys Tyr Leu Lys Thr Val Glu Pro Phe
50 55 60

Leu Asn Gln Glu Gln Leu Glu Ser Thr Lys Ala Lys Val Ala Glu Phe
65 70 75 80

Val Arg Pro Gly Gly Ala Gly Glu Ala Leu Gln Ala Arg Leu Asn Asn
85 90 95

Phe Ala Ala Asp Lys Asp Asn Trp Leu Ala Glu Phe Trp Asp Asp Tyr
100 105 110

Ala Tyr Met Ser Tyr Arg Asp Pro Val Val Pro Tyr Val Ser Tyr Phe
115 120 125

Phe Ser His Lys Asp Val Lys Asn Ile Ile Gly Gln Asp Gln Leu Leu
130 135 140

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 54 of 3077

```
Lys Ala Thr Leu Ile Ala Tyr Tyr Thr Ile Glu Phe Gln Glu Lys Val
145                 150                 155                 160

Leu Asp Glu Ser Leu Asp Pro Glu Val Ile Lys Gly Asn Pro Phe Cys
                165                 170                 175

Met Asn Ala Phe Lys Tyr Met Phe Asn Asn Ser Arg Val Pro Ala Glu
            180                 185                 190

Gly Ser Asp Ile Thr Gln His Tyr Asn Gly Glu Glu Asn Gln Phe Phe
        195                 200                 205

Val Val Ile Tyr Lys Asn Asn Phe Tyr Lys Val Pro Thr His Lys Asn
    210                 215                 220

Gly Gln Arg Leu Thr Lys Gly Glu Ile Tyr Ser Tyr Leu Gln Glu Ile
225                 230                 235                 240

Lys Asn Asp Ala Thr Pro Lys Gly Leu Gly Leu Gly Ala Leu Thr Ser
                245                 250                 255

Leu Asn Arg Asp Glu Trp Leu Ser Ala Tyr Asn Asn Leu Leu Lys Ser
            260                 265                 270

Pro Ile Asn Glu Ala Ser Leu Gly Ser Ile Phe Ala Ser Ser Phe Val
        275                 280                 285

Ile Ala Leu Asp Ser Asn Asn Pro Val Thr Ile Glu Glu Lys Ser Lys
    290                 295                 300

Asn Cys Trp His Gly Asp Gly Gln Asn Arg Phe Phe Asp Lys Pro Leu
305                 310                 315                 320

Glu Phe Phe Val Ser Ala Asn Gly Asn Ser Gly Phe Leu Gly Glu His
                325                 330                 335

Ser Arg Met Asp Ala Thr Pro Thr Val Gln Leu Asn Asn Thr Ile Tyr
            340                 345                 350

Lys Gln Ile Leu Glu Thr Asn Pro Asn Asp Leu Ile Val Glu Ile Gly
        355                 360                 365

Ser Ser Ala Pro Arg Phe Gly Asn Ala Glu Ile Leu Pro Phe Asp Ile
    370                 375                 380

Asn Pro Thr Thr Arg Ala Asn Ile Lys Asp Ala Ile Ala Lys Phe Asp
```

385 390 395 400

Ala Thr Ile Ala Ala His Asp Glu Glu Ile Phe Gln His Tyr Gly Tyr
405 410 415

Gly Lys Gly Leu Ile Lys Lys Phe Lys Val Ser Pro Asp Ala Tyr Val
420 425 430

Gln Leu Leu Met Gln Leu Ala Tyr Phe Lys Tyr Thr Gly Lys Ile Arg
435 440 445

Pro Thr Tyr Glu Ser Ala Ala Thr Arg Lys Phe Leu Lys Gly Arg Thr
450 455 460

Glu Thr Gly Arg Thr Val Ser Asn Glu Ser Lys Lys Phe Val Glu Thr
465 470 475 480

Trp Ser Asp Pro Asn Ala Ser Ser Ala Asp Lys Val Ala Thr Phe Gln
485 490 495

Ala Ala Ala Lys Gln His Val Ala Tyr Leu Ser Ala Ala Ala Asp Gly
500 505 510

Lys Gly Val Asp Arg His Leu Phe Gly Leu Lys Gln Met Ile Gln Pro
515 520 525

Gly Glu Pro Ile Pro Glu Ile Phe Thr Asp Pro Ile Phe Ser Tyr Ser
530 535 540

Gln Thr Trp Tyr Ile Ser Ser Ser Gln Val Pro Ser Glu Phe Phe Gln
545 550 555 560

Ser Trp Gly Trp Ser Gln Val Ile Asp Asp Gly Phe Gly Leu Ala Tyr
565 570 575

Leu Ile Asn Asn Asp Trp Ile His Val His Ile Ser Cys Lys Arg Gly
580 585 590

Asn Gly Leu Gln Ser Asp His Leu Lys Trp Tyr Leu Val Asp Ser Ala
595 600 605

Asn Glu Met Lys Asp Val Leu Thr Lys Gly Leu Leu Thr Asp Ala Lys
610 615 620

<210> 6
<211> 766
<212> PRT
<213> Candida viswanathii

<400> 6

```
Met Ser Thr Tyr Gln Phe Gln Glu Thr Leu Glu Lys Leu Pro Ile Pro
1               5                   10                  15

Asp Leu Asn Gln Thr Cys Ala Asn Tyr Leu Asn Val Leu Arg Pro Leu
            20                  25                  30

Gln Thr Glu Gln Glu His Ile Lys Thr Lys Thr Ala Val Glu Asn Phe
        35                  40                  45

Leu Lys Asn Gly Thr Gly Gln Tyr Leu Asp Ala Ala Leu Arg Glu Tyr
    50                  55                  60

Ala Gln Thr Arg Pro Ser Tyr Ile Glu Gln Phe Trp Tyr Asp Ala Tyr
65                  70                  75                  80

Leu Asn Tyr Asp Ser Pro Val Val Leu Asn Leu Asn Pro Phe Phe Leu
                85                  90                  95

Leu Glu Asp Asp Pro Phe Thr Asn Glu Ser Ser Ser Val Asn Pro Gln
            100                 105                 110

Val Lys Arg Ala Thr Ser Leu Val Met Ser Ser Leu Lys Phe Ile Gln
        115                 120                 125

Ala Leu Lys Asn Glu Thr Leu Ser Val Asp Thr Leu Lys Gly Gly Lys
    130                 135                 140

Pro Leu Cys Met Tyr Gln Tyr Thr Lys Leu Phe Gly Ala Ser Arg Ile
145                 150                 155                 160

Pro Ser Glu Asp Gly Cys Val Met Gln Ser Asp Pro Ala Ser Asn His
                165                 170                 175

Ile Val Val Met Ser Lys Ser Gln Leu Tyr Trp Phe Asp Val Leu Asp
            180                 185                 190

Ser Lys Asn Asn Leu Ile Leu Ser Glu Ala Glu Leu Asn Val Asn Phe
        195                 200                 205

Gln Ser Ile Ile His Asp Ser Leu Arg Thr Pro Ser Asp Glu Ile Ala
    210                 215                 220

Lys Ser Ser Phe Gly Val Leu Thr Thr Glu Asn Arg Arg Ile Trp Ala
225                 230                 235                 240

Asn Val Arg His Asn Leu Met Ser Thr Thr Asn Asn Lys Val Asn His
```


LCY Biotechnology Holding, Inc.
Ex. 1002
Page 57 of 3077

245 250 255

Glu Val Leu Ser Ile Ile Asp Ser Ala Leu Phe Val Leu Cys Phe Asp
260 265 270

Asp Ile Val Ile Asn Asp Leu Ser Glu Leu Ser Lys Asn Met Leu Cys
275 280 285

Gly Leu Ser Ile Leu Asp Asn Gly Ile Gln Val Gly Thr Cys Thr Asn
290 295 300

Arg Trp Tyr Asp Lys Leu Gln Ile Ile Val Thr Lys Asn Ala Lys Ala
305 310 315 320

Gly Ile Asn Phe Glu His Thr Gly Val Asp Gly His Thr Val Leu Arg
325 330 335

Phe Val Ser Asp Ile Tyr Thr Asp Ser Ile Leu Ser Phe Ala Asn Ser
340 345 350

Ile Asn Ser Asn Ala Pro Ser Leu Trp Asn Thr Thr Thr Asn Tyr Lys
355 360 365

Glu Leu Asp Gly Glu Asp Leu Ile Thr Val Pro Arg Lys Leu Glu Trp
370 375 380

Glu Leu Thr Pro Asp Leu Ser Leu Ala Leu Arg Phe Gly Glu Thr Arg
385 390 395 400

Leu Ser Asp Leu Ile Asn Gln Asn Glu Phe Arg His Leu Glu Phe Lys
405 410 415

Asn Tyr Gly Ser Thr Gln Ile Lys Lys Met Lys Phe Ser Pro Asp Ala
420 425 430

Phe Val Gln Met Ala Phe Gln Ala Thr Tyr Tyr Ala Leu Tyr Gly Lys
435 440 445

Val Glu Cys Thr Tyr Glu Pro Ala Met Thr Lys Gln Phe Tyr His Gly
450 455 460

Arg Thr Glu Ala Ile Arg Thr Val Ser Gln Glu Ser Asn Leu Phe Val
465 470 475 480

Arg Lys Phe Phe Asp Ser Thr Val Ser Ile Gln Lys Lys Leu Glu Tyr
485 490 495

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 58 of 3077

```
Leu Thr Gln Ala Cys Thr Lys His Ser Gln Gln Thr Arg Met Ser Ser
            500                 505                 510

Ala Gly Gln Gly Val Asp Arg His Leu Tyr Ala Leu Phe Cys Ile Trp
        515                 520                 525

Lys Arg Tyr Leu His Asp Ala Glu Asp Asp Asp Asp Asp Ala Ser Lys
    530                 535                 540

Asp Gln Pro Ser Ser Thr Glu Leu Ser Arg Asp Asp Thr Ile Ile Ala
545                 550                 555                 560

Gln Thr Asn Gly Asp Ser Thr Asp Asp Asp Asp Ser Thr Thr Val Val
                565                 570                 575

Gly His Gly Ala Asn His His His Asn Leu Lys Ser Ala Asp Leu Leu
            580                 585                 590

Lys Thr Ile Pro Glu Ile Phe Ala Asp Asn Gly Trp Asp Lys Leu Asn
        595                 600                 605

Asn Thr Ile Ile Ser Thr Ser Asn Cys Gly Asn Pro Ser Leu Arg Leu
    610                 615                 620

Phe Gly Phe Gly Pro Val Ser Ala Asn Gly Phe Gly Ile Gly Tyr Ile
625                 630                 635                 640

Leu Lys Asp Asp Ser Ile Ser Ile Cys Ala Ser Ser Lys His Arg Gln
                645                 650                 655

Thr Gln Arg Phe Leu Val Thr Leu Asn Ser Tyr Leu Leu Glu Ile Gln
            660                 665                 670

Asn Ile Trp Lys Gln Ala Gln Lys Met Glu Lys Leu Lys Ala Glu Leu
        675                 680                 685

Ala Lys Ala Val Glu Asp Ala Lys Pro Lys Gly Gln Lys Ser Gly Glu
    690                 695                 700

Ala Glu Gly Ile Ile Glu Pro Lys Pro Asn Asn Leu Ser Thr Leu Leu
705                 710                 715                 720

Gly Gly Tyr Gly Tyr Phe Asp Met Gly Asp Glu Asp Ile Lys Ser Arg
                725                 730                 735

Gly Gln Ser Pro Glu Pro Pro Phe Leu His Arg Ala Gly Ser Gly Phe
            740                 745                 750
```

Ser Ile Arg Glu Ile Gly Lys Lys Leu Arg Leu Ser Glu Tyr
755 760 765

<210> 7
<211> 30
<212> PRT
<213> Candida viswanathii

<400> 7
Met Leu Ser Arg Thr Thr Leu Arg Val Ala Arg Gln Gln Thr Arg Leu
1 5 10 15

Leu Ser Thr Ser Arg Ile Leu Phe Asn Ser Lys Thr Asp Gln
20 25 30

<210> 8
<211> 25
<212> PRT
<213> Candida viswanathii

<400> 8
Met Ser Ala Leu Arg Ser Phe Gln Arg Ser Ser Asn Val Ala Lys Ser
1 5 10 15

Thr Leu Lys Asn Ser Val Arg Thr Tyr
20 25

<210> 9
<211> 33
<212> PRT
<213> Candida viswanathii

<400> 9
Met Phe Asn Phe Lys Leu Ser Gln Gln Val Leu Lys Asn Ser Thr Lys
1 5 10 15

Ser Ile Met Pro Ile Leu Lys Lys Pro Phe Ser Thr Ser His Ala Lys
20 25 30

Gly

<210> 10
<211> 795
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
polypeptide

<400> 10
Met Leu Ser Arg Thr Thr Leu Arg Val Ala Arg Gln Gln Thr Arg Leu
1 5 10 15

```
Leu Ser Thr Ser Arg Ile Leu Phe Asn Ser Lys Thr Asp Gln Ser Thr
            20                  25                  30

Tyr Gln Phe Gln Glu Thr Leu Glu Lys Leu Pro Ile Pro Asp Leu Asn
        35                  40                  45

Gln Thr Cys Ala Asn Tyr Leu Asn Val Leu Arg Pro Leu Gln Thr Glu
    50                  55                  60

Gln Glu His Ile Lys Thr Lys Thr Ala Val Glu Asn Phe Leu Lys Asn
65                  70                  75                  80

Gly Thr Gly Gln Tyr Leu Asp Ala Ala Leu Arg Glu Tyr Ala Gln Thr
                85                  90                  95

Arg Pro Ser Tyr Ile Glu Gln Phe Trp Tyr Asp Ala Tyr Leu Asn Tyr
            100                 105                 110

Asp Ser Pro Val Val Leu Asn Leu Asn Pro Phe Phe Leu Leu Glu Asp
        115                 120                 125

Asp Pro Phe Thr Asn Glu Ser Ser Ser Val Asn Pro Gln Val Lys Arg
    130                 135                 140

Ala Thr Ser Leu Val Met Ser Ser Leu Lys Phe Ile Gln Ala Leu Lys
145                 150                 155                 160

Asn Glu Thr Leu Ser Val Asp Thr Leu Lys Gly Gly Lys Pro Leu Cys
                165                 170                 175

Met Tyr Gln Tyr Thr Lys Leu Phe Gly Ala Ser Arg Ile Pro Ser Glu
            180                 185                 190

Asp Gly Cys Val Met Gln Ser Asp Pro Ala Ser Asn His Ile Val Val
        195                 200                 205

Met Ser Lys Ser Gln Leu Tyr Trp Phe Asp Val Leu Asp Ser Lys Asn
    210                 215                 220

Asn Leu Ile Leu Ser Glu Ala Glu Leu Asn Val Asn Phe Gln Ser Ile
225                 230                 235                 240

Ile His Asp Ser Leu Arg Thr Pro Ser Asp Glu Ile Ala Lys Ser Ser
                245                 250                 255

Phe Gly Val Leu Thr Thr Glu Asn Arg Arg Ile Trp Ala Asn Val Arg
```

260 265 270

His Asn Leu Met Ser Thr Thr Asn Asn Lys Val Asn His Glu Val Leu
275 280 285

Ser Ile Ile Asp Ser Ala Leu Phe Val Leu Cys Phe Asp Asp Ile Val
290 295 300

Ile Asn Asp Leu Ser Glu Leu Ser Lys Asn Met Leu Cys Gly Leu Ser
305 310 315 320

Ile Leu Asp Asn Gly Ile Gln Val Gly Thr Cys Thr Asn Arg Trp Tyr
325 330 335

Asp Lys Leu Gln Ile Ile Val Thr Lys Asn Ala Lys Ala Gly Ile Asn
340 345 350

Phe Glu His Thr Gly Val Asp Gly His Thr Val Leu Arg Phe Val Ser
355 360 365

Asp Ile Tyr Thr Asp Ser Ile Leu Ser Phe Ala Asn Ser Ile Asn Ser
370 375 380

Asn Ala Pro Ser Leu Trp Asn Thr Thr Thr Asn Tyr Lys Glu Leu Asp
385 390 395 400

Gly Glu Asp Leu Ile Thr Val Pro Arg Lys Leu Glu Trp Glu Leu Thr
405 410 415

Pro Asp Leu Ser Leu Ala Leu Arg Phe Gly Glu Thr Arg Leu Ser Asp
420 425 430

Leu Ile Asn Gln Asn Glu Phe Arg His Leu Glu Phe Lys Asn Tyr Gly
435 440 445

Ser Thr Gln Ile Lys Lys Met Lys Phe Ser Pro Asp Ala Phe Val Gln
450 455 460

Met Ala Phe Gln Ala Thr Tyr Tyr Ala Leu Tyr Gly Lys Val Glu Cys
465 470 475 480

Thr Tyr Glu Pro Ala Met Thr Lys Gln Phe Tyr His Gly Arg Thr Glu
485 490 495

Ala Ile Arg Thr Val Ser Gln Glu Ser Asn Leu Phe Val Arg Lys Phe
500 505 510

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 62 of 3077

```
Phe Asp Ser Thr Val Ser Ile Gln Lys Lys Leu Glu Tyr Leu Thr Gln
        515                 520                 525

Ala Cys Thr Lys His Ser Gln Gln Thr Arg Met Ser Ser Ala Gly Gln
    530                 535                 540

Gly Val Asp Arg His Leu Tyr Ala Leu Phe Cys Ile Trp Lys Arg Tyr
545                 550                 555                 560

Leu His Asp Ala Glu Asp Asp Asp Asp Asp Ala Ser Lys Asp Gln Pro
                565                 570                 575

Ser Ser Thr Glu Leu Ser Arg Asp Asp Thr Ile Ile Ala Gln Thr Asn
            580                 585                 590

Gly Asp Ser Thr Asp Asp Asp Asp Ser Thr Thr Val Val Gly His Gly
        595                 600                 605

Ala Asn His His His Asn Leu Lys Ser Ala Asp Leu Leu Lys Thr Ile
    610                 615                 620

Pro Glu Ile Phe Ala Asp Asn Gly Trp Asp Lys Leu Asn Asn Thr Ile
625                 630                 635                 640

Ile Ser Thr Ser Asn Cys Gly Asn Pro Ser Leu Arg Leu Phe Gly Phe
                645                 650                 655

Gly Pro Val Ser Ala Asn Gly Phe Gly Ile Gly Tyr Ile Leu Lys Asp
            660                 665                 670

Asp Ser Ile Ser Ile Cys Ala Ser Ser Lys His Arg Gln Thr Gln Arg
        675                 680                 685

Phe Leu Val Thr Leu Asn Ser Tyr Leu Leu Glu Ile Gln Asn Ile Trp
    690                 695                 700

Lys Gln Ala Gln Lys Met Glu Lys Leu Lys Ala Glu Leu Ala Lys Ala
705                 710                 715                 720

Val Glu Asp Ala Lys Pro Lys Gly Gln Lys Ser Gly Glu Ala Glu Gly
                725                 730                 735

Ile Ile Glu Pro Lys Pro Asn Asn Leu Ser Thr Leu Leu Gly Gly Tyr
            740                 745                 750

Gly Tyr Phe Asp Met Gly Asp Glu Asp Ile Lys Ser Arg Gly Gln Ser
        755                 760                 765
```

```
Pro Glu Pro Pro Phe Leu His Arg Ala Gly Ser Gly Phe Ser Ile Arg
    770                 775                 780

Glu Ile Gly Lys Lys Leu Arg Leu Ser Glu Tyr
785                 790                 795


<210> 11
<211> 790
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
      polypeptide

<400> 11
Met Ser Ala Leu Arg Ser Phe Gln Arg Ser Ser Asn Val Ala Lys Ser
1               5                   10                  15

Thr Leu Lys Asn Ser Val Arg Thr Tyr Ser Thr Tyr Gln Phe Gln Glu
            20                  25                  30

Thr Leu Glu Lys Leu Pro Ile Pro Asp Leu Asn Gln Thr Cys Ala Asn
        35                  40                  45

Tyr Leu Asn Val Leu Arg Pro Leu Gln Thr Glu Gln Glu His Ile Lys
    50                  55                  60

Thr Lys Thr Ala Val Glu Asn Phe Leu Lys Asn Gly Thr Gly Gln Tyr
65                  70                  75                  80

Leu Asp Ala Ala Leu Arg Glu Tyr Ala Gln Thr Arg Pro Ser Tyr Ile
                85                  90                  95

Glu Gln Phe Trp Tyr Asp Ala Tyr Leu Asn Tyr Asp Ser Pro Val Val
            100                 105                 110

Leu Asn Leu Asn Pro Phe Phe Leu Leu Glu Asp Asp Pro Phe Thr Asn
        115                 120                 125

Glu Ser Ser Ser Val Asn Pro Gln Val Lys Arg Ala Thr Ser Leu Val
    130                 135                 140

Met Ser Ser Leu Lys Phe Ile Gln Ala Leu Lys Asn Glu Thr Leu Ser
145                 150                 155                 160

Val Asp Thr Leu Lys Gly Gly Lys Pro Leu Cys Met Tyr Gln Tyr Thr
                165                 170                 175
```


LCY Biotechnology Holding, Inc.
Ex. 1002
Page 64 of 3077

```
Lys Leu Phe Gly Ala Ser Arg Ile Pro Ser Glu Asp Gly Cys Val Met
            180                 185                 190

Gln Ser Asp Pro Ala Ser Asn His Ile Val Val Met Ser Lys Ser Gln
        195                 200                 205

Leu Tyr Trp Phe Asp Val Leu Asp Ser Lys Asn Asn Leu Ile Leu Ser
    210                 215                 220

Glu Ala Glu Leu Asn Val Asn Phe Gln Ser Ile Ile His Asp Ser Leu
225                 230                 235                 240

Arg Thr Pro Ser Asp Glu Ile Ala Lys Ser Ser Phe Gly Val Leu Thr
                245                 250                 255

Thr Glu Asn Arg Arg Ile Trp Ala Asn Val Arg His Asn Leu Met Ser
            260                 265                 270

Thr Thr Asn Asn Lys Val Asn His Glu Val Leu Ser Ile Ile Asp Ser
        275                 280                 285

Ala Leu Phe Val Leu Cys Phe Asp Asp Ile Val Ile Asn Asp Leu Ser
    290                 295                 300

Glu Leu Ser Lys Asn Met Leu Cys Gly Leu Ser Ile Leu Asp Asn Gly
305                 310                 315                 320

Ile Gln Val Gly Thr Cys Thr Asn Arg Trp Tyr Asp Lys Leu Gln Ile
                325                 330                 335

Ile Val Thr Lys Asn Ala Lys Ala Gly Ile Asn Phe Glu His Thr Gly
            340                 345                 350

Val Asp Gly His Thr Val Leu Arg Phe Val Ser Asp Ile Tyr Thr Asp
        355                 360                 365

Ser Ile Leu Ser Phe Ala Asn Ser Ile Asn Ser Asn Ala Pro Ser Leu
    370                 375                 380

Trp Asn Thr Thr Thr Asn Tyr Lys Glu Leu Asp Gly Glu Asp Leu Ile
385                 390                 395                 400

Thr Val Pro Arg Lys Leu Glu Trp Glu Leu Thr Pro Asp Leu Ser Leu
                405                 410                 415

Ala Leu Arg Phe Gly Glu Thr Arg Leu Ser Asp Leu Ile Asn Gln Asn
            420                 425                 430
```


LCY Biotechnology Holding, Inc.
Ex. 1002
Page 65 of 3077

```
Glu Phe Arg His Leu Glu Phe Lys Asn Tyr Gly Ser Thr Gln Ile Lys
        435                 440                 445

Lys Met Lys Phe Ser Pro Asp Ala Phe Val Gln Met Ala Phe Gln Ala
    450                 455                 460

Thr Tyr Tyr Ala Leu Tyr Gly Lys Val Glu Cys Thr Tyr Glu Pro Ala
465                 470                 475                 480

Met Thr Lys Gln Phe Tyr His Gly Arg Thr Glu Ala Ile Arg Thr Val
                485                 490                 495

Ser Gln Glu Ser Asn Leu Phe Val Arg Lys Phe Phe Asp Ser Thr Val
            500                 505                 510

Ser Ile Gln Lys Lys Leu Glu Tyr Leu Thr Gln Ala Cys Thr Lys His
        515                 520                 525

Ser Gln Gln Thr Arg Met Ser Ser Ala Gly Gln Gly Val Asp Arg His
    530                 535                 540

Leu Tyr Ala Leu Phe Cys Ile Trp Lys Arg Tyr Leu His Asp Ala Glu
545                 550                 555                 560

Asp Asp Asp Asp Asp Ala Ser Lys Asp Gln Pro Ser Ser Thr Glu Leu
                565                 570                 575

Ser Arg Asp Asp Thr Ile Ile Ala Gln Thr Asn Gly Asp Ser Thr Asp
            580                 585                 590

Asp Asp Asp Ser Thr Thr Val Val Gly His Gly Ala Asn His His His
        595                 600                 605

Asn Leu Lys Ser Ala Asp Leu Leu Lys Thr Ile Pro Glu Ile Phe Ala
    610                 615                 620

Asp Asn Gly Trp Asp Lys Leu Asn Asn Thr Ile Ile Ser Thr Ser Asn
625                 630                 635                 640

Cys Gly Asn Pro Ser Leu Arg Leu Phe Gly Phe Gly Pro Val Ser Ala
                645                 650                 655

Asn Gly Phe Gly Ile Gly Tyr Ile Leu Lys Asp Asp Ser Ile Ser Ile
            660                 665                 670

Cys Ala Ser Ser Lys His Arg Gln Thr Gln Arg Phe Leu Val Thr Leu
        675                 680                 685
```


LCY Biotechnology Holding, Inc.
Ex. 1002
Page 66 of 3077

```
Asn Ser Tyr Leu Leu Glu Ile Gln Asn Ile Trp Lys Gln Ala Gln Lys
    690                 695                 700

Met Glu Lys Leu Lys Ala Glu Leu Ala Lys Ala Val Glu Asp Ala Lys
705                 710                 715                 720

Pro Lys Gly Gln Lys Ser Gly Glu Ala Glu Gly Ile Ile Glu Pro Lys
                725                 730                 735

Pro Asn Asn Leu Ser Thr Leu Leu Gly Gly Tyr Gly Tyr Phe Asp Met
            740                 745                 750

Gly Asp Glu Asp Ile Lys Ser Arg Gly Gln Ser Pro Glu Pro Pro Phe
        755                 760                 765

Leu His Arg Ala Gly Ser Gly Phe Ser Ile Arg Glu Ile Gly Lys Lys
    770                 775                 780

Leu Arg Leu Ser Glu Tyr
785                 790


<210> 12
<211> 798
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
      polypeptide

<400> 12
Met Phe Asn Phe Lys Leu Ser Gln Gln Val Leu Lys Asn Ser Thr Lys
1               5                   10                  15

Ser Ile Met Pro Ile Leu Lys Lys Pro Phe Ser Thr Ser His Ala Lys
            20                  25                  30

Gly Ser Thr Tyr Gln Phe Gln Glu Thr Leu Glu Lys Leu Pro Ile Pro
        35                  40                  45

Asp Leu Asn Gln Thr Cys Ala Asn Tyr Leu Asn Val Leu Arg Pro Leu
    50                  55                  60

Gln Thr Glu Gln Glu His Ile Lys Thr Lys Thr Ala Val Glu Asn Phe
65                  70                  75                  80

Leu Lys Asn Gly Thr Gly Gln Tyr Leu Asp Ala Ala Leu Arg Glu Tyr
                85                  90                  95
```


LCY Biotechnology Holding, Inc.
Ex. 1002
Page 67 of 3077

```
Ala Gln Thr Arg Pro Ser Tyr Ile Glu Gln Phe Trp Tyr Asp Ala Tyr
            100                 105                 110

Leu Asn Tyr Asp Ser Pro Val Val Leu Asn Leu Asn Pro Phe Phe Leu
        115                 120                 125

Leu Glu Asp Asp Pro Phe Thr Asn Glu Ser Ser Ser Val Asn Pro Gln
    130                 135                 140

Val Lys Arg Ala Thr Ser Leu Val Met Ser Ser Leu Lys Phe Ile Gln
145                 150                 155                 160

Ala Leu Lys Asn Glu Thr Leu Ser Val Asp Thr Leu Lys Gly Gly Lys
                165                 170                 175

Pro Leu Cys Met Tyr Gln Tyr Thr Lys Leu Phe Gly Ala Ser Arg Ile
            180                 185                 190

Pro Ser Glu Asp Gly Cys Val Met Gln Ser Asp Pro Ala Ser Asn His
        195                 200                 205

Ile Val Val Met Ser Lys Ser Gln Leu Tyr Trp Phe Asp Val Leu Asp
    210                 215                 220

Ser Lys Asn Asn Leu Ile Leu Ser Glu Ala Glu Leu Asn Val Asn Phe
225                 230                 235                 240

Gln Ser Ile Ile His Asp Ser Leu Arg Thr Pro Ser Asp Glu Ile Ala
                245                 250                 255

Lys Ser Ser Phe Gly Val Leu Thr Thr Glu Asn Arg Arg Ile Trp Ala
            260                 265                 270

Asn Val Arg His Asn Leu Met Ser Thr Thr Asn Asn Lys Val Asn His
        275                 280                 285

Glu Val Leu Ser Ile Ile Asp Ser Ala Leu Phe Val Leu Cys Phe Asp
    290                 295                 300

Asp Ile Val Ile Asn Asp Leu Ser Glu Leu Ser Lys Asn Met Leu Cys
305                 310                 315                 320

Gly Leu Ser Ile Leu Asp Asn Gly Ile Gln Val Gly Thr Cys Thr Asn
                325                 330                 335

Arg Trp Tyr Asp Lys Leu Gln Ile Ile Val Thr Lys Asn Ala Lys Ala
            340                 345                 350
```


LCY Biotechnology Holding, Inc.
Ex. 1002
Page 68 of 3077

```
Gly Ile Asn Phe Glu His Thr Gly Val Asp Gly His Thr Val Leu Arg
        355                 360                 365

Phe Val Ser Asp Ile Tyr Thr Asp Ser Ile Leu Ser Phe Ala Asn Ser
    370                 375                 380

Ile Asn Ser Asn Ala Pro Ser Leu Trp Asn Thr Thr Thr Asn Tyr Lys
385                 390                 395                 400

Glu Leu Asp Gly Glu Asp Leu Ile Thr Val Pro Arg Lys Leu Glu Trp
                405                 410                 415

Glu Leu Thr Pro Asp Leu Ser Leu Ala Leu Arg Phe Gly Glu Thr Arg
            420                 425                 430

Leu Ser Asp Leu Ile Asn Gln Asn Glu Phe Arg His Leu Glu Phe Lys
        435                 440                 445

Asn Tyr Gly Ser Thr Gln Ile Lys Lys Met Lys Phe Ser Pro Asp Ala
    450                 455                 460

Phe Val Gln Met Ala Phe Gln Ala Thr Tyr Tyr Ala Leu Tyr Gly Lys
465                 470                 475                 480

Val Glu Cys Thr Tyr Glu Pro Ala Met Thr Lys Gln Phe Tyr His Gly
                485                 490                 495

Arg Thr Glu Ala Ile Arg Thr Val Ser Gln Glu Ser Asn Leu Phe Val
            500                 505                 510

Arg Lys Phe Phe Asp Ser Thr Val Ser Ile Gln Lys Lys Leu Glu Tyr
        515                 520                 525

Leu Thr Gln Ala Cys Thr Lys His Ser Gln Gln Thr Arg Met Ser Ser
    530                 535                 540

Ala Gly Gln Gly Val Asp Arg His Leu Tyr Ala Leu Phe Cys Ile Trp
545                 550                 555                 560

Lys Arg Tyr Leu His Asp Ala Glu Asp Asp Asp Asp Asp Ala Ser Lys
                565                 570                 575

Asp Gln Pro Ser Ser Thr Glu Leu Ser Arg Asp Asp Thr Ile Ile Ala
            580                 585                 590

Gln Thr Asn Gly Asp Ser Thr Asp Asp Asp Asp Ser Thr Thr Val Val
```

595 600 605

Gly His Gly Ala Asn His His His Asn Leu Lys Ser Ala Asp Leu Leu
610 615 620

Lys Thr Ile Pro Glu Ile Phe Ala Asp Asn Gly Trp Asp Lys Leu Asn
625 630 635 640

Asn Thr Ile Ile Ser Thr Ser Asn Cys Gly Asn Pro Ser Leu Arg Leu
645 650 655

Phe Gly Phe Gly Pro Val Ser Ala Asn Gly Phe Gly Ile Gly Tyr Ile
660 665 670

Leu Lys Asp Asp Ser Ile Ser Ile Cys Ala Ser Ser Lys His Arg Gln
675 680 685

Thr Gln Arg Phe Leu Val Thr Leu Asn Ser Tyr Leu Leu Glu Ile Gln
690 695 700

Asn Ile Trp Lys Gln Ala Gln Lys Met Glu Lys Leu Lys Ala Glu Leu
705 710 715 720

Ala Lys Ala Val Glu Asp Ala Lys Pro Lys Gly Gln Lys Ser Gly Glu
725 730 735

Ala Glu Gly Ile Ile Glu Pro Lys Pro Asn Asn Leu Ser Thr Leu Leu
740 745 750

Gly Gly Tyr Gly Tyr Phe Asp Met Gly Asp Glu Asp Ile Lys Ser Arg
755 760 765

Gly Gln Ser Pro Glu Pro Pro Phe Leu His Arg Ala Gly Ser Gly Phe
770 775 780

Ser Ile Arg Glu Ile Gly Lys Lys Leu Arg Leu Ser Glu Tyr
785 790 795

<210> 13
<211> 769
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 13
Met Ser Thr Tyr Gln Phe Gln Glu Thr Leu Glu Lys Leu Pro Ile Pro
1 5 10 15

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 70 of 3077

```
Asp Leu Asn Gln Thr Cys Ala Asn Tyr Leu Asn Val Leu Arg Pro Leu
            20                  25                  30

Gln Thr Glu Gln Glu His Ile Lys Thr Lys Thr Ala Val Glu Asn Phe
        35                  40                  45

Leu Lys Asn Gly Thr Gly Gln Tyr Leu Asp Ala Ala Leu Arg Glu Tyr
    50                  55                  60

Ala Gln Thr Arg Pro Ser Tyr Ile Glu Gln Phe Trp Tyr Asp Ala Tyr
65                  70                  75                  80

Leu Asn Tyr Asp Ser Pro Val Val Leu Asn Leu Asn Pro Phe Phe Leu
                85                  90                  95

Leu Glu Asp Asp Pro Phe Thr Asn Glu Ser Ser Ser Val Asn Pro Gln
            100                 105                 110

Val Lys Arg Ala Thr Ser Leu Val Met Ser Ser Leu Lys Phe Ile Gln
        115                 120                 125

Ala Leu Lys Asn Glu Thr Leu Ser Val Asp Thr Leu Lys Gly Gly Lys
    130                 135                 140

Pro Leu Cys Met Tyr Gln Tyr Thr Lys Leu Phe Gly Ala Ser Arg Ile
145                 150                 155                 160

Pro Ser Glu Asp Gly Cys Val Met Gln Ser Asp Pro Ala Ser Asn His
                165                 170                 175

Ile Val Val Met Ser Lys Ser Gln Leu Tyr Trp Phe Asp Val Leu Asp
            180                 185                 190

Ser Lys Asn Asn Leu Ile Leu Ser Glu Ala Glu Leu Asn Val Asn Phe
        195                 200                 205

Gln Ser Ile Ile His Asp Ser Leu Arg Thr Pro Ser Asp Glu Ile Ala
    210                 215                 220

Lys Ser Ser Phe Gly Val Leu Thr Thr Glu Asn Arg Arg Ile Trp Ala
225                 230                 235                 240

Asn Val Arg His Asn Leu Met Ser Thr Thr Asn Asn Lys Val Asn His
                245                 250                 255

Glu Val Leu Ser Ile Ile Asp Ser Ala Leu Phe Val Leu Cys Phe Asp
```

260 265 270

Asp Ile Val Ile Asn Asp Leu Ser Glu Leu Ser Lys Asn Met Leu Cys
275 280 285

Gly Leu Ser Ile Leu Asp Asn Gly Ile Gln Val Gly Thr Cys Thr Asn
290 295 300

Arg Trp Tyr Asp Lys Leu Gln Ile Ile Val Thr Lys Asn Ala Lys Ala
305 310 315 320

Gly Ile Asn Phe Glu His Thr Gly Val Asp Gly His Thr Val Leu Arg
325 330 335

Phe Val Ser Asp Ile Tyr Thr Asp Ser Ile Leu Ser Phe Ala Asn Ser
340 345 350

Ile Asn Ser Asn Ala Pro Ser Leu Trp Asn Thr Thr Thr Asn Tyr Lys
355 360 365

Glu Leu Asp Gly Glu Asp Leu Ile Thr Val Pro Arg Lys Leu Glu Trp
370 375 380

Glu Leu Thr Pro Asp Leu Ser Leu Ala Leu Arg Phe Gly Glu Thr Arg
385 390 395 400

Leu Ser Asp Leu Ile Asn Gln Asn Glu Phe Arg His Leu Glu Phe Lys
405 410 415

Asn Tyr Gly Ser Thr Gln Ile Lys Lys Met Lys Phe Ser Pro Asp Ala
420 425 430

Phe Val Gln Met Ala Phe Gln Ala Thr Tyr Tyr Ala Leu Tyr Gly Lys
435 440 445

Val Glu Cys Thr Tyr Glu Pro Ala Met Thr Lys Gln Phe Tyr His Gly
450 455 460

Arg Thr Glu Ala Ile Arg Thr Val Ser Gln Glu Ser Asn Leu Phe Val
465 470 475 480

Arg Lys Phe Phe Asp Ser Thr Val Ser Ile Gln Lys Lys Leu Glu Tyr
485 490 495

Leu Thr Gln Ala Cys Thr Lys His Ser Gln Gln Thr Arg Met Ser Ser
500 505 510

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 72 of 3077

```
Ala Gly Gln Gly Val Asp Arg His Leu Tyr Ala Leu Phe Cys Ile Trp
        515                 520                 525

Lys Arg Tyr Leu His Asp Ala Glu Asp Asp Asp Asp Asp Ala Ser Lys
    530                 535                 540

Asp Gln Pro Ser Ser Thr Glu Leu Ser Arg Asp Asp Thr Ile Ile Ala
545                 550                 555                 560

Gln Thr Asn Gly Asp Ser Thr Asp Asp Asp Asp Ser Thr Thr Val Val
                565                 570                 575

Gly His Gly Ala Asn His His His Asn Leu Lys Ser Ala Asp Leu Leu
            580                 585                 590

Lys Thr Ile Pro Glu Ile Phe Ala Asp Asn Gly Trp Asp Lys Leu Asn
        595                 600                 605

Asn Thr Ile Ile Ser Thr Ser Asn Cys Gly Asn Pro Ser Leu Arg Leu
    610                 615                 620

Phe Gly Phe Gly Pro Val Ser Ala Asn Gly Phe Gly Ile Gly Tyr Ile
625                 630                 635                 640

Leu Lys Asp Asp Ser Ile Ser Ile Cys Ala Ser Ser Lys His Arg Gln
                645                 650                 655

Thr Gln Arg Phe Leu Val Thr Leu Asn Ser Tyr Leu Leu Glu Ile Gln
            660                 665                 670

Asn Ile Trp Lys Gln Ala Gln Lys Met Glu Lys Leu Lys Ala Glu Leu
        675                 680                 685

Ala Lys Ala Val Glu Asp Ala Lys Pro Lys Gly Gln Lys Ser Gly Glu
    690                 695                 700

Ala Glu Gly Ile Ile Glu Pro Lys Pro Asn Asn Leu Ser Thr Leu Leu
705                 710                 715                 720

Gly Gly Tyr Gly Tyr Phe Asp Met Gly Asp Glu Asp Ile Lys Ser Arg
                725                 730                 735

Gly Gln Ser Pro Glu Pro Pro Phe Leu His Arg Ala Gly Ser Gly Phe
            740                 745                 750

Ser Ile Arg Glu Ile Gly Lys Lys Leu Arg Leu Ser Glu Tyr Pro Lys
        755                 760                 765
```

Leu

<210> 14
<211> 284
<212> PRT
<213> Candida viswanathii

<400> 14
Met Asp Asp Val Asp Ser Ala Leu Ala Asp Asn Val Lys Ser Phe Ala
1 5 10 15

Ala Gly Gly Phe Gly Gly Ile Cys Ala Val Leu Thr Gly His Pro Phe
20 25 30

Asp Leu Val Lys Val Arg Leu Gln Thr Gly Leu Tyr Lys Ser Ser Val
35 40 45

Gln Cys Val Lys Glu Thr Ile Ala Lys Asp Gly Leu Phe Gly Leu Tyr
50 55 60

Arg Gly Val Leu Pro Pro Leu Leu Gly Val Thr Pro Met Phe Ala Val
65 70 75 80

Ser Phe Trp Gly Tyr Asp Val Gly Lys Lys Leu Val Ser Ser Phe Thr
85 90 95

Gly Lys Ser Val Asp Lys Phe Glu Ile Lys Asp Ile Ser Thr Ala Gly
100 105 110

Phe Ile Ser Ala Ile Pro Thr Thr Leu Val Ala Ala Pro Phe Glu Arg
115 120 125

Val Lys Val Met Met Gln Ile Gln Glu Gly Ala Lys Ser Lys Ser Met
130 135 140

Gly Ala Val Val Ala Glu Met Tyr Arg Thr Gly Gly Ile Arg Ser Ile
145 150 155 160

Phe Lys Gly Thr Val Ala Thr Leu Ala Arg Asp Gly Pro Gly Ser Ala
165 170 175

Leu Tyr Phe Ala Thr Tyr Glu Trp Val Lys Lys Glu Leu Thr Ala Pro
180 185 190

Gly Glu Asp Leu Ser Leu Phe Ala Ile Thr Thr Ala Gly Gly Phe Ala
195 200 205

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 74 of 3077

```
Gly Ile Ala Met Trp Leu Gly Val Phe Pro Ile Asp Thr Ile Lys Ser
    210                 215                 220

Thr Gln Gln Ser Ser Asn Val Lys Val Ser Ile Val Gln Ala Thr Lys
225                 230                 235                 240

Asn Ile Tyr Ala Lys Gly Gly Ile Lys Ala Phe Phe Pro Gly Val Gly
                245                 250                 255

Pro Ala Leu Ala Arg Ala Phe Pro Ala Asn Ala Ala Thr Phe Leu Gly
            260                 265                 270

Val Glu Leu Ala Arg Lys Phe Leu Asp Lys Val Ile
        275                 280


<210> 15
<211> 519
<212> PRT
<213> Candida viswanathii

<400> 15
Met Ser Ala Ile Leu Lys Gln Arg Val Arg Tyr Ala Pro Tyr Leu Lys
1               5                   10                  15

Lys Leu Arg Thr Ala Glu Gln Cys Val Asp Leu Phe Lys Asn Gly Gln
            20                  25                  30

Tyr Leu Gly Trp Ser Gly Phe Thr Gly Val Gly Ala Pro Lys Ala Val
        35                  40                  45

Pro Ala Ala Leu Val Asp His Val Glu Lys Asn Lys Leu Gln Gly Gln
    50                  55                  60

Met Gly Phe His Leu Phe Val Gly Ala Ser Ala Gly Pro Glu Glu Ser
65                  70                  75                  80

Arg Trp Ala Glu Asn Asn Met Leu Leu Ser Arg Ser Pro His Gln Val
                85                  90                  95

Gly Lys Pro Ile Ala Ala Ala Ile Asn Asp Gly Arg Thr Gln Phe Phe
            100                 105                 110

Asp Lys His Leu Ser Met Phe Pro Gln Asp Leu Thr Tyr Gly Phe Tyr
        115                 120                 125

Thr Lys Asn Lys Pro Asn Gly Ser Asn Leu Asp Tyr Thr Ile Ile Glu
    130                 135                 140

Ala Thr Ala Ile Thr Glu Asp Gly Ala Ile Ile Pro Gly Pro Ala Val
```

145 150 155 160

Gly Ala Ser Pro Glu Met Leu Ser Val Ser Asp Asn Ile Ile Ile Glu
165 170 175

Val Asn Thr Lys Thr Pro Ser Phe Glu Gly Ile His Asp Ile Asp Leu
180 185 190

Pro Val Asn Pro Pro Phe Arg Gln Pro Tyr Pro His Thr Thr Ala Asp
195 200 205

Tyr Arg Ile Gly Arg Thr Ala Ile Pro Val Asp Pro Glu Lys Val Ile
210 215 220

Ala Ile Val Glu Thr Thr Glu Gly Asp Lys Val Pro Pro Asn Thr Pro
225 230 235 240

Ser Asp Ala Gln Ser Gln Ala Ile Ala Gly His Leu Ile Glu Phe Leu
245 250 255

Glu Asn Glu Val Lys Gln Gly Arg Leu Pro Glu Asn Leu His Pro Leu
260 265 270

Gln Ser Gly Ile Gly Asn Ile Ala Asn Ala Val Val Glu Gly Leu Ala
275 280 285

Gln Ser Asn Phe Lys Asn Leu Thr Val Trp Thr Glu Val Leu Gln Asp
290 295 300

Ser Phe Leu Asp Phe Phe Glu Ser Gly Ser Leu Asp Phe Ala Thr Ala
305 310 315 320

Thr Ser Ile Arg Leu Thr Glu Glu Gly Phe Lys Lys Phe Tyr Glu Asn
325 330 335

Trp Asp Val Tyr Ser Lys Lys Leu Cys Leu Arg Ser Gln Val Val Ser
340 345 350

Asn Ser Pro Glu Ile Ile Arg Arg Leu Gly Val Leu Ala Phe Asn Thr
355 360 365

Pro Val Glu Val Asp Ile Tyr Ala His Ala Asn Ser Thr Asn Val Met
370 375 380

Gly Ser Arg Met Leu Asn Gly Leu Gly Gly Ser Gly Asp Phe Leu Arg
385 390 395 400

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 76 of 3077

```
Asn Ala Lys Leu Ser Val Met His Thr Pro Ser Ala Arg Pro Thr Lys
                405                 410                 415

Val Asp Pro Thr Gly Val Ser Cys Ile Val Pro Met Ala Thr His Val
            420                 425                 430

Asp Gln Thr Glu His Asp Leu Asp Val Val Val Thr Glu Gln Gly Leu
        435                 440                 445

Ala Asp Leu Arg Gly Leu Ser Pro Lys Glu Arg Ala Gln Val Ile Ile
    450                 455                 460

Asn Asn Cys Ala His Pro Asp Tyr Gln Ala Gln Leu Gln Asp Tyr Phe
465                 470                 475                 480

Asp Arg Ala Val Phe Tyr Ala Thr Lys Lys Lys Met Leu His Glu Pro
                485                 490                 495

His Met Leu Asn Glu Ala Phe Ala Met His Leu Asn Leu Gln Glu Asn
            500                 505                 510

Gly Thr Met Lys Leu Asn Lys
        515


<210> 16
<211> 515
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
      polypeptide

<400> 16
Met Ala Pro Tyr Leu Lys Lys Leu Arg Thr Ala Glu Gln Cys Val Asp
1               5                   10                  15

Leu Phe Lys Asn Gly Gln Tyr Leu Gly Trp Ser Gly Phe Thr Gly Val
            20                  25                  30

Gly Ala Pro Lys Ala Val Pro Ala Ala Leu Val Asp His Val Glu Lys
        35                  40                  45

Asn Lys Leu Gln Gly Gln Met Gly Phe His Leu Phe Val Gly Ala Ser
    50                  55                  60

Ala Gly Pro Glu Glu Ser Arg Trp Ala Glu Asn Asn Met Leu Leu Ser
65                  70                  75                  80

Arg Ser Pro His Gln Val Gly Lys Pro Ile Ala Ala Ala Ile Asn Asp
```

85 90 95

Gly Arg Thr Gln Phe Phe Asp Lys His Leu Ser Met Phe Pro Gln Asp
100 105 110

Leu Thr Tyr Gly Phe Tyr Thr Lys Asn Lys Pro Asn Gly Ser Asn Leu
115 120 125

Asp Tyr Thr Ile Ile Glu Ala Thr Ala Ile Thr Glu Asp Gly Ala Ile
130 135 140

Ile Pro Gly Pro Ala Val Gly Ala Ser Pro Glu Met Leu Ser Val Ser
145 150 155 160

Asp Asn Ile Ile Ile Glu Val Asn Thr Lys Thr Pro Ser Phe Glu Gly
165 170 175

Ile His Asp Ile Asp Leu Pro Val Asn Pro Pro Phe Arg Gln Pro Tyr
180 185 190

Pro His Thr Thr Ala Asp Tyr Arg Ile Gly Arg Thr Ala Ile Pro Val
195 200 205

Asp Pro Glu Lys Val Ile Ala Ile Val Glu Thr Thr Glu Gly Asp Lys
210 215 220

Val Pro Pro Asn Thr Pro Ser Asp Ala Gln Ser Gln Ala Ile Ala Gly
225 230 235 240

His Leu Ile Glu Phe Leu Glu Asn Glu Val Lys Gln Gly Arg Leu Pro
245 250 255

Glu Asn Leu His Pro Leu Gln Ser Gly Ile Gly Asn Ile Ala Asn Ala
260 265 270

Val Val Glu Gly Leu Ala Gln Ser Asn Phe Lys Asn Leu Thr Val Trp
275 280 285

Thr Glu Val Leu Gln Asp Ser Phe Leu Asp Phe Phe Glu Ser Gly Ser
290 295 300

Leu Asp Phe Ala Thr Ala Thr Ser Ile Arg Leu Thr Glu Glu Gly Phe
305 310 315 320

Lys Lys Phe Tyr Glu Asn Trp Asp Val Tyr Ser Lys Lys Leu Cys Leu
325 330 335

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 78 of 3077

```
Arg Ser Gln Val Val Ser Asn Ser Pro Glu Ile Ile Arg Arg Leu Gly
            340                 345                 350

Val Leu Ala Phe Asn Thr Pro Val Glu Val Asp Ile Tyr Ala His Ala
        355                 360                 365

Asn Ser Thr Asn Val Met Gly Ser Arg Met Leu Asn Gly Leu Gly Gly
    370                 375                 380

Ser Gly Asp Phe Leu Arg Asn Ala Lys Leu Ser Val Met His Thr Pro
385                 390                 395                 400

Ser Ala Arg Pro Thr Lys Val Asp Pro Thr Gly Val Ser Cys Ile Val
                405                 410                 415

Pro Met Ala Thr His Val Asp Gln Thr Glu His Asp Leu Asp Val Val
            420                 425                 430

Val Thr Glu Gln Gly Leu Ala Asp Leu Arg Gly Leu Ser Pro Lys Glu
        435                 440                 445

Arg Ala Gln Val Ile Ile Asn Asn Cys Ala His Pro Asp Tyr Gln Ala
    450                 455                 460

Gln Leu Gln Asp Tyr Phe Asp Arg Ala Val Phe Tyr Ala Thr Lys Lys
465                 470                 475                 480

Lys Met Leu His Glu Pro His Met Leu Asn Glu Ala Phe Ala Met His
                485                 490                 495

Leu Asn Leu Gln Glu Asn Gly Thr Met Lys Leu Asn Lys Gly Arg Arg
            500                 505                 510

Ala Lys Leu
        515


<210> 17
<211> 2274
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
      polypeptide

<400> 17
Met Arg Cys Gln Val Ser Gln Pro Ser Arg Phe Thr Asn Leu Leu Val
1               5                   10                  15

His Arg Leu Pro Arg Thr Leu Leu Asn Tyr Pro Val Val Asn Thr Leu
```

20 25 30

Phe Ile Pro Arg Arg His Tyr Ser Leu Asn Phe Ser Phe Lys Asn Leu
35 40 45

Leu Lys Lys Met Thr Asp Leu Ser Pro Ser Pro Thr Asp Ser Leu Asn
50 55 60

Tyr Thr Gln Leu His Ser Ser Leu Pro Ser His Phe Leu Gly Gly Asn
65 70 75 80

Ser Val Leu Thr Ala Glu Pro Ser Ala Val Thr Asp Phe Val Lys Thr
85 90 95

His Gln Gly His Thr Val Ile Thr Lys Val Leu Ile Ala Asn Asn Gly
100 105 110

Ile Gly Ala Val Lys Glu Ile Arg Ser Val Arg Lys Trp Ala Tyr Glu
115 120 125

Thr Phe Gly Asp Glu Arg Ala Ile Gln Phe Val Ala Met Ala Thr Pro
130 135 140

Glu Asp Met Glu Ala Asn Ala Glu Tyr Ile Arg Met Ala Asp Gln Phe
145 150 155 160

Val Glu Val Pro Gly Gly Thr Asn Asn Asn Asn Tyr Ala Asn Val Asp
165 170 175

Leu Ile Val Glu Ile Ala Glu Arg Thr Asp Val His Ala Val Trp Ala
180 185 190

Gly Trp Gly His Ala Ser Glu Asn Pro Leu Leu Pro Glu Arg Leu Ala
195 200 205

Ala Ser Pro Lys Lys Ile Val Phe Ile Gly Pro Pro Gly Ser Ala Met
210 215 220

Arg Ser Leu Gly Asp Lys Ile Ser Ser Thr Ile Val Ala Gln His Ala
225 230 235 240

Lys Val Pro Cys Ile Pro Trp Ser Gly Thr Gly Val Glu Glu Val His
245 250 255

Val Asp Pro Glu Thr Lys Leu Val Ser Val Asp Asp His Val Tyr Ala
260 265 270

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 80 of 3077

```
Lys Gly Cys Cys Thr Ser Pro Glu Asp Gly Leu Glu Lys Ala Lys Arg
        275                 280                 285

Ile Gly Phe Pro Val Met Val Lys Ala Ser Glu Gly Gly Gly Gly Lys
    290                 295                 300

Gly Ile Arg Lys Val Asp His Glu Lys Asp Phe Ile Ser Leu Tyr Asn
305                 310                 315                 320

Gln Ala Ala Asn Glu Ile Pro Gly Ser Pro Ile Phe Ile Met Lys Leu
                325                 330                 335

Ala Gly Asp Ala Arg His Leu Glu Val Gln Leu Phe Ala Asp Gln Tyr
            340                 345                 350

Gly Thr Asn Ile Ser Leu Phe Gly Arg Asp Cys Ser Val Gln Arg Arg
        355                 360                 365

His Gln Lys Ile Ile Glu Glu Ala Pro Val Thr Ile Ala Asn Lys Asp
    370                 375                 380

Thr Phe Val Glu Met Glu Lys Ala Ala Val Arg Leu Gly Lys Leu Val
385                 390                 395                 400

Gly Tyr Val Ser Ala Gly Thr Val Glu Tyr Leu Tyr Ser Tyr Ala Glu
                405                 410                 415

Asp Lys Phe Tyr Phe Leu Glu Leu Asn Pro Arg Leu Gln Val Glu His
            420                 425                 430

Pro Thr Thr Glu Met Val Ser Gly Val Asn Leu Pro Ala Ala Gln Leu
        435                 440                 445

Gln Ile Ala Met Gly Leu Pro Met His Arg Ile Arg Asp Ile Arg Leu
    450                 455                 460

Leu Tyr Gly Val Asp Pro His Ser Ala Thr Glu Ile Asp Phe Glu Phe
465                 470                 475                 480

Lys Ser Pro Asn Ser Leu Ile Thr Gln Arg Lys Pro Ala Pro Lys Gly
                485                 490                 495

His Cys Thr Ala Cys Arg Ile Thr Ser Glu Asp Pro Gly Glu Gly Phe
            500                 505                 510

Lys Pro Ser Gly Gly Thr Leu His Glu Leu Asn Phe Arg Ser Ser Ser
        515                 520                 525
```

```
Asn Val Trp Gly Tyr Phe Ser Val Ala Asn Gln Ser Ser Ile His Ser
    530                 535                 540

Phe Ala Asp Ser Gln Phe Gly His Ile Phe Ala Phe Gly Glu Asn Arg
545                 550                 555                 560

Gln Ala Ser Arg Lys His Met Ile Val Ala Leu Lys Glu Leu Ser Ile
                565                 570                 575

Arg Gly Asp Phe Arg Thr Thr Val Glu Tyr Leu Ile Lys Leu Leu Glu
            580                 585                 590

Thr Pro Asp Phe Ala Asp Asn Thr Ile Thr Thr Gly Trp Leu Asp Glu
        595                 600                 605

Leu Ile Thr Lys Lys Leu Thr Ala Glu Arg Pro Asp Pro Ile Val Ala
    610                 615                 620

Val Val Cys Gly Ala Val Thr Lys Ala His Ile Gln Ala Glu Glu Asp
625                 630                 635                 640

Lys Lys Glu Tyr Ile Glu Ser Leu Glu Lys Gly Gln Val Pro Asn Lys
                645                 650                 655

Ser Leu Leu Lys Thr Ile Phe Pro Val Glu Phe Ile Tyr Glu Gly Glu
            660                 665                 670

Arg Tyr Lys Phe Thr Ala Thr Lys Ser Ser Glu Asp Lys Tyr Thr Leu
        675                 680                 685

Phe Leu Asn Gly Ser Arg Cys Val Ile Gly Ala Arg Ser Leu Ser Asp
    690                 695                 700

Gly Gly Leu Leu Cys Ala Leu Asp Gly Lys Ser His Ser Val Tyr Trp
705                 710                 715                 720

Lys Glu Glu Ala Ala Ala Thr Arg Leu Ser Val Asp Gly Lys Thr Cys
                725                 730                 735

Leu Leu Glu Val Glu Asn Asp Pro Thr Gln Leu Arg Thr Pro Ser Pro
            740                 745                 750

Gly Lys Leu Val Lys Tyr Leu Val Glu Ser Gly Glu His Val Asp Ala
        755                 760                 765

Gly Gln Ser Tyr Ala Glu Val Glu Val Met Lys Met Cys Met Pro Leu
    770                 775                 780
```

```
Ile Ala Gln Glu Asn Gly Thr Val Gln Leu Leu Lys Gln Pro Gly Ser
785                 790                 795                 800

Thr Leu Asn Ala Gly Asp Ile Leu Ala Ile Leu Ala Leu Asp Asp Pro
                805                 810                 815

Ser Lys Val Lys His Ala Lys Pro Tyr Glu Gly Thr Leu Pro Glu Met
            820                 825                 830

Gly Asp Pro Thr Val Thr Gly Ser Lys Pro Ala His Leu Phe Gln His
        835                 840                 845

Tyr Asp Thr Ile Leu Lys Asn Ile Leu Ala Gly Tyr Asp Asn Gln Val
    850                 855                 860

Ile Leu Asn Ser Thr Leu Lys Asn Met Met Glu Ile Leu Lys Asn Lys
865                 870                 875                 880

Glu Leu Pro Tyr Ser Glu Trp Arg Leu Gln Ile Ser Ala Leu His Ser
                885                 890                 895

Arg Ile Pro Pro Lys Leu Asp Glu Ala Leu Thr Ser Leu Ile Glu Arg
            900                 905                 910

Thr Glu Ser Arg Gly Ala Glu Phe Pro Ala Arg Gln Ile Leu Lys Leu
        915                 920                 925

Val Asn Lys Thr Leu Gly Glu Pro Gly Asn Glu Leu Leu Gly Asp Val
    930                 935                 940

Val Ala Pro Leu Val Ser Ile Ala Asn Arg Tyr Gln Asn Gly Leu Val
945                 950                 955                 960

Glu His Glu Tyr Asp Tyr Phe Ala Ser Leu Val Asn Glu Tyr Cys Asn
                965                 970                 975

Val Glu His Phe Phe Ser Gly Glu Asn Val Arg Glu Glu Asp Val Ile
            980                 985                 990

Leu Arg Leu Arg Asp Glu Asn Lys  Ser Asp Leu Lys Lys  Val Ile Ser
        995                 1000                1005

Ile Cys  Leu Ser His Ser Arg  Val Ser Ala Lys Asn  Asn Leu Ile
    1010                1015                1020

Leu Ala  Ile Leu Glu Ala Tyr  Glu Pro Leu Leu Gln  Ser Asn Ser
```

1025 1030 1035

Ser Thr Ala Val Ala Ile Arg Asp Ser Leu Lys Lys Ile Val Gln
1040 1045 1050

Leu Asp Ser Arg Ala Cys Ala Lys Val Gly Leu Lys Ala Arg Glu
1055 1060 1065

Leu Leu Ile Gln Cys Ser Leu Pro Ser Ile Lys Glu Arg Ser Asp
1070 1075 1080

Gln Leu Glu His Ile Leu Arg Ser Ala Val Val Glu Thr Ser Tyr
1085 1090 1095

Gly Glu Val Phe Ala Lys His Arg Glu Pro Lys Leu Glu Ile Ile
1100 1105 1110

Gln Glu Val Val Glu Ser Lys His Val Val Phe Asp Val Leu Ser
1115 1120 1125

Gln Phe Leu Val His Gln Asp Cys Trp Val Ala Ile Ala Ala Ala
1130 1135 1140

Glu Val Tyr Val Arg Arg Ser Tyr Arg Ala Tyr Asp Leu Gly Lys
1145 1150 1155

Ile Asp Tyr His Ile His Asp Arg Leu Pro Ile Val Glu Trp Lys
1160 1165 1170

Phe Lys Leu Ala Gln Ile Ala Gly Ser Arg Tyr Asn Ala Val Gln
1175 1180 1185

Ser Ala Ser Val Gly Asp Asp Ser Thr Thr Met Lys His Ala Ala
1190 1195 1200

Ser Val Ser Asp Leu Ser Phe Val Val Asp Ser Lys Ser Glu Ser
1205 1210 1215

Thr Ser Arg Thr Gly Val Leu Val Pro Ala Arg His Leu Asp Asp
1220 1225 1230

Val Asp Glu Ile Leu Ser Ala Ala Leu Glu Tyr Phe Gln Pro Ser
1235 1240 1245

Asp Ala Leu Ser Phe Gln Ala Lys Gly Glu Arg Pro Glu Leu Leu
1250 1255 1260

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 84 of 3077

```
Asn Val  Leu Asn Ile Val Ile  Thr Asp Ile Asp Gly  Tyr Ser Asp
    1265                1270                1275

Glu Asp  Glu Cys Leu Lys Arg  Ile His Glu Ile Leu  Asn Glu Tyr
    1280                1285                1290

Glu Asp  Asp Leu Val Phe Ala  Gly Val Arg Arg Val  Thr Phe Val
    1295                1300                1305

Phe Ala  His Gln Val Gly Ser  Tyr Pro Lys Tyr Tyr  Thr Phe Thr
    1310                1315                1320

Gly Pro  Val Tyr Glu Glu Asn  Lys Val Ile Arg His  Ile Glu Pro
    1325                1330                1335

Ala Leu  Ala Phe Gln Leu Glu  Leu Gly Arg Leu Ala  Asn Phe Asp
    1340                1345                1350

Ile Lys  Pro Ile Phe Thr Asn  Asn Arg Asn Ile His  Val Tyr Glu
    1355                1360                1365

Ala Ile  Gly Lys Asn Ala Pro  Ser Asp Lys Arg Phe  Phe Thr Arg
    1370                1375                1380

Gly Ile  Ile Arg Gly Gly Val  Leu Lys Asp Glu Ile  Ser Leu Thr
    1385                1390                1395

Glu Tyr  Leu Ile Ala Glu Ser  Asn Arg Leu Ile Ser  Asp Ile Leu
    1400                1405                1410

Asp Thr  Leu Glu Val Ile Asp  Thr Ser Asn Ser Asp  Leu Asn His
    1415                1420                1425

Ile Phe  Ile Asn Phe Ser Asn  Val Phe Asn Val Gln  Pro Ala Asp
    1430                1435                1440

Val Glu  Ala Ala Phe Ala Ser  Phe Leu Glu Arg Phe  Gly Arg Arg
    1445                1450                1455

Leu Trp  Arg Leu Arg Val Thr  Gly Ala Glu Ile Arg  Ile Val Cys
    1460                1465                1470

Thr Asp  Pro Gln Gly Asn Ser  Phe Pro Leu Arg Ala  Ile Ile Asn
    1475                1480                1485

Asn Val  Ser Gly Tyr Val Val  Lys Ser Glu Leu Tyr  Leu Glu Val
    1490                1495                1500
```


LCY Biotechnology Holding, Inc.
Ex. 1002
Page 85 of 3077

```
Lys Asn  Pro Lys Gly Asp Trp  Val Phe Lys Ser Ile  Gly His Pro
    1505                1510                1515

Gly Ser  Met His Leu Gln Pro  Ile Ser Thr Pro Tyr  Pro Val Lys
    1520                1525                1530

Glu Ser  Leu Gln Pro Lys Arg  Tyr Arg Ala His Asn  Met Gly Thr
    1535                1540                1545

Thr Phe  Val Tyr Asp Phe Pro  Glu Leu Phe Arg Gln  Ala Thr Ile
    1550                1555                1560

Ser Gln  Trp Lys Lys His Gly  Lys Lys Ala Pro Lys  Asp Val Phe
    1565                1570                1575

Thr Ser  Leu Glu Leu Ile Thr  Asp Glu Asn Asp Ala  Leu Val Ala
    1580                1585                1590

Val Glu  Arg Asp Pro Gly Ala  Asn Lys Ile Gly Met  Val Gly Phe
    1595                1600                1605

Lys Val  Thr Ala Lys Thr Pro  Glu Tyr Pro Arg Gly  Arg Ser Phe
    1610                1615                1620

Ile Ile  Val Ala Asn Asp Ile  Thr His Lys Ile Gly  Ser Phe Gly
    1625                1630                1635

Pro Asp  Glu Asp Glu Tyr Phe  Asn Lys Cys Thr Asp  Leu Ala Arg
    1640                1645                1650

Lys Leu  Gly Val Pro Arg Ile  Tyr Leu Ser Ala Asn  Ser Gly Ala
    1655                1660                1665

Arg Ile  Gly Val Ala Glu Glu  Leu Ile Pro Leu Tyr  Gln Val Ala
    1670                1675                1680

Trp Asn  Glu Glu Gly Asn Pro  Asp Lys Gly Phe Arg  Tyr Leu Tyr
    1685                1690                1695

Leu Asn  Pro Asp Ala Lys Glu  Ala Leu Glu Lys Asp  Gly Lys Gly
    1700                1705                1710

Asp Thr  Ile Val Thr Glu Arg  Ile Val Glu Asp Gly  Gln Glu Arg
    1715                1720                1725

His Val  Ile Lys Ala Ile Ile  Gly Ala Glu Asn Gly  Leu Gly Val
    1730                1735                1740
```

```
Glu Cys  Leu Lys Gly Ser Gly  Leu Ile Ala Gly Ala  Thr Ser Arg
    1745                1750                1755

Ala Tyr  Arg Asp Ile Phe Thr  Ile Thr Leu Val Thr  Cys Arg Ser
    1760                1765                1770

Val Gly  Ile Gly Ala Tyr Leu  Val Arg Leu Gly Gln  Arg Ala Ile
    1775                1780                1785

Gln Ile  Glu Gly Gln Pro Ile  Ile Leu Thr Gly Ala  Pro Ala Ile
    1790                1795                1800

Asn Lys  Leu Leu Gly Arg Glu  Val Tyr Ser Ser Asn  Leu Gln Leu
    1805                1810                1815

Gly Gly  Thr Gln Ile Met Tyr  Asn Asn Gly Val Ser  His Leu Thr
    1820                1825                1830

Ala Ser  Asp Asp Leu Ala Gly  Val Glu Lys Ile Met  Glu Trp Leu
    1835                1840                1845

Ser Tyr  Val Pro Ala Lys Arg  Gly Met Pro Val Pro  Ile Leu Glu
    1850                1855                1860

Ser Glu  Asp Thr Trp Asp Arg  Asp Ile Asp Tyr Tyr  Pro Pro Lys
    1865                1870                1875

Gln Glu  Ala Phe Asp Ile Arg  Trp Met Ile Glu Gly  Lys Gln Val
    1880                1885                1890

Glu Gly  Glu Glu Phe Glu Ser  Gly Leu Phe Asp Lys  Gly Ser Phe
    1895                1900                1905

Gln Glu  Thr Leu Ser Gly Trp  Ala Lys Gly Val Val  Val Gly Arg
    1910                1915                1920

Ala Arg  Leu Gly Gly Ile Pro  Ile Gly Val Ile Gly  Val Glu Thr
    1925                1930                1935

Arg Thr  Ile Glu Asn Met Ile  Pro Ala Asp Pro Ala  Asn Pro Ser
    1940                1945                1950

Ser Thr  Glu Ala Leu Ile Gln  Glu Ala Gly Gln Val  Trp Tyr Pro
    1955                1960                1965

Asn Ser  Ala Phe Lys Thr Ala  Gln Ala Ile Asn Asp  Phe Asn Asn
```


LCY Biotechnology Holding, Inc.
Ex. 1002
Page 87 of 3077

1970 1975 1980

Gly Glu Gln Leu Pro Leu Met Ile Leu Ala Asn Trp Arg Gly Phe
1985 1990 1995

Ser Gly Gly Gln Arg Asp Met Tyr Asn Glu Val Leu Lys Tyr Gly
2000 2005 2010

Ser Phe Ile Val Asp Ala Leu Val Asp Phe Lys Gln Pro Ile Phe
2015 2020 2025

Thr Tyr Ile Pro Pro Asn Gly Glu Leu Arg Gly Gly Ser Trp Val
2030 2035 2040

Val Val Asp Pro Thr Ile Asn Ser Asp Met Met Glu Met Tyr Ala
2045 2050 2055

Asp Val Asp Ser Arg Ala Gly Val Leu Glu Pro Glu Gly Met Val
2060 2065 2070

Gly Ile Lys Tyr Arg Arg Asp Lys Leu Leu Ala Thr Met Gln Arg
2075 2080 2085

Leu Asp Pro Thr Tyr Ala Gln Leu Lys Glu Lys Leu Asn Asp Ser
2090 2095 2100

Ser Leu Ser Pro Glu Glu His Ala Gln Val Ser Thr Lys Ile Val
2105 2110 2115

Lys Arg Glu Lys Ala Leu Leu Pro Ile Tyr Ala Gln Ile Ser Val
2120 2125 2130

Gln Phe Ala Asp Leu His Asp Arg Ser Gly Arg Met Met Ala Lys
2135 2140 2145

Gly Val Ile Arg Lys Glu Ile Lys Trp Val Asp Ala Arg Arg Phe
2150 2155 2160

Phe Phe Trp Arg Leu Arg Arg Arg Leu Asn Glu Glu Tyr Val Leu
2165 2170 2175

Lys Leu Ile Gly Glu Gln Val Lys Asn Ala Asn Lys Leu Glu Lys
2180 2185 2190

Val Ala Arg Leu Lys Ser Trp Met Pro Thr Val Asp Tyr Asp Asp
2195 2200 2205

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 88 of 3077

```
Asp Gln  Ala Val Ser Thr Trp  Ile Glu Glu Asn His  Ala Lys Leu
    2210                 2215                 2220


Gln Lys  Arg Val Glu Glu Leu  Arg Gln Glu Lys Asn  Lys Ser Asp
    2225                 2230                 2235


Ile Val  Lys Leu Leu Gln Glu  Asp Pro Ser Asn Ala  Ala Ser Val
    2240                 2245                 2250


Met Arg  Asp Phe Val Asp Arg  Leu Ser Asp Glu Glu  Lys Glu Lys
    2255                 2260                 2265


Phe Leu  Lys Ser Leu Asn
    2270
```

<210> 18
<211> 2223
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 18

```
Met Thr Asp Leu Ser Pro Ser Pro Thr Asp Ser Leu Asn Tyr Thr Gln
1               5                   10                  15


Leu His Ser Ser Leu Pro Ser His Phe Leu Gly Gly Asn Ser Val Leu
            20                  25                  30


Thr Ala Glu Pro Ser Ala Val Thr Asp Phe Val Lys Thr His Gln Gly
        35                  40                  45


His Thr Val Ile Thr Lys Val Leu Ile Ala Asn Asn Gly Ile Gly Ala
    50                  55                  60


Val Lys Glu Ile Arg Ser Val Arg Lys Trp Ala Tyr Glu Thr Phe Gly
65                  70                  75                  80


Asp Glu Arg Ala Ile Gln Phe Val Ala Met Ala Thr Pro Glu Asp Met
                85                  90                  95


Glu Ala Asn Ala Glu Tyr Ile Arg Met Ala Asp Gln Phe Val Glu Val
            100                 105                 110


Pro Gly Gly Thr Asn Asn Asn Asn Tyr Ala Asn Val Asp Leu Ile Val
        115                 120                 125


Glu Ile Ala Glu Arg Thr Asp Val His Ala Val Trp Ala Gly Trp Gly
```

130 135 140

His Ala Ser Glu Asn Pro Leu Leu Pro Glu Arg Leu Ala Ala Ser Pro
145 150 155 160

Lys Lys Ile Val Phe Ile Gly Pro Pro Gly Ser Ala Met Arg Ser Leu
165 170 175

Gly Asp Lys Ile Ser Ser Thr Ile Val Ala Gln His Ala Lys Val Pro
180 185 190

Cys Ile Pro Trp Ser Gly Thr Gly Val Glu Glu Val His Val Asp Pro
195 200 205

Glu Thr Lys Leu Val Ser Val Asp Asp His Val Tyr Ala Lys Gly Cys
210 215 220

Cys Thr Ser Pro Glu Asp Gly Leu Glu Lys Ala Lys Arg Ile Gly Phe
225 230 235 240

Pro Val Met Val Lys Ala Ser Glu Gly Gly Gly Gly Lys Gly Ile Arg
245 250 255

Lys Val Asp His Glu Lys Asp Phe Ile Ser Leu Tyr Asn Gln Ala Ala
260 265 270

Asn Glu Ile Pro Gly Ser Pro Ile Phe Ile Met Lys Leu Ala Gly Asp
275 280 285

Ala Arg His Leu Glu Val Gln Leu Phe Ala Asp Gln Tyr Gly Thr Asn
290 295 300

Ile Ser Leu Phe Gly Arg Asp Cys Ser Val Gln Arg Arg His Gln Lys
305 310 315 320

Ile Ile Glu Glu Ala Pro Val Thr Ile Ala Asn Lys Asp Thr Phe Val
325 330 335

Glu Met Glu Lys Ala Ala Val Arg Leu Gly Lys Leu Val Gly Tyr Val
340 345 350

Ser Ala Gly Thr Val Glu Tyr Leu Tyr Ser Tyr Ala Glu Asp Lys Phe
355 360 365

Tyr Phe Leu Glu Leu Asn Pro Arg Leu Gln Val Glu His Pro Thr Thr
370 375 380

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 90 of 3077

```
Glu Met Val Ser Gly Val Asn Leu Pro Ala Ala Gln Leu Gln Ile Ala
385                 390                 395                 400

Met Gly Leu Pro Met His Arg Ile Arg Asp Ile Arg Leu Leu Tyr Gly
                405                 410                 415

Val Asp Pro His Ser Ala Thr Glu Ile Asp Phe Glu Phe Lys Ser Pro
            420                 425                 430

Asn Ser Leu Ile Thr Gln Arg Lys Pro Ala Pro Lys Gly His Cys Thr
        435                 440                 445

Ala Cys Arg Ile Thr Ser Glu Asp Pro Gly Glu Gly Phe Lys Pro Ser
    450                 455                 460

Gly Gly Thr Leu His Glu Leu Asn Phe Arg Ser Ser Ser Asn Val Trp
465                 470                 475                 480

Gly Tyr Phe Ser Val Ala Asn Gln Ser Ser Ile His Ser Phe Ala Asp
                485                 490                 495

Ser Gln Phe Gly His Ile Phe Ala Phe Gly Glu Asn Arg Gln Ala Ser
            500                 505                 510

Arg Lys His Met Ile Val Ala Leu Lys Glu Leu Ser Ile Arg Gly Asp
        515                 520                 525

Phe Arg Thr Thr Val Glu Tyr Leu Ile Lys Leu Leu Glu Thr Pro Asp
    530                 535                 540

Phe Ala Asp Asn Thr Ile Thr Thr Gly Trp Leu Asp Glu Leu Ile Thr
545                 550                 555                 560

Lys Lys Leu Thr Ala Glu Arg Pro Asp Pro Ile Val Ala Val Val Cys
                565                 570                 575

Gly Ala Val Thr Lys Ala His Ile Gln Ala Glu Glu Asp Lys Lys Glu
            580                 585                 590

Tyr Ile Glu Ser Leu Glu Lys Gly Gln Val Pro Asn Lys Ser Leu Leu
        595                 600                 605

Lys Thr Ile Phe Pro Val Glu Phe Ile Tyr Glu Gly Glu Arg Tyr Lys
    610                 615                 620

Phe Thr Ala Thr Lys Ser Ser Glu Asp Lys Tyr Thr Leu Phe Leu Asn
625                 630                 635                 640
```

```
Gly Ser Arg Cys Val Ile Gly Ala Arg Ser Leu Ser Asp Gly Gly Leu
                645                 650                 655

Leu Cys Ala Leu Asp Gly Lys Ser His Ser Val Tyr Trp Lys Glu Glu
            660                 665                 670

Ala Ala Ala Thr Arg Leu Ser Val Asp Gly Lys Thr Cys Leu Leu Glu
        675                 680                 685

Val Glu Asn Asp Pro Thr Gln Leu Arg Thr Pro Ser Pro Gly Lys Leu
    690                 695                 700

Val Lys Tyr Leu Val Glu Ser Gly Glu His Val Asp Ala Gly Gln Ser
705                 710                 715                 720

Tyr Ala Glu Val Glu Val Met Lys Met Cys Met Pro Leu Ile Ala Gln
                725                 730                 735

Glu Asn Gly Thr Val Gln Leu Leu Lys Gln Pro Gly Ser Thr Leu Asn
            740                 745                 750

Ala Gly Asp Ile Leu Ala Ile Leu Ala Leu Asp Asp Pro Ser Lys Val
        755                 760                 765

Lys His Ala Lys Pro Tyr Glu Gly Thr Leu Pro Glu Met Gly Asp Pro
    770                 775                 780

Thr Val Thr Gly Ser Lys Pro Ala His Leu Phe Gln His Tyr Asp Thr
785                 790                 795                 800

Ile Leu Lys Asn Ile Leu Ala Gly Tyr Asp Asn Gln Val Ile Leu Asn
                805                 810                 815

Ser Thr Leu Lys Asn Met Met Glu Ile Leu Lys Asn Lys Glu Leu Pro
            820                 825                 830

Tyr Ser Glu Trp Arg Leu Gln Ile Ser Ala Leu His Ser Arg Ile Pro
        835                 840                 845

Pro Lys Leu Asp Glu Ala Leu Thr Ser Leu Ile Glu Arg Thr Glu Ser
    850                 855                 860

Arg Gly Ala Glu Phe Pro Ala Arg Gln Ile Leu Lys Leu Val Asn Lys
865                 870                 875                 880

Thr Leu Gly Glu Pro Gly Asn Glu Leu Leu Gly Asp Val Val Ala Pro
                885                 890                 895
```

```
Leu Val Ser Ile Ala Asn Arg Tyr Gln Asn Gly Leu Val Glu His Glu
            900                 905                 910

Tyr Asp Tyr Phe Ala Ser Leu Val Asn Glu Tyr Cys Asn Val Glu His
        915                 920                 925

Phe Phe Ser Gly Glu Asn Val Arg Glu Glu Asp Val Ile Leu Arg Leu
    930                 935                 940

Arg Asp Glu Asn Lys Ser Asp Leu Lys Lys Val Ile Ser Ile Cys Leu
945                 950                 955                 960

Ser His Ser Arg Val Ser Ala Lys Asn Asn Leu Ile Leu Ala Ile Leu
                965                 970                 975

Glu Ala Tyr Glu Pro Leu Leu Gln Ser Asn Ser Ser Thr Ala Val Ala
            980                 985                 990

Ile Arg Asp Ser Leu Lys Lys Ile  Val Gln Leu Asp Ser  Arg Ala Cys
        995                 1000                1005

Ala Lys  Val Gly Leu Lys Ala  Arg Glu Leu Leu Ile  Gln Cys Ser
    1010                1015                1020

Leu Pro  Ser Ile Lys Glu Arg  Ser Asp Gln Leu Glu  His Ile Leu
    1025                1030                1035

Arg Ser  Ala Val Val Glu Thr  Ser Tyr Gly Glu Val  Phe Ala Lys
    1040                1045                1050

His Arg  Glu Pro Lys Leu Glu  Ile Ile Gln Glu Val  Val Glu Ser
    1055                1060                1065

Lys His  Val Val Phe Asp Val  Leu Ser Gln Phe Leu  Val His Gln
    1070                1075                1080

Asp Cys  Trp Val Ala Ile Ala  Ala Ala Glu Val Tyr  Val Arg Arg
    1085                1090                1095

Ser Tyr  Arg Ala Tyr Asp Leu  Gly Lys Ile Asp Tyr  His Ile His
    1100                1105                1110

Asp Arg  Leu Pro Ile Val Glu  Trp Lys Phe Lys Leu  Ala Gln Ile
    1115                1120                1125

Ala Gly  Ser Arg Tyr Asn Ala  Val Gln Ser Ala Ser  Val Gly Asp
```

1130 1135 1140

Asp Ser Thr Thr Met Lys His Ala Ala Ser Val Ser Asp Leu Ser
1145 1150 1155

Phe Val Val Asp Ser Lys Ser Glu Ser Thr Ser Arg Thr Gly Val
1160 1165 1170

Leu Val Pro Ala Arg His Leu Asp Asp Val Asp Glu Ile Leu Ser
1175 1180 1185

Ala Ala Leu Glu Tyr Phe Gln Pro Ser Asp Ala Leu Ser Phe Gln
1190 1195 1200

Ala Lys Gly Glu Arg Pro Glu Leu Leu Asn Val Leu Asn Ile Val
1205 1210 1215

Ile Thr Asp Ile Asp Gly Tyr Ser Asp Glu Asp Glu Cys Leu Lys
1220 1225 1230

Arg Ile His Glu Ile Leu Asn Glu Tyr Glu Asp Asp Leu Val Phe
1235 1240 1245

Ala Gly Val Arg Arg Val Thr Phe Val Phe Ala His Gln Val Gly
1250 1255 1260

Ser Tyr Pro Lys Tyr Tyr Thr Phe Thr Gly Pro Val Tyr Glu Glu
1265 1270 1275

Asn Lys Val Ile Arg His Ile Glu Pro Ala Leu Ala Phe Gln Leu
1280 1285 1290

Glu Leu Gly Arg Leu Ala Asn Phe Asp Ile Lys Pro Ile Phe Thr
1295 1300 1305

Asn Asn Arg Asn Ile His Val Tyr Glu Ala Ile Gly Lys Asn Ala
1310 1315 1320

Pro Ser Asp Lys Arg Phe Phe Thr Arg Gly Ile Ile Arg Gly Gly
1325 1330 1335

Val Leu Lys Asp Glu Ile Ser Leu Thr Glu Tyr Leu Ile Ala Glu
1340 1345 1350

Ser Asn Arg Leu Ile Ser Asp Ile Leu Asp Thr Leu Glu Val Ile
1355 1360 1365

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 94 of 3077

```
Asp Thr  Ser Asn Ser Asp Leu  Asn His Ile Phe Ile  Asn Phe Ser
    1370                1375                1380

Asn Val  Phe Asn Val Gln Pro  Ala Asp Val Glu Ala  Ala Phe Ala
    1385                1390                1395

Ser Phe  Leu Glu Arg Phe Gly  Arg Arg Leu Trp Arg  Leu Arg Val
    1400                1405                1410

Thr Gly  Ala Glu Ile Arg Ile  Val Cys Thr Asp Pro  Gln Gly Asn
    1415                1420                1425

Ser Phe  Pro Leu Arg Ala Ile  Ile Asn Asn Val Ser  Gly Tyr Val
    1430                1435                1440

Val Lys  Ser Glu Leu Tyr Leu  Glu Val Lys Asn Pro  Lys Gly Asp
    1445                1450                1455

Trp Val  Phe Lys Ser Ile Gly  His Pro Gly Ser Met  His Leu Gln
    1460                1465                1470

Pro Ile  Ser Thr Pro Tyr Pro  Val Lys Glu Ser Leu  Gln Pro Lys
    1475                1480                1485

Arg Tyr  Arg Ala His Asn Met  Gly Thr Thr Phe Val  Tyr Asp Phe
    1490                1495                1500

Pro Glu  Leu Phe Arg Gln Ala  Thr Ile Ser Gln Trp  Lys Lys His
    1505                1510                1515

Gly Lys  Lys Ala Pro Lys Asp  Val Phe Thr Ser Leu  Glu Leu Ile
    1520                1525                1530

Thr Asp  Glu Asn Asp Ala Leu  Val Ala Val Glu Arg  Asp Pro Gly
    1535                1540                1545

Ala Asn  Lys Ile Gly Met Val  Gly Phe Lys Val Thr  Ala Lys Thr
    1550                1555                1560

Pro Glu  Tyr Pro Arg Gly Arg  Ser Phe Ile Ile Val  Ala Asn Asp
    1565                1570                1575

Ile Thr  His Lys Ile Gly Ser  Phe Gly Pro Asp Glu  Asp Glu Tyr
    1580                1585                1590

Phe Asn  Lys Cys Thr Asp Leu  Ala Arg Lys Leu Gly  Val Pro Arg
    1595                1600                1605
```

```
Ile Tyr  Leu Ser Ala Asn Ser  Gly Ala Arg Ile Gly  Val Ala Glu
    1610                 1615                 1620

Glu Leu  Ile Pro Leu Tyr Gln  Val Ala Trp Asn Glu  Glu Gly Asn
    1625                 1630                 1635

Pro Asp  Lys Gly Phe Arg Tyr  Leu Tyr Leu Asn Pro  Asp Ala Lys
    1640                 1645                 1650

Glu Ala  Leu Glu Lys Asp Gly  Lys Gly Asp Thr Ile  Val Thr Glu
    1655                 1660                 1665

Arg Ile  Val Glu Asp Gly Gln  Glu Arg His Val Ile  Lys Ala Ile
    1670                 1675                 1680

Ile Gly  Ala Glu Asn Gly Leu  Gly Val Glu Cys Leu  Lys Gly Ser
    1685                 1690                 1695

Gly Leu  Ile Ala Gly Ala Thr  Ser Arg Ala Tyr Arg  Asp Ile Phe
    1700                 1705                 1710

Thr Ile  Thr Leu Val Thr Cys  Arg Ser Val Gly Ile  Gly Ala Tyr
    1715                 1720                 1725

Leu Val  Arg Leu Gly Gln Arg  Ala Ile Gln Ile Glu  Gly Gln Pro
    1730                 1735                 1740

Ile Ile  Leu Thr Gly Ala Pro  Ala Ile Asn Lys Leu  Leu Gly Arg
    1745                 1750                 1755

Glu Val  Tyr Ser Ser Asn Leu  Gln Leu Gly Gly Thr  Gln Ile Met
    1760                 1765                 1770

Tyr Asn  Asn Gly Val Ser His  Leu Thr Ala Ser Asp  Asp Leu Ala
    1775                 1780                 1785

Gly Val  Glu Lys Ile Met Glu  Trp Leu Ser Tyr Val  Pro Ala Lys
    1790                 1795                 1800

Arg Gly  Met Pro Val Pro Ile  Leu Glu Ser Glu Asp  Thr Trp Asp
    1805                 1810                 1815

Arg Asp  Ile Asp Tyr Tyr Pro  Pro Lys Gln Glu Ala  Phe Asp Ile
    1820                 1825                 1830

Arg Trp  Met Ile Glu Gly Lys  Gln Val Glu Gly Glu  Glu Phe Glu
    1835                 1840                 1845
```

```
Ser Gly  Leu Phe Asp Lys Gly  Ser Phe Gln Glu Thr  Leu Ser Gly
    1850                 1855                 1860

Trp Ala  Lys Gly Val Val Val  Gly Arg Ala Arg Leu  Gly Gly Ile
    1865                 1870                 1875

Pro Ile  Gly Val Ile Gly Val  Glu Thr Arg Thr Ile  Glu Asn Met
    1880                 1885                 1890

Ile Pro  Ala Asp Pro Ala Asn  Pro Ser Ser Thr Glu  Ala Leu Ile
    1895                 1900                 1905

Gln Glu  Ala Gly Gln Val Trp  Tyr Pro Asn Ser Ala  Phe Lys Thr
    1910                 1915                 1920

Ala Gln  Ala Ile Asn Asp Phe  Asn Asn Gly Glu Gln  Leu Pro Leu
    1925                 1930                 1935

Met Ile  Leu Ala Asn Trp Arg  Gly Phe Ser Gly Gly  Gln Arg Asp
    1940                 1945                 1950

Met Tyr  Asn Glu Val Leu Lys  Tyr Gly Ser Phe Ile  Val Asp Ala
    1955                 1960                 1965

Leu Val  Asp Phe Lys Gln Pro  Ile Phe Thr Tyr Ile  Pro Pro Asn
    1970                 1975                 1980

Gly Glu  Leu Arg Gly Gly Ser  Trp Val Val Val Asp  Pro Thr Ile
    1985                 1990                 1995

Asn Ser  Asp Met Met Glu Met  Tyr Ala Asp Val Asp  Ser Arg Ala
    2000                 2005                 2010

Gly Val  Leu Glu Pro Glu Gly  Met Val Gly Ile Lys  Tyr Arg Arg
    2015                 2020                 2025

Asp Lys  Leu Leu Ala Thr Met  Gln Arg Leu Asp Pro  Thr Tyr Ala
    2030                 2035                 2040

Gln Leu  Lys Glu Lys Leu Asn  Asp Ser Ser Leu Ser  Pro Glu Glu
    2045                 2050                 2055

His Ala  Gln Val Ser Thr Lys  Ile Val Lys Arg Glu  Lys Ala Leu
    2060                 2065                 2070

Leu Pro  Ile Tyr Ala Gln Ile  Ser Val Gln Phe Ala  Asp Leu His
```

2075 2080 2085

Asp Arg Ser Gly Arg Met Met Ala Lys Gly Val Ile Arg Lys Glu
2090 2095 2100

Ile Lys Trp Val Asp Ala Arg Arg Phe Phe Phe Trp Arg Leu Arg
2105 2110 2115

Arg Arg Leu Asn Glu Glu Tyr Val Leu Lys Leu Ile Gly Glu Gln
2120 2125 2130

Val Lys Asn Ala Asn Lys Leu Glu Lys Val Ala Arg Leu Lys Ser
2135 2140 2145

Trp Met Pro Thr Val Asp Tyr Asp Asp Asp Gln Ala Val Ser Thr
2150 2155 2160

Trp Ile Glu Glu Asn His Ala Lys Leu Gln Lys Arg Val Glu Glu
2165 2170 2175

Leu Arg Gln Glu Lys Asn Lys Ser Asp Ile Val Lys Leu Leu Gln
2180 2185 2190

Glu Asp Pro Ser Asn Ala Ala Ser Val Met Arg Asp Phe Val Asp
2195 2200 2205

Arg Leu Ser Asp Glu Glu Lys Glu Lys Phe Leu Lys Ser Leu Asn
2210 2215 2220

<210> 19
<211> 2223
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 19
Met Thr Asp Leu Ser Pro Ser Pro Thr Asp Ser Leu Asn Tyr Thr Gln
1 5 10 15

Leu His Ser Ser Leu Pro Ser His Phe Leu Gly Gly Asn Ser Val Leu
20 25 30

Thr Ala Glu Pro Ser Ala Val Thr Asp Phe Val Lys Thr His Gln Gly
35 40 45

His Thr Val Ile Thr Lys Val Leu Ile Ala Asn Asn Gly Ile Gly Ala
50 55 60

```
Val Lys Glu Ile Arg Ser Val Arg Lys Trp Ala Tyr Glu Thr Phe Gly
65                  70                  75                  80

Asp Glu Arg Ala Ile Gln Phe Val Ala Met Ala Thr Pro Glu Asp Met
                85                  90                  95

Glu Ala Asn Ala Glu Tyr Ile Arg Met Ala Asp Gln Phe Val Glu Val
            100                 105                 110

Pro Gly Gly Thr Asn Asn Asn Asn Tyr Ala Asn Val Asp Leu Ile Val
        115                 120                 125

Glu Ile Ala Glu Arg Thr Asp Val His Ala Val Trp Ala Gly Trp Gly
    130                 135                 140

His Ala Ser Glu Asn Pro Leu Leu Pro Glu Arg Leu Ala Ala Ser Pro
145                 150                 155                 160

Lys Lys Ile Val Phe Ile Gly Pro Pro Gly Ser Ala Met Arg Ser Leu
                165                 170                 175

Gly Asp Lys Ile Ser Ser Thr Ile Val Ala Gln His Ala Lys Val Pro
            180                 185                 190

Cys Ile Pro Trp Ser Gly Thr Gly Val Glu Glu Val His Val Asp Pro
        195                 200                 205

Glu Thr Lys Leu Val Ser Val Asp Asp His Val Tyr Ala Lys Gly Cys
    210                 215                 220

Cys Thr Ser Pro Glu Asp Gly Leu Glu Lys Ala Lys Arg Ile Gly Phe
225                 230                 235                 240

Pro Val Met Val Lys Ala Ser Glu Gly Gly Gly Gly Lys Gly Ile Arg
                245                 250                 255

Lys Val Asp His Glu Lys Asp Phe Ile Ser Leu Tyr Asn Gln Ala Ala
            260                 265                 270

Asn Glu Ile Pro Gly Ser Pro Ile Phe Ile Met Lys Leu Ala Gly Asp
        275                 280                 285

Ala Arg His Leu Glu Val Gln Leu Phe Ala Asp Gln Tyr Gly Thr Asn
    290                 295                 300

Ile Ser Leu Phe Gly Arg Asp Cys Ser Val Gln Arg Arg His Gln Lys
```


LCY Biotechnology Holding, Inc.
Ex. 1002
Page 99 of 3077

305 310 315 320

Ile Ile Glu Glu Ala Pro Val Thr Ile Ala Asn Lys Asp Thr Phe Val
325 330 335

Glu Met Glu Lys Ala Ala Val Arg Leu Gly Lys Leu Val Gly Tyr Val
340 345 350

Ser Ala Gly Thr Val Glu Tyr Leu Tyr Ser Tyr Ala Glu Asp Lys Phe
355 360 365

Tyr Phe Leu Glu Leu Asn Pro Arg Leu Gln Val Glu His Pro Thr Thr
370 375 380

Glu Met Val Ser Gly Val Asn Leu Pro Ala Ala Gln Leu Gln Ile Ala
385 390 395 400

Met Gly Leu Pro Met His Arg Ile Arg Asp Ile Arg Leu Leu Tyr Gly
405 410 415

Val Asp Pro His Ser Ala Thr Glu Ile Asp Phe Glu Phe Lys Ser Pro
420 425 430

Asn Ser Leu Ile Thr Gln Arg Lys Pro Ala Pro Lys Gly His Cys Thr
435 440 445

Ala Cys Arg Ile Thr Ser Glu Asp Pro Gly Glu Gly Phe Lys Pro Ser
450 455 460

Gly Gly Thr Leu His Glu Leu Asn Phe Arg Ser Ser Ser Asn Val Trp
465 470 475 480

Gly Tyr Phe Ser Val Ala Asn Gln Ser Ser Ile His Ser Phe Ala Asp
485 490 495

Ser Gln Phe Gly His Ile Phe Ala Phe Gly Glu Asn Arg Gln Ala Ser
500 505 510

Arg Lys His Met Ile Val Ala Leu Lys Glu Leu Ser Ile Arg Gly Asp
515 520 525

Phe Arg Thr Thr Val Glu Tyr Leu Ile Lys Leu Leu Glu Thr Pro Asp
530 535 540

Phe Ala Asp Asn Thr Ile Thr Thr Gly Trp Leu Asp Glu Leu Ile Thr
545 550 555 560

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 100 of 3077

```
Lys Lys Leu Thr Ala Glu Arg Pro Asp Pro Ile Val Ala Val Val Cys
                565                 570                 575

Gly Ala Val Thr Lys Ala His Ile Gln Ala Glu Glu Asp Lys Lys Glu
            580                 585                 590

Tyr Ile Glu Ser Leu Glu Lys Gly Gln Val Pro Asn Lys Ser Leu Leu
        595                 600                 605

Lys Thr Ile Phe Pro Val Glu Phe Ile Tyr Glu Gly Glu Arg Tyr Lys
    610                 615                 620

Phe Thr Ala Thr Lys Ser Ser Glu Asp Lys Tyr Thr Leu Phe Leu Asn
625                 630                 635                 640

Gly Ser Arg Cys Val Ile Gly Ala Arg Ser Leu Ser Asp Gly Gly Leu
                645                 650                 655

Leu Cys Ala Leu Asp Gly Lys Ser His Ser Val Tyr Trp Lys Glu Glu
            660                 665                 670

Ala Ala Ala Thr Arg Leu Ser Val Asp Gly Lys Thr Cys Leu Leu Glu
        675                 680                 685

Val Glu Asn Asp Pro Thr Gln Leu Arg Thr Pro Ser Pro Gly Lys Leu
    690                 695                 700

Val Lys Tyr Leu Val Glu Ser Gly Glu His Val Asp Ala Gly Gln Ser
705                 710                 715                 720

Tyr Ala Glu Val Glu Val Met Lys Met Cys Met Pro Leu Ile Ala Gln
                725                 730                 735

Glu Asn Gly Thr Val Gln Leu Leu Lys Gln Pro Gly Ser Thr Leu Asn
            740                 745                 750

Ala Gly Asp Ile Leu Ala Ile Leu Ala Leu Asp Asp Pro Ser Lys Val
        755                 760                 765

Lys His Ala Lys Pro Tyr Glu Gly Thr Leu Pro Glu Met Gly Asp Pro
    770                 775                 780

Thr Val Thr Gly Ser Lys Pro Ala His Leu Phe Gln His Tyr Asp Thr
785                 790                 795                 800

Ile Leu Lys Asn Ile Leu Ala Gly Tyr Asp Asn Gln Val Ile Leu Asn
                805                 810                 815
```

```
Ser Thr Leu Lys Asn Met Met Glu Ile Leu Lys Asn Lys Glu Leu Pro
            820                 825                 830

Tyr Ser Glu Trp Arg Leu Gln Ile Ser Ala Leu His Ser Arg Ile Pro
        835                 840                 845

Pro Lys Leu Asp Glu Ala Leu Thr Ser Leu Ile Glu Arg Thr Glu Ser
    850                 855                 860

Arg Gly Ala Glu Phe Pro Ala Arg Gln Ile Leu Lys Leu Val Asn Lys
865                 870                 875                 880

Thr Leu Gly Glu Pro Gly Asn Glu Leu Leu Gly Asp Val Val Ala Pro
                885                 890                 895

Leu Val Ser Ile Ala Asn Arg Tyr Gln Asn Gly Leu Val Glu His Glu
            900                 905                 910

Tyr Asp Tyr Phe Ala Ser Leu Val Asn Glu Tyr Cys Asn Val Glu His
        915                 920                 925

Phe Phe Ser Gly Glu Asn Val Arg Glu Glu Asp Val Ile Leu Arg Leu
    930                 935                 940

Arg Asp Glu Asn Lys Ser Asp Leu Lys Lys Val Ile Ser Ile Cys Leu
945                 950                 955                 960

Ser His Ser Arg Val Ser Ala Lys Asn Asn Leu Ile Leu Ala Ile Leu
                965                 970                 975

Glu Ala Tyr Glu Pro Leu Leu Gln Ser Asn Ser Ser Thr Ala Val Ala
            980                 985                 990

Ile Arg Asp Ser Leu Lys Lys Ile  Val Gln Leu Asp Ser  Arg Ala Cys
        995                 1000                1005

Ala Lys  Val Gly Leu Lys Ala  Arg Glu Leu Leu Ile  Gln Cys Ser
    1010                1015                1020

Leu Pro  Ser Ile Lys Glu Arg  Ser Asp Gln Leu Glu  His Ile Leu
    1025                1030                1035

Arg Ser  Ala Val Val Glu Thr  Ser Tyr Gly Glu Val  Phe Ala Lys
    1040                1045                1050

His Arg  Glu Pro Lys Leu Glu  Ile Ile Gln Glu Val  Val Glu Ser
    1055                1060                1065
```

```
Lys His  Val Val Phe Asp Val  Leu Ser Gln Phe Leu  Val His Gln
    1070                1075                1080

Asp Cys  Trp Val Ala Ile Ala  Ala Ala Glu Val Tyr  Val Arg Arg
    1085                1090                1095

Ser Tyr  Arg Ala Tyr Asp Leu  Gly Lys Ile Asp Tyr  His Ile His
    1100                1105                1110

Asp Arg  Leu Pro Ile Val Glu  Trp Lys Phe Lys Leu  Ala Gln Ile
    1115                1120                1125

Ala Gly  Ser Arg Tyr Asn Ala  Val Gln Ser Ala Ser  Val Gly Asp
    1130                1135                1140

Asp Ser  Thr Thr Met Lys His  Ala Ala Ala Val Ser  Asp Leu Ser
    1145                1150                1155

Phe Val  Val Asp Ser Lys Ser  Glu Ser Thr Ser Arg  Thr Gly Val
    1160                1165                1170

Leu Val  Pro Ala Arg His Leu  Asp Asp Val Asp Glu  Ile Leu Ser
    1175                1180                1185

Ala Ala  Leu Glu Tyr Phe Gln  Pro Ser Asp Ala Leu  Ser Phe Gln
    1190                1195                1200

Ala Lys  Gly Glu Arg Pro Glu  Leu Leu Asn Val Leu  Asn Ile Val
    1205                1210                1215

Ile Thr  Asp Ile Asp Gly Tyr  Ser Asp Glu Asp Glu  Cys Leu Lys
    1220                1225                1230

Arg Ile  His Glu Ile Leu Asn  Glu Tyr Glu Asp Asp  Leu Val Phe
    1235                1240                1245

Ala Gly  Val Arg Arg Val Thr  Phe Val Phe Ala His  Gln Val Gly
    1250                1255                1260

Ser Tyr  Pro Lys Tyr Tyr Thr  Phe Thr Gly Pro Val  Tyr Glu Glu
    1265                1270                1275

Asn Lys  Val Ile Arg His Ile  Glu Pro Ala Leu Ala  Phe Gln Leu
    1280                1285                1290

Glu Leu  Gly Arg Leu Ala Asn  Phe Asp Ile Lys Pro  Ile Phe Thr
```

1295 1300 1305

Asn Asn Arg Asn Ile His Val Tyr Glu Ala Ile Gly Lys Asn Ala
1310 1315 1320

Pro Ser Asp Lys Arg Phe Phe Thr Arg Gly Ile Ile Arg Gly Gly
1325 1330 1335

Val Leu Lys Asp Glu Ile Ser Leu Thr Glu Tyr Leu Ile Ala Glu
1340 1345 1350

Ser Asn Arg Leu Ile Ser Asp Ile Leu Asp Thr Leu Glu Val Ile
1355 1360 1365

Asp Thr Ser Asn Ser Asp Leu Asn His Ile Phe Ile Asn Phe Ser
1370 1375 1380

Asn Val Phe Asn Val Gln Pro Ala Asp Val Glu Ala Ala Phe Ala
1385 1390 1395

Ser Phe Leu Glu Arg Phe Gly Arg Arg Leu Trp Arg Leu Arg Val
1400 1405 1410

Thr Gly Ala Glu Ile Arg Ile Val Cys Thr Asp Pro Gln Gly Asn
1415 1420 1425

Ser Phe Pro Leu Arg Ala Ile Ile Asn Asn Val Ser Gly Tyr Val
1430 1435 1440

Val Lys Ser Glu Leu Tyr Leu Glu Val Lys Asn Pro Lys Gly Asp
1445 1450 1455

Trp Val Phe Lys Ser Ile Gly His Pro Gly Ser Met His Leu Gln
1460 1465 1470

Pro Ile Ser Thr Pro Tyr Pro Val Lys Glu Ser Leu Gln Pro Lys
1475 1480 1485

Arg Tyr Arg Ala His Asn Met Gly Thr Thr Phe Val Tyr Asp Phe
1490 1495 1500

Pro Glu Leu Phe Arg Gln Ala Thr Ile Ser Gln Trp Lys Lys His
1505 1510 1515

Gly Lys Lys Ala Pro Lys Asp Val Phe Thr Ser Leu Glu Leu Ile
1520 1525 1530

```
Thr Asp  Glu Asn Asp Ala Leu  Val Ala Val Glu Arg  Asp Pro Gly
    1535                1540                1545

Ala Asn  Lys Ile Gly Met Val  Gly Phe Lys Val Thr  Ala Lys Thr
    1550                1555                1560

Pro Glu  Tyr Pro Arg Gly Arg  Ser Phe Ile Ile Val  Ala Asn Asp
    1565                1570                1575

Ile Thr  His Lys Ile Gly Ser  Phe Gly Pro Asp Glu  Asp Glu Tyr
    1580                1585                1590

Phe Asn  Lys Cys Thr Asp Leu  Ala Arg Lys Leu Gly  Val Pro Arg
    1595                1600                1605

Ile Tyr  Leu Ser Ala Asn Ser  Gly Ala Arg Ile Gly  Val Ala Glu
    1610                1615                1620

Glu Leu  Ile Pro Leu Tyr Gln  Val Ala Trp Asn Glu  Glu Gly Asn
    1625                1630                1635

Pro Asp  Lys Gly Phe Arg Tyr  Leu Tyr Leu Asn Pro  Asp Ala Lys
    1640                1645                1650

Glu Ala  Leu Glu Lys Asp Gly  Lys Gly Asp Thr Ile  Val Thr Glu
    1655                1660                1665

Arg Ile  Val Glu Asp Gly Gln  Glu Arg His Val Ile  Lys Ala Ile
    1670                1675                1680

Ile Gly  Ala Glu Asn Gly Leu  Gly Val Glu Cys Leu  Lys Gly Ser
    1685                1690                1695

Gly Leu  Ile Ala Gly Ala Thr  Ser Arg Ala Tyr Arg  Asp Ile Phe
    1700                1705                1710

Thr Ile  Thr Leu Val Thr Cys  Arg Ser Val Gly Ile  Gly Ala Tyr
    1715                1720                1725

Leu Val  Arg Leu Gly Gln Arg  Ala Ile Gln Ile Glu  Gly Gln Pro
    1730                1735                1740

Ile Ile  Leu Thr Gly Ala Pro  Ala Ile Asn Lys Leu  Leu Gly Arg
    1745                1750                1755

Glu Val  Tyr Ser Ser Asn Leu  Gln Leu Gly Gly Thr  Gln Ile Met
    1760                1765                1770
```

```
Tyr Asn  Asn Gly Val Ser His  Leu Thr Ala Ser Asp  Asp Leu Ala
    1775                 1780                 1785

Gly Val  Glu Lys Ile Met Glu  Trp Leu Ser Tyr Val  Pro Ala Lys
    1790                 1795                 1800

Arg Gly  Met Pro Val Pro Ile  Leu Glu Ser Glu Asp  Thr Trp Asp
    1805                 1810                 1815

Arg Asp  Ile Asp Tyr Tyr Pro  Pro Lys Gln Glu Ala  Phe Asp Ile
    1820                 1825                 1830

Arg Trp  Met Ile Glu Gly Lys  Gln Val Glu Gly Glu  Glu Phe Glu
    1835                 1840                 1845

Ser Gly  Leu Phe Asp Lys Gly  Ser Phe Gln Glu Thr  Leu Ser Gly
    1850                 1855                 1860

Trp Ala  Lys Gly Val Val Val  Gly Arg Ala Arg Leu  Gly Gly Ile
    1865                 1870                 1875

Pro Ile  Gly Val Ile Gly Val  Glu Thr Arg Thr Ile  Glu Asn Met
    1880                 1885                 1890

Ile Pro  Ala Asp Pro Ala Asn  Pro Ser Ser Thr Glu  Ala Leu Ile
    1895                 1900                 1905

Gln Glu  Ala Gly Gln Val Trp  Tyr Pro Asn Ser Ala  Phe Lys Thr
    1910                 1915                 1920

Ala Gln  Ala Ile Asn Asp Phe  Asn Asn Gly Glu Gln  Leu Pro Leu
    1925                 1930                 1935

Met Ile  Leu Ala Asn Trp Arg  Gly Phe Ser Gly Gly  Gln Arg Asp
    1940                 1945                 1950

Met Tyr  Asn Glu Val Leu Lys  Tyr Gly Ser Phe Ile  Val Asp Ala
    1955                 1960                 1965

Leu Val  Asp Phe Lys Gln Pro  Ile Phe Thr Tyr Ile  Pro Pro Asn
    1970                 1975                 1980

Gly Glu  Leu Arg Gly Gly Ser  Trp Val Val Val Asp  Pro Thr Ile
    1985                 1990                 1995

Asn Ser  Asp Met Met Glu Met  Tyr Ala Asp Val Asp  Ser Arg Ala
    2000                 2005                 2010
```

```
Gly Val  Leu Glu Pro Glu Gly  Met Val Gly Ile Lys  Tyr Arg Arg
    2015                2020                2025

Asp Lys  Leu Leu Ala Thr Met  Gln Arg Leu Asp Pro  Thr Tyr Ala
    2030                2035                2040

Gln Leu  Lys Glu Lys Leu Asn  Asp Ser Ser Leu Ser  Pro Glu Glu
    2045                2050                2055

His Ala  Gln Val Ser Thr Lys  Ile Val Lys Arg Glu  Lys Ala Leu
    2060                2065                2070

Leu Pro  Ile Tyr Ala Gln Ile  Ser Val Gln Phe Ala  Asp Leu His
    2075                2080                2085

Asp Arg  Ser Gly Arg Met Met  Ala Lys Gly Val Ile  Arg Lys Glu
    2090                2095                2100

Ile Lys  Trp Val Asp Ala Arg  Arg Phe Phe Phe Trp  Arg Leu Arg
    2105                2110                2115

Arg Arg  Leu Asn Glu Glu Tyr  Val Leu Lys Leu Ile  Gly Glu Gln
    2120                2125                2130

Val Lys  Asn Ala Asn Lys Leu  Glu Lys Val Ala Arg  Leu Lys Ser
    2135                2140                2145

Trp Met  Pro Thr Val Asp Tyr  Asp Asp Asp Gln Ala  Val Ser Thr
    2150                2155                2160

Trp Ile  Glu Glu Asn His Ala  Lys Leu Gln Lys Arg  Val Glu Glu
    2165                2170                2175

Leu Arg  Gln Glu Lys Asn Lys  Ser Asp Ile Val Lys  Leu Leu Gln
    2180                2185                2190

Glu Asp  Pro Ser Asn Ala Ala  Ser Val Met Arg Asp  Phe Val Asp
    2195                2200                2205

Arg Leu  Ser Asp Glu Glu Lys  Glu Lys Phe Leu Lys  Ser Leu Asn
    2210                2215                2220
```

<210> 20
<211> 675
<212> PRT
<213> Candida viswanathii

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 107 of 3077

<400> 20

```
Met Pro Glu Ser Thr Gln Gln Ser His Ile Ser Leu Asp His Glu Lys
1               5                   10                  15

Met Gln Gln Pro Pro Thr Gly Phe Thr Glu Arg Ser Thr Ala Lys Pro
            20                  25                  30

Asn Leu Ala Asp Phe Asp Thr Tyr Lys Lys Leu Tyr Lys Gln Ser Val
        35                  40                  45

Glu Asn Pro Asn Glu Phe Phe Thr Glu Gln Ala Asn Lys Asn Leu Asp
    50                  55                  60

Trp Phe Lys Pro Phe Asp Leu Ala Arg Phe Pro Val Asp Pro Lys Asp
65                  70                  75                  80

Asp Phe Lys Asn Gly Asp Leu Pro Ala Trp Phe Ile Asn Gly Gln Leu
                85                  90                  95

Asn Ala Ser Tyr Asn Ala Val Asp Arg Trp Ala Ile Lys Asn Pro Asp
            100                 105                 110

Lys Pro Ala Ile Ile Tyr Glu Gly Asp Glu Pro Asp Ser Gly Arg Ile
        115                 120                 125

Ile Thr Tyr Gly Glu Leu Leu Lys Glu Val Ser Lys Leu Ala Gln Thr
    130                 135                 140

Leu Thr Lys Leu Gly Val Lys Lys Gly Asp Ser Val Ala Val Tyr Leu
145                 150                 155                 160

Pro Met Ile Pro Glu Ala Ile Val Thr Leu Leu Ala Ile Val Arg Ile
                165                 170                 175

Gly Ala Val His Ser Val Val Phe Ala Gly Phe Ser Ser Ala Ser Leu
            180                 185                 190

Arg Asp Arg Ile Leu Asp Ala Asp Ser Arg Ile Val Ile Thr Ala Asp
        195                 200                 205

Glu Ser Lys Arg Gly Gly Lys Thr Ile Glu Thr Lys Lys Ile Val Asp
    210                 215                 220

Asp Ala Leu Lys Glu Cys Pro Lys Val Arg Asn Val Ile Val Phe Lys
225                 230                 235                 240

Arg Thr Gly Asn Ser His Val Pro Phe Ser Ala Gly Arg Asp Leu Trp
                245                 250                 255
```

```
Trp His Glu Glu Met Ala Lys Tyr Gly Pro Tyr Phe Pro Pro Val Pro
            260                 265                 270

Val Asn Ser Glu Asp Pro Leu Phe Leu Leu Tyr Thr Ser Gly Ser Thr
        275                 280                 285

Gly Lys Pro Lys Gly Val Gln His Asn Thr Ala Gly Tyr Leu Leu Gly
    290                 295                 300

Ala Ile Leu Thr Thr Lys Tyr Thr Phe Asp Val His Glu Glu Asp Val
305                 310                 315                 320

Leu Phe Thr Ala Gly Asp Val Gly Trp Ile Thr Gly His Thr Tyr Cys
                325                 330                 335

Val Tyr Gly Pro Leu Leu Cys Gly Ala Thr Thr Val Val Phe Glu Gly
            340                 345                 350

Thr Pro Ala Tyr Pro Asn Tyr Ser Arg Tyr Trp Glu Ile Val Asp Gln
        355                 360                 365

Tyr Lys Val Asn Gln Phe Tyr Val Ala Pro Thr Ala Leu Arg Leu Leu
    370                 375                 380

Lys Arg Ala Gly Thr Lys Tyr Val Glu Lys Tyr Asp Leu Ser Ser Leu
385                 390                 395                 400

Arg Val Leu Gly Ser Val Gly Glu Pro Ile Ala Ala Glu Val Trp His
                405                 410                 415

Trp Tyr Asn Asp Asn Ile Gly Arg Gly Lys Ala His Ile Val Asp Thr
            420                 425                 430

Tyr Trp Gln Thr Glu Ser Gly Ser His Leu Leu Thr Pro Leu Ala Gly
        435                 440                 445

Val Thr Pro Thr Lys Pro Gly Ser Ala Ser Leu Pro Phe Phe Gly Val
    450                 455                 460

Val Pro Lys Ile Leu Asp Pro Thr Thr Gly Glu Glu Leu Glu Gly Asn
465                 470                 475                 480

Asp Val Glu Gly Val Leu Ala Ile Lys Ser Ala Trp Pro Ser Ile Thr
                485                 490                 495

Arg Gly Ile Tyr Asn Asp Tyr Asn Arg Phe Ile Glu Thr Tyr Leu Ala
```

500 505 510

Pro Tyr His Asp His Tyr Phe Ser Gly Asp Gly Ala Ala Arg Asp Asn
515 520 525

Asp Gly Phe Tyr Trp Ile Leu Gly Arg Val Asp Asp Val Val Asn Val
530 535 540

Ser Gly His Arg Leu Ser Thr Ala Glu Ile Glu Ala Ala Leu Ile Glu
545 550 555 560

His Pro Leu Val Ala Glu Ser Ala Val Val Gly Tyr Ala Asp Asp Leu
565 570 575

Thr Gly Gln Ala Val Ala Ala Tyr Val Ser Leu Lys Lys Asp Lys Val
580 585 590

Ile Gly Asp Asp Met Glu Ala Val Lys Lys Glu Leu Ile Leu Thr Val
595 600 605

Arg Lys Glu Ile Gly Pro Phe Ala Ala Pro Lys Leu Ile Leu Leu Val
610 615 620

Asp Asp Leu Pro Lys Thr Arg Ser Gly Lys Ile Met Arg Arg Ile Leu
625 630 635 640

Arg Lys Val Leu Ala Gly Glu Glu Asp Gln Leu Gly Asp Ile Ser Thr
645 650 655

Leu Ser Asn Pro Gly Val Val Gln Gln Ile Ile Glu Ile Val His Ala
660 665 670

Cys Lys Lys
675

<210> 21
<211> 671
<212> PRT
<213> Candida viswanathii

<400> 21
Met Thr Thr Asp Ser Asn Thr His Lys Val Val His Glu Ala Asn Gly
1 5 10 15

Val Lys Leu Arg Gln Thr Pro Lys Glu Phe Phe Gln Arg Gln Pro Asn
20 25 30

Lys Gly His Ile Gln Asp Thr Glu Glu Tyr Lys Lys Leu Tyr Glu Gln
35 40 45

```
Ser Ile Lys Asp Pro Gln Gly Phe Phe Gly Pro Leu Ala Lys Glu Leu
    50                  55                  60

Leu Ser Trp Asp Ser Asp Phe His Thr Val Lys Ser Gly Thr Leu Lys
65                  70                  75                  80

Asn Gly Asp Ala Ala Trp Phe Leu Gly Gly Lys Leu Asn Ala Ser Tyr
                85                  90                  95

Asn Cys Val Ala Arg His Ala Leu Ala Asn Pro Asn Lys Pro Ala Ile
            100                 105                 110

Ile Tyr Glu Ala Asp Glu Glu Arg Asp Ser Tyr Ile Leu Thr Tyr Gly
        115                 120                 125

Asp Leu Leu Arg Glu Val Ser Lys Val Ala Gly Val Leu His Ser Trp
    130                 135                 140

Gly Ile Arg Lys Gly Asp Thr Val Ala Val Tyr Leu Pro Met Asn Ala
145                 150                 155                 160

Gln Ala Leu Ile Ala Met Leu Ala Ile Ala Arg Leu Gly Ala Ala His
                165                 170                 175

Ser Val Ile Phe Ala Gly Phe Ser Ser Gly Ser Ile Lys Asp Arg Val
            180                 185                 190

Asn Asp Ala Ser Cys Lys Ala Leu Ile Thr Cys Asp Glu Gly Arg Arg
        195                 200                 205

Gly Gly Arg Thr Thr Asn Ile Lys Lys Leu Cys Asp Glu Ala Leu Lys
    210                 215                 220

Gln Cys Pro Thr Val Lys His Val Leu Val His Arg Arg Thr Gly Asn
225                 230                 235                 240

Pro Glu Ile Lys Leu Thr Glu Gly Arg Asp Tyr Tyr Trp Asp Glu Glu
                245                 250                 255

Thr Ala Lys Phe Ser Gly Tyr Phe Pro Pro Val Pro Val Asp Ser Glu
            260                 265                 270

Asp Pro Leu Phe Leu Leu Tyr Thr Ser Gly Ser Thr Gly Thr Pro Lys
        275                 280                 285

Gly Val Val His Thr Thr Ala Gly Tyr Leu Leu Gly Ala Ala Leu Ser
```

290 295 300

Thr Lys Tyr Ile Phe Asp Val His Pro Glu Asp Val Phe Phe Thr Ala
305 310 315 320

Gly Asp Val Gly Trp Ile Thr Gly His Thr Tyr Ala Leu Tyr Gly Pro
325 330 335

Leu Leu Leu Gly Val Pro Ser Val Val Phe Glu Gly Thr Pro Ala Tyr
340 345 350

Pro Asp Tyr Gly Arg Phe Trp Gln Ile Val Glu Lys His Lys Ala Thr
355 360 365

His Phe Tyr Val Ala Pro Thr Ala Leu Arg Leu Leu Arg Lys Ala Gly
370 375 380

Glu Gln Glu Val Gly Lys Tyr Asp Leu Ser Ser Leu Arg Thr Leu Gly
385 390 395 400

Ser Val Gly Glu Pro Ile Ser Pro Asp Ile Trp Glu Trp Tyr Asn Glu
405 410 415

Phe Val Gly Lys Asp Gln Cys His Ile Ser Asp Thr Tyr Trp Gln Thr
420 425 430

Glu Ser Gly Ser His Leu Ile Ala Pro Leu Ala Gly Ala Ile Ala Asn
435 440 445

Lys Pro Gly Ser Ala Ser Tyr Pro Phe Phe Gly Ile Glu Ala Ala Leu
450 455 460

Ile Asp Pro Val Thr Gly Val Glu Ile Lys Gly Asn Asp Val Glu Gly
465 470 475 480

Val Leu Val Val Lys Asp His Trp Pro Ser Met Ala Arg Thr Val Phe
485 490 495

Asn Asn His Val Lys Tyr Met Asp Thr Tyr Met Asn Pro Tyr Pro Gly
500 505 510

Tyr Tyr Phe Thr Gly Asp Gly Ala Ala Arg Asp Asn Asp Gly Tyr Tyr
515 520 525

Trp Ile Arg Gly Arg Val Asp Asp Val Val Asn Val Ser Gly His Arg
530 535 540

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 112 of 3077

```
Leu Ser Thr Ala Glu Ile Glu Ser Ala Leu Ile Glu Asp Pro Arg Val
545                 550                 555                 560

Gly Glu Ser Ala Val Val Gly Ile Asn Asp Asp Leu Thr Gly Gln Ala
                565                 570                 575

Val Ile Ala Tyr Val Ala Leu Lys Ser Gly Asp Val Ala Asp Glu Asp
            580                 585                 590

Ala Leu Arg Lys Glu Leu Ile Leu Leu Val Arg Lys Glu Ile Gly Pro
        595                 600                 605

Phe Ala Ala Pro Lys Ser Val Ile Leu Val Gln Asp Leu Pro Lys Thr
    610                 615                 620

Arg Ser Gly Lys Ile Met Arg Arg Ile Leu Arg Lys Val Ser Ser Asn
625                 630                 635                 640

Glu Ala Asp Gln Leu Gly Asp Ile Thr Thr Leu Gln Asn Pro Gly Ser
                645                 650                 655

Val Glu Gly Met Ile Ser Ala Phe Gly Ala Gln Phe Gly Arg Lys
            660                 665                 670


<210> 22
<211> 383
<212> PRT
<213> Candida viswanathii

<400> 22
Met Leu Thr Leu Thr Ser Gly Pro Asn Pro Leu Pro Asp Phe Glu Glu
1               5                   10                  15

Ala Leu Arg Val Ile Lys Val Asp Asp Thr His Tyr Val Gly Ala His
            20                  25                  30

Ser Leu Arg Leu Pro Val Lys Gly Gly Arg Gly Val Tyr Gly Gly His
        35                  40                  45

Met Ile Ala Gln Ser Leu Leu Val Gly Ile Glu Ser Thr Arg Asp Asp
    50                  55                  60

Lys Thr Asn Lys Val Phe Ile Pro Asp Ser Tyr His Ser Tyr Phe Ile
65                  70                  75                  80

Gly Ala Gly Asn Ala Lys Ile Pro Met Asn Tyr Thr Val Glu Lys Leu
                85                  90                  95

Tyr Asp Asp Glu Asn Val Ser Lys Arg Phe Ile Ile Ala Glu Gln Lys
```

```
            100                 105                 110

Gly Arg His Arg Leu Thr Cys Leu Val Thr Leu Arg Arg Pro Gly Thr
        115                 120                 125

Lys Pro Phe His Asp Ser Asp Asn Leu Asp Ile Ser Ile Pro Val Pro
    130                 135                 140

Lys Ile Gln Leu Lys His Pro Asp Pro Asp Lys Leu His Gln Val Gln
145                 150                 155                 160

His Thr Asp Phe Ile Arg Asn Ala Phe Gly Lys Glu Leu Met Asp Tyr
                165                 170                 175

Arg Glu Cys Pro Glu Glu Asn Glu Leu Tyr Ala Ala Glu Arg Trp Leu
            180                 185                 190

Thr Val Phe Thr Gly Ile Arg Asn Gln Pro Lys Pro Gly Ala Ser Leu
        195                 200                 205

Glu Thr Val Val Glu Glu Leu Pro Asp Ala Gln Gly Gln Met His Thr
    210                 215                 220

Val Glu Lys Ser Ile Leu Arg Pro Lys Asp Ser Gln Ser Phe Lys Asp
225                 230                 235                 240

Pro Ile Tyr Asn Phe Val Gly Leu Ala Asp Leu Ser Asp Ser Ala Phe
                245                 250                 255

Leu Thr Thr Met Ala Arg Ile Leu His Ile Pro Trp Ala Pro Ser Ile
            260                 265                 270

Glu Ile Asp Asp Thr Tyr Asp Pro Ala Arg Asp Ala Thr Tyr Ile Met
        275                 280                 285

Arg Ser Thr Leu Asn Ala Ala His Ile Phe His Tyr Asn Ala Met Ser
    290                 295                 300

Leu Asp His His Ile Tyr Phe His Asn Glu Asp Tyr Thr Ser Asp Asp
305                 310                 315                 320

Gly Ser Gly Phe Asp Ile Cys Lys Asp Trp Leu Ala Phe Thr Tyr Gln
                325                 330                 335

Met Lys Arg Leu Ser Asn Asn Arg Thr Leu Val Arg Gly Phe Leu Phe
            340                 345                 350
```

Asn Glu Lys His Lys Cys Ile Ala Thr Val Val Gln Glu Gly Leu Thr
355 360 365

Ile Met Gln Asn Gly Val Gly Arg Thr Ala Asp Lys Ser Arg Leu
370 375 380

<210> 23
<211> 326
<212> PRT
<213> Candida viswanathii

<400> 23
Met Ile Glu Asn Ile Ser Gly Asn Gly Asn Tyr Pro Gln Asn His Glu
1 5 10 15

Val Asp Leu Glu Lys Glu Phe Gly Val Glu Lys Ile Gly Ile Asn Leu
20 25 30

Tyr Arg Gly Lys Ser Pro Ile Pro Lys Pro Asp Arg Arg Ser Arg Gly
35 40 45

Ala Tyr Gly Gly Tyr Leu Ala Gly Gln Ala Leu Leu Val Ala Met Lys
50 55 60

Ser Thr Pro Pro Glu Tyr Arg Pro His Ser Phe His Ser Tyr Phe Ile
65 70 75 80

Lys Ala Val Asn Asp Lys Glu Thr Leu Glu Trp Arg Val Glu Glu Thr
85 90 95

Ser Asn Gly Arg Asn Tyr Ala Asn Arg Ser Leu Gln Ala Phe Gln Ala
100 105 110

Gly Asn Leu Val Tyr Thr Ala Asn Val Ser Leu Thr Lys Lys Asn Ser
115 120 125

Ala Lys Lys Ala Glu Glu Ala Thr Gly Val Lys Pro Phe Glu Phe Gln
130 135 140

Gly Lys Pro His Glu Trp Phe Glu Lys His Lys Arg Asp Asp Leu Pro
145 150 155 160

Leu Ala Thr Pro Ser Ser Ser Leu Leu Ile Tyr His Lys Phe Phe Pro
165 170 175

Glu Val Val Ser Leu Glu Ala Ser Lys Glu Glu Glu Ser Lys Pro Ala
180 185 190

Ala Asp Arg Glu Leu Ser Trp Tyr Phe Lys Trp Gly Ile Asn Asn Glu

195 200 205

Glu Gly His His Gln Pro Leu Val Asn Leu Asn Ser Asp Tyr Gln Tyr
210 215 220

Val Gly Met Ala Ala Leu Thr Asp Ala Val Tyr Leu Asn Arg Leu Leu
225 230 235 240

Arg Ile Leu Arg Val Glu Asp Ala Asp His Thr Gln Leu Val His Tyr
245 250 255

Phe Ser Val Ser Leu Asp His Thr Met Tyr Phe His Asp Asp Asp Phe
260 265 270

Asp Val Thr Lys Trp Met Gly Phe Thr Phe Lys Val Thr Arg Phe Ser
275 280 285

His Asn Arg Ala Leu Cys Gln Gly Glu Val Tyr Asn Asp Lys Gly Val
290 295 300

His Val Cys Thr Ile Val Gln Glu Gly Leu Met Met Leu Asn Gly Leu
305 310 315 320

Glu Gln Gly Ala Lys Leu
325

<210> 24
<211> 326
<212> PRT
<213> Candida viswanathii

<400> 24
Met Ile Glu Ser Ile Ser Glu Asn Gly Asn Tyr Pro Gln Asn His Glu
1 5 10 15

Val Asp Leu Glu Lys Glu Phe Gly Val Glu Lys Ile Gly Ile Asn Leu
20 25 30

Tyr Arg Gly Lys Ser Pro Ile Pro Lys Pro Asp Arg Arg Ser Arg Gly
35 40 45

Ala Tyr Gly Gly Tyr Leu Ala Gly Gln Ala Leu Leu Val Ala Met Lys
50 55 60

Ser Thr Pro Pro Asp Tyr Arg Pro His Ser Phe His Ser Tyr Phe Ile
65 70 75 80

Lys Ala Val Asn Asp Lys Val Thr Leu Glu Trp Arg Val Glu Glu Thr
85 90 95

```
Ser Asn Gly Arg Asn Tyr Ala Asn Arg Ala Leu Gln Ala Phe Gln Ala
            100                 105                 110

Gly Asp Leu Val Tyr Thr Ala Asn Val Ser Leu Thr Lys Lys Asn Ser
        115                 120                 125

Ala Lys Lys Thr Glu Glu Ala Thr Gly Val Lys Pro Phe Glu Phe Gln
    130                 135                 140

Gly Lys Pro His Glu Trp Phe Asp Lys His Lys Ile Asp Asp Leu Pro
145                 150                 155                 160

Leu Ala Thr Pro Ser Ser Ser Leu Leu Ile Tyr His Lys Phe Phe Pro
                165                 170                 175

Glu Val Val Ser Leu Glu Ala Ser Lys Glu Glu Glu Ser Lys Pro Ala
            180                 185                 190

Ala Asp Arg Glu Leu Ser Trp Tyr Phe Lys Trp Gly Ile Asn Asn Glu
        195                 200                 205

Glu Gly His His Gln Pro Leu Val Asn Leu Asn Ser Asp Tyr Gln Tyr
    210                 215                 220

Val Gly Met Ala Ala Leu Thr Asp Ala Val Tyr Leu Asn Arg Leu Leu
225                 230                 235                 240

Arg Ile Leu Arg Val Gln Asp Ala Asp His Thr Gln Leu Val His Tyr
                245                 250                 255

Phe Ser Val Ser Leu Asp His Thr Met Tyr Phe His Asp Asp Asp Phe
            260                 265                 270

Asp Val Thr Lys Trp Met Gly Phe Thr Phe Lys Val Thr Arg Phe Ser
        275                 280                 285

His Asn Arg Ala Leu Cys Gln Gly Glu Val Tyr Asn Asp Lys Gly Val
    290                 295                 300

His Val Cys Thr Ile Val Gln Glu Gly Leu Met Met Leu Asn Gly Leu
305                 310                 315                 320

Glu Glu Gly Ala Lys Leu
                325
```

<210> 25

<211> 279
<212> PRT
<213> Candida viswanathii

<400> 25

```
Met Ser His Pro Thr Pro Glu Glu Val Tyr Gly Val Thr Lys Val Ala
1               5                  10                  15

Glu Asn Lys Tyr Val Gly Asn Arg Pro Leu Asn Lys Pro Thr Pro Lys
            20                  25                  30

Thr Arg Gly Val Tyr Gly Gly Asn Phe Cys Ala Gln Ala Ile Leu Val
        35                  40                  45

Ala Ile Glu Ser Ala Pro Glu Gly Phe Thr Pro His Ser Ile His Ser
    50                  55                  60

Asn Phe Ile Arg Gly Gly Asp Pro Glu Val Pro Val Glu Trp Glu Val
65                  70                  75                  80

Glu Val Ile Ser Asn Gly Lys Ser Phe Ala Asn Arg Ile Val Lys Gly
                85                  90                  95

Val Gln His Gly Ile Val Val Tyr Val Ala Thr Val Ser Leu Thr Asn
            100                 105                 110

Lys Asn Ser Thr Thr Arg Asn Glu Ser Phe Thr Tyr Asp Thr Pro Pro
        115                 120                 125

Asp Glu Thr Val Lys Thr Tyr Gly Asn Ala Glu Leu Asp Thr Tyr Tyr
    130                 135                 140

Gln Gly Trp Leu Tyr Leu Glu Val Lys Asn Tyr Pro Lys Gln Leu His
145                 150                 155                 160

Ser His Gln Ile Ser Tyr Ser Val Lys Trp Gly Pro Glu Asn Asp Ala
                165                 170                 175

Trp Lys Asp Ala Ser Gln Thr Tyr Gln Phe Val Gly Leu Ala Ala Ile
            180                 185                 190

Ser Asp Val Leu Asp Leu Gly Gln Ile Leu Arg Asn Leu Asp Ile His
        195                 200                 205

Leu Ser Thr Pro Lys Phe Asn Val Ser Leu Asp His Ser Val Tyr Phe
    210                 215                 220

His Gly Ala Asp Phe Asp Val Thr Arg Trp Ser Thr Thr Thr Ile Arg
225                 230                 235                 240
```

```
Met Thr Lys Leu Ala His Gly Arg Ala Leu Ile Glu Gly Glu Met Tyr
                245                 250                 255

Ser Asp Gln Gly Arg His Ile Ala Ser Ile Val Gln Glu Arg Leu Tyr
            260                 265                 270

Ile Ala Glu Ser Pro Lys Leu
        275
```

<210> 26
<211> 298
<212> PRT
<213> Candida viswanathii

<400> 26

```
Met Thr Ser Ile Ala Pro Ala Thr Asp Ile Pro Ala Val Asp Ile Ser
1               5                   10                  15

Thr Ile Tyr Asp Val Lys Gln Ile Asp Ala Asn Arg Tyr Arg Gly Val
            20                  25                  30

Arg Pro Leu Gln Lys Gln Ala Arg Glu Phe Arg Gly Val Phe Gly Gly
        35                  40                  45

Asn Leu Val Ala Gln Ser Val Val Val Ala Leu Arg Ser Val Pro Ala
    50                  55                  60

Gly Phe His Pro Asn Ser Val His Ala Tyr Tyr Val Arg Ala Val Thr
65                  70                  75                  80

Asp Glu Thr Pro Ile Glu Trp Glu Val Glu Glu Thr Thr Thr Gly Arg
                85                  90                  95

Thr Phe Ala Asn Arg Ser Ile Arg Gly Leu Gln Asn Lys Lys Val Val
            100                 105                 110

Phe Thr Ala Ser Ile Ser Leu Thr Arg Lys Asn Ser Asn Ala Glu Val
        115                 120                 125

Ile Gln Lys Thr Gly His Pro Ser Leu Gln Phe Gln Arg Asp Thr Glu
    130                 135                 140

Pro Tyr Tyr Glu Gln Met Arg Ala Arg Pro Gly Glu Cys Lys Val Leu
145                 150                 155                 160

Tyr Ile Asn Ala Asn Thr His Leu Thr Val Arg Gln Phe Pro Glu Val
                165                 170                 175
```


LCY Biotechnology Holding, Inc.
Ex. 1002
Page 119 of 3077

```
Ser Ser Arg Asp Val Phe Ser Phe Leu Val Arg Phe Gly Asn Asp Gly
            180                 185                 190

Arg Glu Arg Ile Val Gly Met Thr Pro Glu Tyr Gln Tyr Ala Ala Leu
        195                 200                 205

Ala Ala Leu Ser Asp Trp Val Arg Leu Arg Phe Tyr Phe Asp Asn Met
    210                 215                 220

Gly Val Asp Val Gln Thr Ser Phe Asp Val Ser Leu Asp His Ser Ile
225                 230                 235                 240

Tyr Phe His Asp Asp Gly Phe Asp Ala Thr Glu Tyr Leu Val Phe Ser
                245                 250                 255

Val Lys Val Ser Arg Ile Ser His Ser Arg Val Leu Tyr Asp Gly Gln
            260                 265                 270

Ile Phe Asn Asp Arg Gly Ala His Val Ala Ser Ile Arg Gln Glu Arg
        275                 280                 285

Leu Tyr Val Val Ser Asn Lys Pro Lys Phe
    290                 295

<210> 27
<211> 353
<212> PRT
<213> Candida viswanathii

<400> 27
Met Pro Thr Phe Asn Tyr Lys Asp Gly Glu Thr Ile Asp Val Gln Lys
1               5                   10                  15

Glu Phe Gly Val Val Glu Thr Ala Pro Asn Lys Tyr Val Gly Val Lys
            20                  25                  30

Pro Leu Val Lys Pro Met Pro His Val Lys Gly Val Phe Gly Gly Asn
        35                  40                  45

Leu Ala Gly Gln Ala Leu Leu Val Ala Met Lys Ser Val Gly Pro Asp
    50                  55                  60

Phe Ser Pro His Ser Leu His Ser Tyr Phe Ile Arg Ala Gly Ser Asp
65                  70                  75                  80

Gln Thr Pro Val Glu Trp Thr Val Gln Ala Ile Ser Asp Gly Asn Ser
                85                  90                  95
```

```
Phe Cys Asn Arg Phe Ile Lys Gly Val Gln Asn Gly Gln Val Ile Tyr
            100                 105                 110

Ile Ala Asn Val Ser Leu Thr Lys Arg Asn Ser Ala Ala Asp Ala Met
        115                 120                 125

Lys Lys Tyr Glu Glu Tyr His Ala Gln Ile Arg Gln Lys Gly Lys Asp
    130                 135                 140

Gly Asp Ala Asp Glu Glu Asp Glu Asp Asp Asp Asp Glu Asp Asp Asn
145                 150                 155                 160

Ala Pro Ala Lys Pro Phe Gly Phe Gln Thr Pro Ser His Lys Trp Ile
                165                 170                 175

Lys Asp Arg Asp Leu Asp Lys Leu Pro Val Ser Asp Met Glu Ser Asn
            180                 185                 190

Leu Leu Leu Tyr Tyr Lys Leu Pro Pro Glu Phe Val Ser Leu Lys Ser
        195                 200                 205

Ser Thr Glu Glu Glu Ser Leu Pro Val Ser Glu Arg Arg Met Gly Ala
    210                 215                 220

Leu Ala Lys Trp Gly Ile Glu Asn Glu Gln Gly Phe Asn Gln Pro Leu
225                 230                 235                 240

Thr Asn Leu Asp Lys Ser Phe Gln Tyr Val Gly Leu Ala Asn Ile Thr
                245                 250                 255

Asp Gly Leu Tyr Leu Gly Thr Leu Asn Arg Ile Leu Arg Ile Asp Asp
            260                 265                 270

Leu Thr Leu Asp Glu Arg Ala Thr Asn Tyr Phe Ser Val Ser Leu Asp
        275                 280                 285

His Val Ile Tyr Phe His Asp Asp Asp Phe Asp Val Thr Lys Trp Met
    290                 295                 300

Gly Phe Thr Phe Arg Cys Ser Arg Tyr Ser His Asn Arg Val Ile Phe
305                 310                 315                 320

Glu Gly Glu Ile Tyr Ser Asp Lys Gly Val Gln Val Ala Ser Ile Ile
                325                 330                 335

Gln Glu Gly Leu Val Arg Phe Lys Asp Gly Tyr Leu Lys Asn Ala Lys
            340                 345                 350
```

Leu

<210> 28
<211> 349
<212> PRT
<213> Candida viswanathii

<400> 28
Met Ser Thr Thr Gly Ile Tyr Asn Ala Lys Pro Ala Ala Ala Ile Leu
1 5 10 15

Glu Thr Lys Thr Lys Val Lys Leu Val Tyr Thr Glu Gly Pro Lys Leu
20 25 30

Val Tyr Glu Gly Leu His Pro Val Glu Thr Ile Lys Glu Thr Val Arg
35 40 45

Gly Thr Tyr Gly Gly Asp Phe Ile Ala Gln Gly Val Asn Ala Ala Trp
50 55 60

Glu Ser Ile Gly Asn Arg Met Asp Phe Gln Pro His Ser Leu His Ala
65 70 75 80

Tyr Tyr Ile Lys Ser Gly Ser Gln Glu Ser Val Leu Arg Trp Glu Val
85 90 95

Phe Lys Val Ser Asp Ser Arg Ser Phe Ala Asn Arg Met Leu Thr Ala
100 105 110

Tyr Gln Thr His Thr Asn Gln Leu Val Phe Thr Met Gln Ile Ser Phe
115 120 125

Thr Lys Asp Asn Asn Glu Glu Ile Lys Arg Ala Glu Tyr Glu Lys Leu
130 135 140

Leu Ala Ser Gly Gly Lys Ile Arg Ser Ile Pro Phe Val Ile Lys Lys
145 150 155 160

Ala Pro Asn Glu Lys Tyr Phe Lys Leu Arg Asp Arg Ile Asn Glu Leu
165 170 175

His Tyr Ser Glu Tyr Thr Asn Gly Ile Ile Ala Met Ala Ser Pro Leu
180 185 190

Asp Phe His Glu Tyr Ala Thr Glu Met Asn Tyr Asp Thr Leu Gly Asn
195 200 205

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 122 of 3077

```
Lys Glu Tyr Gly Thr Phe Met Lys Ile Leu Asp Asn Tyr Ser Leu Gly
    210                 215                 220

Lys Asp Tyr Ser Lys Gln Ser Phe Leu Gly Leu Ser Phe Leu Ser Asp
225                 230                 235                 240

Gly Val Trp Val Phe Cys Phe Thr Arg Ala Leu Gly Leu Pro Leu Gly
                245                 250                 255

Ser Tyr Glu Lys Glu Phe Phe Arg Val Ser Leu Asp His Thr Val Tyr
            260                 265                 270

Phe His Asp Ala Asn Phe Asp Ser Ser Glu Trp Leu Phe Leu Asp Phe
        275                 280                 285

Arg Phe Val Asn Met Lys Asn Asn Arg Phe Leu Thr Val Leu Asn Tyr
    290                 295                 300

Tyr Thr Leu Gln Gly Lys Leu Ile Ala Thr Val Leu Gln Glu Val Tyr
305                 310                 315                 320

Thr Ser Leu His Gln Gly Ile Ile Asp Lys Ser Gln Glu Ile Ala Val
                325                 330                 335

Lys Ser Gly Asn Lys Lys Gln Val Thr Thr Pro Lys Leu
            340                 345


<210> 29
<211> 351
<212> PRT
<213> Candida viswanathii

<400> 29
Met Asp Lys Leu Gln Ala Glu Val Tyr Glu Ala Glu Pro Pro Val Ala
1               5                   10                  15

Lys Leu Glu Ala Lys Thr Ala Ala Lys Leu Ile Ser Ser Asp Gly Ala
            20                  25                  30

Lys Leu Thr Tyr Glu Gly Val Tyr Pro Val Glu Leu Val Arg Lys Gly
        35                  40                  45

Leu Arg Gly Thr Tyr Gly Gly Asp Phe Ile Ala Gln Gly Ile Asn Val
    50                  55                  60

Ala Trp Glu Ser Ile Gly Asn Lys Thr Asp Phe Gln Pro His Ser Leu
65                  70                  75                  80

His Ala Tyr Phe Val Lys Ala Gly Ser Asp Ser Ser Val Leu Arg Trp
```

85 90 95

Glu Val Leu Lys Val Ser Asp Ser Arg Asn Phe Ala Asn Arg Leu Met
100 105 110

Leu Ala Tyr Gln Thr His Thr Asn Glu Leu Val Phe Thr Met Gln Ile
115 120 125

Ser Phe Thr Lys Asp Asn Asn Glu Glu Ile Lys Arg Ala Glu Tyr Lys
130 135 140

Gln Leu Leu Gln Ser Gly Gly Lys Ile Arg Ser Ile Pro Phe Ala Ile
145 150 155 160

Lys Lys Pro Pro Asn Glu Lys Tyr Phe Lys Leu Lys Asp Lys Val Asp
165 170 175

Asp Leu Pro Tyr Phe Glu His Thr Asn Gly Asn Met Ala Thr Ala Ile
180 185 190

Pro Pro Asp Phe Leu Glu Tyr Ala Thr Glu Met Asn His Asp Thr Val
195 200 205

Gly Asn Lys Glu Phe Gly Ile Phe Met Lys Val Leu Asp Asp Tyr Ser
210 215 220

Leu Gly Lys Asn Tyr Glu Arg Gln Ser Phe Leu Gly Leu Ala Phe Leu
225 230 235 240

Ser Asp Ala Val Trp Leu Ser Ser Phe Thr Pro Ala Leu Gly Leu Pro
245 250 255

Leu Gly Thr Leu Glu Arg Lys Phe Phe Arg Val Ser Leu Asp His Thr
260 265 270

Met Tyr Phe His Asp Ala Asn Phe Asp Ser Ser Glu Trp Ile Phe Val
275 280 285

Asp Phe Arg Phe Val Asn Leu Asn Asn Asn Arg Leu Leu Gly Val Val
290 295 300

Asn Phe Tyr Thr Leu Gln Gly Lys Leu Val Ala Thr Val Ile Gln Glu
305 310 315 320

Ala Tyr Met Phe Leu His Gln Ala Ile Ile Asp Lys Ser Gln Glu Ile
325 330 335

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 124 of 3077

Ala Glu Lys Ser Gly His Lys Lys Gln Val Ile Thr Pro Lys Leu
340 345 350

<210> 30
<211> 351
<212> PRT
<213> Candida viswanathii

<400> 30
Met Asp Lys Leu Gln Ala Glu Val Tyr Asp Ala Glu Pro Pro Val Ala
1 5 10 15

Lys Leu Glu Ala Lys Thr Ala Ala Lys Leu Ile Ser Ala Asp Gly Ala
20 25 30

Lys Leu Ile Tyr Glu Gly Val Tyr Pro Val Glu Leu Val Arg Lys Gly
35 40 45

Leu Arg Gly Thr Tyr Gly Gly Asp Phe Ile Ala Gln Gly Ile Asn Val
50 55 60

Ala Trp Glu Ser Ile Gly Asn Lys Thr Asp Phe Gln Pro His Ser Leu
65 70 75 80

His Ala Tyr Phe Val Lys Ala Gly Ser Asp Gln Ser Val Leu Arg Trp
85 90 95

Glu Val Leu Lys Val Ser Asp Ser Arg Ser Phe Ala Asn Arg Leu Met
100 105 110

Leu Ala Tyr Gln Thr His Thr Asn Gln Leu Val Phe Thr Met Gln Ile
115 120 125

Ser Phe Thr Lys Asp Asn Asn Glu Glu Ile Lys Arg Ala Glu Tyr Lys
130 135 140

Gln Leu Leu Glu Ser Gly Gly Lys Ile Arg Ser Ile Pro Phe Ala Ile
145 150 155 160

Lys Lys Pro Pro Asn Glu Lys Tyr Phe Lys Leu Lys Asp Lys Val Asp
165 170 175

Asp Leu Pro Tyr Phe Glu His Thr Asn Gly Asn Met Ala Thr Ala Ile
180 185 190

Pro Pro Asp Phe Leu Glu Tyr Ala Thr Glu Met Asn His Asp Thr Val
195 200 205

Gly Asn Lys Glu Phe Gly Ile Phe Met Lys Val Leu Asp Asp Tyr Ser

210 215 220

Leu Asp Lys Asn Tyr Glu Arg Gln Ser Phe Leu Gly Leu Ala Phe Leu
225 230 235 240

Ser Asp Ala Val Trp Leu Ser Ser Phe Thr Pro Ala Leu Gly Leu Pro
245 250 255

Leu Gly Thr Phe Glu Arg Lys Phe Phe Arg Val Ser Leu Asp His Thr
260 265 270

Met Tyr Phe His Asp Ala Asn Phe Asp Ser Ser Glu Trp Ile Phe Val
275 280 285

Asp Phe Arg Phe Val Asn Leu Asn Asn Asn Arg Leu Leu Gly Val Val
290 295 300

Asn Phe Tyr Thr Leu Gln Gly Lys Leu Val Ala Thr Val Ile Gln Glu
305 310 315 320

Ala Tyr Met Phe Leu His Gln Gly Ile Ile Asp Gln Ser Gln Glu Ile
325 330 335

Ala Glu Lys Ser Gly His Lys Lys Gln Val Ile Thr Pro Lys Leu
340 345 350

<210> 31
<211> 351
<212> PRT
<213> Candida viswanathii

<400> 31
Met Glu Arg Leu Gln Ala Glu Val Tyr Asp Ala Lys Pro Thr Val Ala
1 5 10 15

Lys Leu Glu Thr Lys Thr Thr Val Lys Leu Val Ala Gln Asp Gly Lys
20 25 30

Arg Leu Val Tyr Glu Ala Ile Tyr Pro Val Glu His Ile Lys Glu Gly
35 40 45

Ile Pro Gly Ala Tyr Gly Gly Asp Thr Leu Ala Gln Gly Met Asn Ala
50 55 60

Ala Trp Asp Ser Leu Gly Asp Lys Lys Asp Phe Gln Pro His Ser Val
65 70 75 80

His Ser Tyr Phe Val Lys Pro Ala Thr Asn Lys Ser Ile Leu Arg Trp
85 90 95

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 126 of 3077

```
Glu Val Ile Lys Val Ser Asp Gly Arg Ser Phe Ala Asn Arg Phe Val
            100                 105                 110

Ser Gly Tyr Gln Thr His Asn Asn Ala Leu Val Phe Thr Met Gln Ile
        115                 120                 125

Ser Phe Thr Lys Tyr Asn Asp Glu Val Val Lys Ile Ala Glu Tyr Asn
    130                 135                 140

Lys Leu Leu Glu Ser Gly Gly Lys Ile Arg Ser Ile Pro Phe Val Ile
145                 150                 155                 160

Lys Lys Ala Pro Asn Glu Lys Tyr Phe Lys Phe Lys Asp Asn Leu Gly
                165                 170                 175

Asp Leu Arg Tyr Ile Glu His Thr Asn Gly Asn Ile Ala Thr Ala Met
            180                 185                 190

Ser Glu Asp Leu Phe Glu Tyr Ala Thr Glu Met Asn His Asp Thr Leu
        195                 200                 205

Gly Asn Gln Glu Phe Gly Ile Phe Met Lys Val Leu Asp Asn Tyr Ser
    210                 215                 220

Leu Gly Ser Asp Tyr Thr Lys Gln Ser Tyr Leu Gly Leu Ala Phe Leu
225                 230                 235                 240

Ser Asp Ala Ile Trp Met Ser Val Cys Ser Arg Ala Leu Gly Leu Pro
                245                 250                 255

Phe Gly Ser Tyr His Arg Gln Phe Phe Gly Val Ser Met Asp His Ser
            260                 265                 270

Met Tyr Phe His Asp Ala Asn Phe Asp Ser Thr Glu Trp Ile Phe Leu
        275                 280                 285

Asp Phe Arg Phe Val Asn Leu Lys Asn Asp Arg Leu Leu Gly Val Ala
    290                 295                 300

Asn Phe Tyr Thr Leu Glu Gly Lys Leu Ile Ser Thr Leu Ile Gln Glu
305                 310                 315                 320

Ala Tyr Leu Phe Leu His Pro Gly Ile Ile Ala Lys Ser Gln Glu Ile
                325                 330                 335

Ala Val Lys Ser Gly Asn Lys Arg Gln Val Lys Met Pro Lys Leu
```

340 345 350

<210> 32
<211> 276
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 32
Met Ser His Pro Thr Pro Glu Glu Val Tyr Gly Val Thr Lys Val Ala
1 5 10 15

Glu Asn Lys Tyr Val Gly Asn Arg Pro Leu Asn Lys Pro Thr Pro Lys
20 25 30

Thr Arg Gly Val Tyr Gly Gly Asn Phe Cys Ala Gln Ala Ile Leu Val
35 40 45

Ala Ile Glu Ser Ala Pro Glu Gly Phe Thr Pro His Ser Ile His Ser
50 55 60

Asn Phe Ile Arg Gly Gly Asp Pro Glu Val Pro Val Glu Trp Glu Val
65 70 75 80

Glu Val Ile Ser Asn Gly Lys Ser Phe Ala Asn Arg Ile Val Lys Gly
85 90 95

Val Gln His Gly Ile Val Val Tyr Val Ala Thr Val Ser Leu Thr Asn
100 105 110

Lys Asn Ser Thr Thr Arg Asn Glu Ser Phe Thr Tyr Asp Thr Pro Pro
115 120 125

Asp Glu Thr Val Lys Thr Tyr Gly Asn Ala Glu Leu Asp Thr Tyr Tyr
130 135 140

Gln Gly Trp Leu Tyr Leu Glu Val Lys Asn Tyr Pro Lys Gln Leu His
145 150 155 160

Ser His Gln Ile Ser Tyr Ser Val Lys Trp Gly Pro Glu Asn Asp Ala
165 170 175

Trp Lys Asp Ala Ser Gln Thr Tyr Gln Phe Val Gly Leu Ala Ala Ile
180 185 190

Ser Asp Val Leu Asp Leu Gly Gln Ile Leu Arg Asn Leu Asp Ile His
195 200 205

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 128 of 3077

```
Leu Ser Thr Pro Lys Phe Asn Val Ser Leu Asp His Ser Val Tyr Phe
    210                 215                 220

His Gly Ala Asp Phe Asp Val Thr Arg Trp Ser Thr Thr Thr Ile Arg
225                 230                 235                 240

Met Thr Lys Leu Ala His Gly Arg Ala Leu Ile Glu Gly Glu Met Tyr
                245                 250                 255

Ser Asp Gln Gly Arg His Ile Ala Ser Ile Val Gln Glu Arg Leu Tyr
            260                 265                 270

Ile Ala Glu Ser
        275


<210> 33
<211> 231
<212> PRT
<213> Candida viswanathii

<400> 33
Met Val Ala Asp Ser Leu Val Tyr His Pro Thr Val Ser Lys Leu Val
1               5                   10                  15

Lys Phe Leu Asp Thr Thr Pro Lys Arg Glu Lys Val Phe Arg Leu Leu
            20                  25                  30

Ser Tyr Leu Ser Arg Phe Leu Gly Tyr Tyr Ala Tyr Arg Lys Gly Tyr
        35                  40                  45

Ser Lys Glu Thr Ile Ala Leu Phe Ala Asn Leu Lys Gly Asn Phe Thr
    50                  55                  60

Phe Ile Arg Lys Ala Met Arg Phe Leu Lys Pro Ile Asn His Leu Gln
65                  70                  75                  80

Leu Ala Ser Lys Ala Tyr Asp Asn Lys Leu Leu Asp Pro Val Leu Gln
                85                  90                  95

Ile Thr Thr Ile Ile Arg Asn Leu Ala Tyr Ala Gly Tyr Leu Thr Ile
            100                 105                 110

Asp Gly Val Ile Phe Phe Lys Leu Leu Gly Leu Ile Asp Ala Lys Lys
        115                 120                 125

Phe Pro Asn Leu Ala Thr Tyr Ala Ser Arg Phe Trp Leu Ile Gly Leu
    130                 135                 140
```

```
Ile Ala Gly Leu Ile Asn Ser Leu Arg Ile Ile Tyr Ser Leu Lys Asp
145                 150                 155                 160

Tyr Glu His Gln Glu Gly Asp Lys Glu Lys Glu Thr Asp Ala Lys Ala
                165                 170                 175

Ile His Thr Lys Leu Tyr Ala Ala Lys Arg Lys Leu Val Trp Asp Leu
            180                 185                 190

Leu Asp Thr Phe Ile Ala Leu Asn Ser Leu Asp Ile Leu His Phe Thr
        195                 200                 205

Glu Gly Asp Val Gly Phe Ala Gly Thr Ile Thr Ser Leu Leu Gly Leu
    210                 215                 220

Glu Asp Leu Trp Lys Ala Thr
225                 230
```

<210> 34
<211> 679
<212> PRT
<213> Candida viswanathii

```
<400> 34
Met Ala Leu Asp Lys Leu Asp Leu Tyr Val Ile Ile Thr Leu Val Val
1               5                   10                  15

Ala Val Ala Ala Tyr Phe Ala Lys Asn Gln Phe Leu Asp Gln Pro Gln
            20                  25                  30

Asp Thr Gly Phe Leu Asn Thr Asp Ser Gly Ser Asn Ser Arg Asp Val
        35                  40                  45

Leu Ser Thr Leu Lys Lys Asn Asn Lys Asn Thr Leu Leu Leu Phe Gly
    50                  55                  60

Ser Gln Thr Gly Thr Ala Glu Asp Tyr Ala Asn Lys Leu Ser Arg Glu
65                  70                  75                  80

Leu His Ser Arg Phe Gly Leu Lys Thr Met Val Ala Asp Phe Ala Asp
                85                  90                  95

Tyr Asp Trp Asp Asn Phe Gly Asp Ile Thr Glu Asp Ile Leu Val Phe
            100                 105                 110

Phe Ile Val Ala Thr Tyr Gly Glu Gly Glu Pro Thr Asp Asn Ala Asp
        115                 120                 125
```

```
Glu Phe His Thr Trp Leu Thr Glu Glu Ala Asp Thr Leu Ser Thr Leu
    130                 135                 140

Arg Tyr Thr Val Phe Gly Leu Gly Asn Ser Thr Tyr Glu Phe Phe Asn
145                 150                 155                 160

Ala Ile Gly Arg Lys Phe Asp Arg Leu Leu Ser Glu Lys Gly Gly Asp
                165                 170                 175

Arg Phe Ala Glu Tyr Ala Glu Gly Asp Asp Gly Thr Gly Thr Leu Asp
            180                 185                 190

Glu Asp Phe Met Ala Trp Lys Asp Asn Val Phe Asp Ala Leu Lys Asn
        195                 200                 205

Asp Leu Asn Phe Glu Glu Lys Glu Leu Lys Tyr Glu Pro Asn Val Lys
    210                 215                 220

Leu Thr Glu Arg Asp Asp Leu Ser Ala Ala Asp Ser Gln Val Ser Leu
225                 230                 235                 240

Gly Glu Pro Asn Lys Lys Tyr Ile Asn Ser Glu Gly Ile Asp Leu Thr
                245                 250                 255

Lys Gly Pro Phe Asp His Thr His Pro Tyr Leu Ala Arg Ile Thr Glu
            260                 265                 270

Thr Arg Glu Leu Phe Ser Ser Lys Glu Arg His Cys Ile His Val Glu
        275                 280                 285

Phe Asp Ile Ser Glu Ser Asn Leu Lys Tyr Thr Thr Gly Asp His Leu
    290                 295                 300

Ala Ile Trp Pro Ser Asn Ser Asp Glu Asn Ile Lys Gln Phe Ala Lys
305                 310                 315                 320

Cys Phe Gly Leu Glu Asp Lys Leu Asp Thr Val Ile Glu Leu Lys Ala
                325                 330                 335

Leu Asp Ser Thr Tyr Thr Ile Pro Phe Pro Thr Pro Ile Thr Tyr Gly
            340                 345                 350

Ala Val Ile Arg His His Leu Glu Ile Ser Gly Pro Val Ser Arg Gln
        355                 360                 365

Phe Phe Leu Ser Ile Ala Gly Phe Ala Pro Asp Glu Glu Thr Lys Lys
    370                 375                 380
```


LCY Biotechnology Holding, Inc.
Ex. 1002
Page 131 of 3077

```
Thr Phe Thr Arg Leu Gly Gly Asp Lys Gln Glu Phe Ala Thr Lys Val
385                 390                 395                 400

Thr Arg Arg Lys Phe Asn Ile Ala Asp Ala Leu Leu Tyr Ser Ser Asn
                405                 410                 415

Asn Thr Pro Trp Ser Asp Val Pro Phe Glu Phe Leu Ile Glu Asn Ile
            420                 425                 430

Gln His Leu Thr Pro Arg Tyr Tyr Ser Ile Ser Ser Ser Ser Leu Ser
        435                 440                 445

Glu Lys Gln Leu Ile Asn Val Thr Ala Val Val Glu Ala Glu Glu Glu
    450                 455                 460

Ala Asp Gly Arg Pro Val Thr Gly Val Val Thr Asn Leu Leu Lys Asn
465                 470                 475                 480

Ile Glu Ile Ala Gln Asn Lys Thr Gly Glu Lys Pro Leu Val His Tyr
                485                 490                 495

Asp Leu Ser Gly Pro Arg Gly Lys Phe Asn Lys Phe Lys Leu Pro Val
            500                 505                 510

His Val Arg Arg Ser Asn Phe Lys Leu Pro Lys Asn Ser Thr Thr Pro
        515                 520                 525

Val Ile Leu Ile Gly Pro Gly Thr Gly Val Ala Pro Leu Arg Gly Phe
    530                 535                 540

Val Arg Glu Arg Val Gln Gln Val Lys Asn Gly Val Asn Val Gly Lys
545                 550                 555                 560

Thr Leu Leu Phe Tyr Gly Cys Arg Asn Ser Asn Glu Asp Phe Leu Tyr
                565                 570                 575

Lys Gln Glu Trp Ala Glu Tyr Ala Ser Val Leu Gly Glu Asn Phe Glu
            580                 585                 590

Met Phe Asn Ala Phe Ser Arg Gln Asp Pro Ser Lys Lys Val Tyr Val
        595                 600                 605

Gln Asp Lys Ile Leu Glu Asn Ser Gln Leu Val His Glu Leu Leu Thr
    610                 615                 620

Glu Gly Ala Ile Ile Tyr Val Cys Gly Asp Ala Ser Arg Met Ala Arg
625                 630                 635                 640
```

```
Asp Val Gln Thr Thr Ile Ser Lys Ile Val Ala Lys Ser Arg Glu Ile
                645                 650                 655

Ser Glu Asp Lys Ala Ala Glu Leu Val Lys Ser Trp Lys Val Gln Asn
            660                 665                 670

Arg Tyr Gln Glu Asp Val Trp
        675
```

<210> 35
<211> 696
<212> PRT
<213> Candida viswanathii

<400> 35

```
Met Gly Ala Pro Leu Thr Val Ala Val Gly Glu Ala Lys Pro Gly Glu
1               5                   10                  15

Thr Ala Pro Arg Arg Lys Ala Ala Gln Lys Met Ala Ser Val Glu Arg
            20                  25                  30

Pro Thr Asp Ser Lys Ala Thr Thr Leu Pro Asp Phe Ile Glu Glu Cys
        35                  40                  45

Phe Ala Arg Asn Gly Thr Arg Asp Ala Met Ala Trp Arg Asp Leu Val
    50                  55                  60

Glu Ile His Val Glu Thr Lys Gln Val Thr Lys Ile Ile Asp Gly Glu
65                  70                  75                  80

Gln Lys Lys Val Asp Lys Asp Trp Ile Tyr Tyr Glu Met Gly Pro Tyr
                85                  90                  95

Asn Tyr Ile Ser Tyr Pro Lys Leu Leu Thr Leu Val Lys Asn Tyr Ser
            100                 105                 110

Lys Gly Leu Leu Glu Leu Gly Leu Ala Pro Asp Gln Glu Ser Lys Leu
        115                 120                 125

Met Ile Phe Ala Ser Thr Ser His Lys Trp Met Gln Thr Phe Leu Ala
    130                 135                 140

Ser Ser Phe Gln Gly Ile Pro Val Val Thr Ala Tyr Asp Thr Leu Gly
145                 150                 155                 160

Glu Ser Gly Leu Thr His Ser Leu Val Gln Thr Glu Ser Asp Ala Val
                165                 170                 175
```

```
Phe Thr Asp Asn Gln Leu Leu Ser Ser Leu Ile Arg Pro Leu Glu Lys
            180                 185                 190

Ala Thr Ser Val Lys Tyr Val Ile His Gly Glu Lys Ile Asp Pro Asn
        195                 200                 205

Asp Lys Arg Gln Gly Gly Lys Ile Tyr Gln Asp Ala Glu Lys Ala Lys
    210                 215                 220

Glu Lys Ile Leu Gln Ile Arg Pro Asp Ile Lys Phe Ile Ser Phe Asp
225                 230                 235                 240

Glu Val Val Ala Leu Gly Glu Gln Ser Ser Lys Glu Leu His Phe Pro
                245                 250                 255

Lys Pro Glu Asp Pro Ile Cys Ile Met Tyr Thr Ser Gly Ser Thr Gly
            260                 265                 270

Ala Pro Lys Gly Val Val Ile Thr Asn Ala Asn Ile Val Ala Ala Val
        275                 280                 285

Gly Gly Ile Ser Thr Asn Ala Thr Arg Asp Leu Val Arg Thr Val Asp
    290                 295                 300

Arg Val Ile Ala Phe Leu Pro Leu Ala His Ile Phe Glu Leu Ala Phe
305                 310                 315                 320

Glu Leu Val Thr Phe Trp Trp Gly Ala Pro Leu Gly Tyr Ala Asn Val
                325                 330                 335

Lys Thr Leu Thr Glu Ala Ser Cys Arg Asn Cys Gln Pro Asp Leu Ile
            340                 345                 350

Glu Phe Lys Pro Thr Ile Met Val Gly Val Ala Ala Val Trp Glu Ser
        355                 360                 365

Val Arg Lys Gly Val Leu Ser Lys Leu Lys Gln Ala Ser Pro Ile Gln
    370                 375                 380

Gln Lys Ile Phe Trp Ala Ala Phe Asn Ala Lys Ser Thr Leu Asn Arg
385                 390                 395                 400

Tyr Gly Leu Pro Gly Gly Gly Leu Phe Asp Ala Val Phe Lys Lys Val
                405                 410                 415

Lys Ala Ala Thr Gly Gly Gln Leu Arg Tyr Val Leu Asn Gly Gly Ser
            420                 425                 430
```

```
Pro Ile Ser Val Asp Ala Gln Val Phe Ile Ser Thr Leu Leu Ala Pro
        435                 440                 445

Met Leu Leu Gly Tyr Gly Leu Thr Glu Thr Cys Ala Asn Thr Thr Ile
    450                 455                 460

Val Glu His Thr Arg Phe Gln Ile Gly Thr Leu Gly Thr Leu Val Gly
465                 470                 475                 480

Ser Val Thr Ala Lys Leu Val Asp Val Ala Asp Ala Gly Tyr Tyr Ala
                485                 490                 495

Lys Asn Asn Gln Gly Glu Ile Trp Leu Lys Gly Gly Pro Val Val Lys
            500                 505                 510

Glu Tyr Tyr Lys Asn Glu Glu Glu Thr Lys Ala Ala Phe Thr Glu Asp
        515                 520                 525

Gly Trp Phe Lys Thr Gly Asp Ile Gly Glu Trp Thr Ala Asp Gly Gly
    530                 535                 540

Leu Asn Ile Ile Asp Arg Lys Lys Asn Leu Val Lys Thr Leu Asn Gly
545                 550                 555                 560

Glu Tyr Ile Ala Leu Glu Lys Leu Glu Ser Ile Tyr Arg Ser Asn His
                565                 570                 575

Leu Ile Leu Asn Leu Cys Val Tyr Ala Asp Gln Thr Lys Val Lys Pro
            580                 585                 590

Ile Ala Ile Val Leu Pro Ile Glu Ala Asn Leu Lys Ser Met Leu Lys
        595                 600                 605

Asp Glu Lys Ile Ile Pro Asp Ala Asp Ser Gln Glu Leu Ser Ser Leu
    610                 615                 620

Val His Asn Lys Lys Val Ala Gln Ala Val Leu Arg His Leu Leu Gln
625                 630                 635                 640

Thr Gly Lys Gln Gln Gly Leu Lys Gly Ile Glu Leu Leu Gln Asn Val
                645                 650                 655

Val Leu Leu Asp Asp Glu Trp Thr Pro Gln Asn Gly Phe Val Thr Ser
            660                 665                 670

Ala Gln Lys Leu Gln Arg Lys Lys Ile Leu Glu Ser Cys Lys Lys Glu
```

675 680 685

Val Glu Glu Ala Tyr Lys Ser Ser
690 695

<210> 36
<211> 662
<212> PRT
<213> Candida viswanathii

<400> 36
Met Pro Thr Glu Leu Gln Lys Glu Arg Glu Leu Thr Lys Phe Asn Pro
1 5 10 15

Lys Glu Leu Asn Tyr Phe Leu Glu Gly Ser Gln Glu Arg Ser Glu Ile
20 25 30

Ile Ser Asn Met Val Glu Gln Met Gln Lys Asp Pro Ile Leu Lys Val
35 40 45

Asp Ala Ser Tyr Tyr Asn Leu Thr Lys Asp Gln Gln Arg Glu Val Thr
50 55 60

Ala Lys Lys Ile Ala Arg Leu Ser Arg Tyr Phe Glu His Glu Tyr Pro
65 70 75 80

Asp Gln Gln Ala Gln Arg Leu Ser Ile Leu Gly Val Phe Asp Pro Gln
85 90 95

Val Phe Thr Arg Ile Gly Val Asn Leu Gly Leu Phe Val Ser Cys Val
100 105 110

Arg Gly Asn Gly Thr Asn Ser Gln Phe Phe Tyr Trp Thr Ile Asn Lys
115 120 125

Gly Ile Asp Lys Leu Arg Gly Ile Tyr Gly Cys Phe Gly Met Thr Glu
130 135 140

Leu Ala His Gly Ser Asn Val Gln Gly Ile Glu Thr Thr Ala Thr Phe
145 150 155 160

Asp Glu Asp Thr Asp Glu Phe Val Ile Asn Thr Pro His Ile Gly Ala
165 170 175

Thr Lys Trp Trp Ile Gly Gly Ala Ala His Ser Ala Thr His Cys Ser
180 185 190

Val Tyr Ala Arg Leu Lys Val Lys Gly Lys Asp Tyr Gly Val Lys Thr
195 200 205

```
Phe Val Val Pro Leu Arg Asp Ser Asn His Asp Leu Glu Pro Gly Val
    210                 215                 220

Thr Val Gly Asp Ile Gly Ala Lys Met Gly Arg Asp Gly Ile Asp Asn
225                 230                 235                 240

Gly Trp Ile Gln Phe Ser Asn Val Arg Ile Pro Arg Phe Phe Met Leu
                245                 250                 255

Gln Lys Tyr Cys Lys Val Ser Arg Leu Gly Glu Val Thr Met Pro Pro
            260                 265                 270

Ser Glu Gln Leu Ser Tyr Ser Ala Leu Ile Gly Gly Arg Val Thr Met
        275                 280                 285

Met Met Asp Ser Tyr Arg Met Thr Ser Arg Phe Ile Thr Ile Ala Leu
    290                 295                 300

Arg Tyr Ala Ile His Arg Arg Gln Phe Lys Lys Lys Asp Thr Asp Thr
305                 310                 315                 320

Ile Glu Thr Lys Leu Ile Asp Tyr Pro Leu His Gln Lys Arg Leu Phe
                325                 330                 335

Pro Phe Leu Ala Ala Ala Tyr Leu Phe Ser Gln Gly Ala Leu Tyr Leu
            340                 345                 350

Glu Gln Thr Met Asn Ala Thr Asn Asp Lys Leu Asp Glu Ala Val Ser
        355                 360                 365

Ala Gly Glu Lys Glu Ala Ile Asp Ala Ala Ile Val Glu Ser Lys Lys
    370                 375                 380

Leu Phe Val Ala Ser Gly Cys Leu Lys Ser Thr Cys Thr Trp Leu Thr
385                 390                 395                 400

Ala Glu Ala Ile Asp Glu Ala Arg Gln Ala Cys Gly Gly His Gly Tyr
                405                 410                 415

Ser Ser Tyr Asn Gly Phe Gly Lys Ala Tyr Ser Asp Trp Val Val Gln
            420                 425                 430

Cys Thr Trp Glu Gly Asp Asn Asn Ile Leu Ala Met Asn Val Ala Lys
        435                 440                 445

Pro Met Val Arg Asp Leu Leu Lys Glu Pro Glu Gln Lys Gly Leu Val
```

```
450                 455                 460

Leu Ser Ser Val Ala Asp Leu Asp Asp Pro Ala Lys Leu Val Lys Ala
465                 470                 475                 480

Phe Asp His Ala Leu Ser Gly Leu Ala Arg Asp Ile Gly Ala Val Ala
                485                 490                 495

Glu Asp Lys Gly Phe Asp Ile Thr Gly Pro Ser Leu Val Leu Val Ser
            500                 505                 510

Lys Leu Asn Ala His Arg Phe Leu Ile Asp Gly Phe Phe Lys Arg Ile
        515                 520                 525

Thr Pro Glu Trp Ser Glu Val Leu Arg Pro Leu Gly Phe Leu Tyr Ala
    530                 535                 540

Asp Trp Ile Leu Thr Asn Phe Gly Ala Thr Phe Leu Gln Tyr Gly Ile
545                 550                 555                 560

Ile Thr Pro Asp Val Ser Arg Lys Ile Ser Ser Glu His Phe Pro Ala
                565                 570                 575

Leu Cys Ala Lys Val Arg Pro Asn Val Val Gly Leu Thr Asp Gly Phe
            580                 585                 590

Asn Leu Thr Asp Met Met Thr Asn Ala Ala Ile Gly Arg Tyr Asp Gly
        595                 600                 605

Asn Val Tyr Glu His Tyr Phe Glu Thr Val Lys Ala Leu Asn Pro Pro
    610                 615                 620

Glu Asn Thr Lys Ala Pro Tyr Ser Lys Ala Leu Glu Asp Met Leu Asn
625                 630                 635                 640

Arg Pro Asp Leu Glu Val Arg Glu Arg Gly Glu Lys Ser Glu Glu Ala
                645                 650                 655

Ala Glu Ile Leu Ser Ser
            660


<210> 37
<211> 662
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
      polypeptide
```

<400> 37

```
Met Pro Thr Glu Leu Gln Lys Glu Arg Glu Leu Thr Lys Phe Asn Pro
1               5                  10                  15

Lys Glu Leu Asn Tyr Phe Leu Glu Gly Ser Gln Glu Arg Ser Glu Ile
            20                  25                  30

Ile Ser Asn Met Val Glu Gln Met Gln Lys Asp Pro Ile Leu Lys Val
        35                  40                  45

Asp Ala Ser Tyr Tyr Asn Leu Thr Lys Asp Gln Gln Arg Glu Val Thr
    50                  55                  60

Ala Lys Lys Ile Ala Arg Leu Ser Arg Tyr Phe Glu His Glu Tyr Pro
65                  70                  75                  80

Asp Gln Gln Ala Gln Arg Leu Ser Ile Leu Gly Val Phe Asp Pro Gln
                85                  90                  95

Val Gly Thr Arg Ile Gly Val Asn Leu Gly Leu Phe Val Ser Cys Val
            100                 105                 110

Arg Gly Asn Gly Thr Asn Ser Gln Phe Phe Tyr Trp Thr Ile Asn Lys
        115                 120                 125

Gly Ile Asp Lys Leu Arg Gly Ile Tyr Gly Cys Phe Gly Met Thr Glu
    130                 135                 140

Leu Ala His Gly Ser Asn Val Gln Gly Ile Glu Thr Thr Ala Thr Phe
145                 150                 155                 160

Asp Glu Asp Thr Asp Glu Phe Val Ile Asn Thr Pro His Ile Gly Ala
                165                 170                 175

Thr Lys Trp Trp Ile Gly Gly Ala Ala His Ser Ala Thr His Cys Ser
            180                 185                 190

Val Tyr Ala Arg Leu Lys Val Lys Gly Lys Asp Tyr Gly Val Lys Thr
        195                 200                 205

Phe Val Val Pro Leu Arg Asp Ser Asn His Asp Leu Glu Pro Gly Val
    210                 215                 220

Thr Val Gly Asp Ile Gly Ala Lys Met Gly Arg Asp Gly Ile Asp Asn
225                 230                 235                 240

Gly Trp Ile Gln Phe Ser Asn Val Arg Ile Pro Arg Phe Phe Met Leu
```

245 250 255

Gln Lys Tyr Cys Lys Val Ser Arg Leu Gly Glu Val Thr Met Pro Pro
260 265 270

Ser Glu Gln Leu Ser Tyr Ser Ala Leu Ile Gly Gly Arg Val Thr Met
275 280 285

Met Met Asp Ser Tyr Arg Met Thr Ser Arg Phe Ile Thr Ile Ala Leu
290 295 300

Arg Tyr Ala Ile His Arg Arg Gln Phe Lys Lys Lys Asp Thr Asp Thr
305 310 315 320

Ile Glu Thr Lys Leu Ile Asp Tyr Pro Leu His Gln Lys Arg Leu Phe
325 330 335

Pro Phe Leu Ala Ala Ala Tyr Leu Phe Ser Gln Gly Ala Leu Tyr Leu
340 345 350

Glu Gln Thr Met Asn Ala Thr Asn Asp Lys Leu Asp Glu Ala Val Ser
355 360 365

Ala Gly Glu Lys Glu Ala Ile Asp Ala Ala Ile Val Glu Ser Lys Lys
370 375 380

Leu Phe Val Ala Ser Gly Cys Leu Lys Ser Thr Cys Thr Trp Leu Thr
385 390 395 400

Ala Glu Ala Ile Asp Glu Ala Arg Gln Ala Cys Gly Gly His Gly Tyr
405 410 415

Ser Ser Tyr Asn Gly Phe Gly Lys Ala Tyr Ser Asp Phe Val Val Gln
420 425 430

Cys Thr Trp Glu Gly Asp Asn Asn Ile Leu Ala Met Asn Val Ala Lys
435 440 445

Pro Met Val Arg Asp Leu Leu Lys Glu Pro Glu Gln Lys Gly Leu Val
450 455 460

Leu Ser Ser Val Ala Asp Leu Asp Asp Pro Ala Lys Leu Val Lys Ala
465 470 475 480

Phe Asp His Ala Leu Ser Gly Leu Ala Arg Asp Ile Gly Ala Val Ala
485 490 495

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 140 of 3077

Glu Asp Lys Gly Phe Asp Ile Thr Gly Pro Ser Leu Val Leu Val Ser
500 505 510

Lys Leu Asn Ala His Arg Phe Leu Ile Asp Gly Phe Phe Lys Arg Ile
515 520 525

Thr Pro Glu Trp Ser Glu Val Leu Arg Pro Leu Gly Phe Leu Tyr Ala
530 535 540

Asp Trp Ile Leu Thr Asn Phe Gly Ala Thr Phe Leu Gln Tyr Gly Ile
545 550 555 560

Ile Thr Pro Asp Val Ser Arg Lys Ile Ser Ser Glu His Phe Pro Ala
565 570 575

Leu Cys Ala Lys Val Arg Pro Asn Val Val Gly Leu Thr Asp Gly Phe
580 585 590

Asn Leu Thr Asp Met Met Thr Asn Ala Ala Ile Gly Arg Tyr Asp Gly
595 600 605

Asn Val Tyr Glu His Tyr Phe Glu Thr Val Lys Ala Leu Asn Pro Pro
610 615 620

Glu Asn Thr Lys Ala Pro Tyr Ser Lys Ala Leu Glu Asp Met Leu Asn
625 630 635 640

Arg Pro Asp Leu Glu Val Arg Glu Arg Gly Glu Lys Ser Glu Glu Ala
645 650 655

Ala Glu Ile Leu Ser Ser
660

<210> 38
<211> 662
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 38
Met Pro Thr Glu Leu Gln Lys Glu Arg Glu Leu Thr Lys Phe Asn Pro
1 5 10 15

Lys Glu Leu Asn Tyr Phe Leu Glu Gly Ser Gln Glu Arg Ser Glu Ile
20 25 30

Ile Ser Asn Met Val Glu Gln Met Gln Lys Asp Pro Ile Leu Lys Val

35 40 45

Asp Ala Ser Tyr Tyr Asn Leu Thr Lys Asp Gln Gln Arg Glu Val Thr
50 55 60

Ala Lys Lys Ile Ala Arg Leu Ser Arg Tyr Phe Glu His Glu Tyr Pro
65 70 75 80

Asp Gln Gln Ala Gln Arg Leu Ser Ile Leu Gly Val Phe Asp Pro Gln
85 90 95

Val Phe Thr Arg Ile Gly Val Asn Leu Gly Leu Phe Val Ser Cys Val
100 105 110

Arg Gly Asn Gly Thr Asn Ser Gln Phe Phe Tyr Trp Thr Ile Asn Lys
115 120 125

Gly Ile Asp Lys Leu Arg Gly Ile Tyr Gly Cys Phe Gly Met Thr Glu
130 135 140

Leu Ala His Gly Ser Asn Val Gln Gly Ile Glu Thr Thr Ala Thr Phe
145 150 155 160

Asp Glu Asp Thr Asp Glu Phe Val Ile Asn Thr Pro His Ile Gly Ala
165 170 175

Thr Lys Trp Trp Ile Gly Gly Ala Ala His Ser Ala Thr His Cys Ser
180 185 190

Val Tyr Ala Arg Leu Lys Val Lys Gly Lys Asp Tyr Gly Val Lys Thr
195 200 205

Phe Val Val Pro Leu Arg Asp Ser Asn His Asp Leu Glu Pro Gly Val
210 215 220

Thr Val Gly Asp Ile Gly Ala Lys Met Gly Arg Asp Gly Ile Asp Asn
225 230 235 240

Gly Trp Ile Gln Phe Ser Asn Val Arg Ile Pro Arg Phe Phe Met Leu
245 250 255

Gln Lys Tyr Cys Lys Val Ser Arg Leu Gly Glu Val Thr Met Pro Pro
260 265 270

Ser Glu Gln Leu Ser Tyr Ser Ala Leu Ile Gly Gly Arg Val Thr Met
275 280 285

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 142 of 3077

```
Met Met Asp Ser Tyr Arg Met Thr Ser Arg Phe Ile Thr Ile Ala Leu
    290                 295                 300

Arg Tyr Ala Ile His Arg Arg Gln Phe Lys Lys Lys Asp Thr Asp Thr
305                 310                 315                 320

Ile Glu Thr Lys Leu Ile Asp Tyr Pro Leu His Gln Lys Arg Leu Phe
                325                 330                 335

Pro Phe Leu Ala Ala Ala Tyr Leu Phe Ser Gln Gly Ala Leu Tyr Leu
            340                 345                 350

Glu Gln Thr Met Asn Ala Thr Asn Asp Lys Leu Asp Glu Ala Val Ser
        355                 360                 365

Ala Gly Glu Lys Glu Ala Ile Asp Ala Ala Ile Val Glu Ser Lys Lys
    370                 375                 380

Leu Phe Val Ala Ser Gly Cys Leu Lys Ser Thr Cys Thr Trp Leu Thr
385                 390                 395                 400

Ala Glu Ala Ile Asp Glu Ala Arg Gln Ala Cys Gly Gly His Gly Tyr
                405                 410                 415

Ser Ser Tyr Asn Gly Phe Gly Lys Ala Tyr Ser Asp Trp Val Val Gln
            420                 425                 430

Cys Thr Trp Glu Gly Asp Asn Asn Ile Leu Ala Met Asn Val Ala Lys
        435                 440                 445

Pro Met Val Arg Asp Leu Leu Lys Glu Pro Glu Gln Lys Gly Leu Val
    450                 455                 460

Leu Ser Ser Val Ala Asp Leu Asp Asp Pro Ala Lys Leu Val Lys Ala
465                 470                 475                 480

Phe Asp His Ala Leu Ser Gly Leu Ala Arg Asp Ile Gly Ala Val Ala
                485                 490                 495

Glu Asp Lys Gly Phe Asp Ile Thr Gly Pro Ser Leu Val Leu Val Ser
            500                 505                 510

Lys Leu Asn Ala His Arg Phe Leu Ile Asp Gly Phe Phe Lys Arg Ile
        515                 520                 525

Thr Pro Glu Trp Ser Glu Val Leu Arg Pro Leu Gly Phe Leu Tyr Ala
    530                 535                 540
```

```
Asp Trp Ile Leu Thr Asn Phe Gly Ala Thr Phe Leu Gln Tyr Gly Ile
545                 550                 555                 560

Ile Thr Pro Asp Val Ser Arg Lys Ile Ser Ser Glu His Phe Pro Ala
                565                 570                 575

Leu Cys Ala Lys Val Arg Pro Asn Val Val Gly Leu Thr Asp Gly Phe
            580                 585                 590

Asn Leu Thr Asp Met Met Thr Asn Ala Ala Ile Gly Arg Tyr Asp Gly
        595                 600                 605

Asn Val Tyr Glu His Tyr Phe Glu Thr Val Lys Ala Leu Asn Pro Pro
    610                 615                 620

Glu Asn Thr Lys Ala Pro Tyr Ser Lys Ala Leu Glu Asp Met Leu Asn
625                 630                 635                 640

Arg Pro Asp Leu Glu Val Arg Glu Arg Gly Glu Lys Ser Glu Glu Ala
                645                 650                 655

Ala Glu Ile Leu Ser Ser
            660

<210> 39
<211> 709
<212> PRT
<213> Candida viswanathii

<400> 39
Met Thr Phe Thr Lys Lys Asn Val Ser Val Ser Gln Gly Pro Asp Pro
1               5                   10                  15

Arg Ser Ser Ile Gln Lys Glu Arg Asp Ser Ser Lys Trp Asn Pro Gln
            20                  25                  30

Gln Met Asn Tyr Phe Leu Glu Gly Ser Val Glu Arg Ser Glu Leu Met
        35                  40                  45

Lys Ala Leu Ala Gln Gln Met Glu Arg Asp Pro Ile Leu Phe Thr Asp
    50                  55                  60

Gly Ser Tyr Tyr Asp Leu Thr Lys Asp Gln Gln Arg Glu Leu Thr Ala
65                  70                  75                  80

Val Lys Ile Asn Arg Ile Ala Arg Tyr Arg Glu Gln Glu Ser Ile Asp
                85                  90                  95
```


LCY Biotechnology Holding, Inc.
Ex. 1002
Page 144 of 3077

```
Thr Phe Asn Lys Arg Leu Ser Leu Ile Gly Ile Phe Asp Pro Gln Val
            100                 105                 110

Gly Thr Arg Ile Gly Val Asn Leu Gly Leu Phe Leu Ser Cys Ile Arg
        115                 120                 125

Gly Asn Gly Thr Thr Ser Gln Leu Asn Tyr Trp Ala Asn Glu Lys Glu
    130                 135                 140

Thr Ala Asp Val Lys Gly Ile Tyr Gly Cys Phe Gly Met Thr Glu Leu
145                 150                 155                 160

Ala His Gly Ser Asn Val Ala Gly Leu Glu Thr Thr Ala Thr Phe Asp
                165                 170                 175

Lys Glu Ser Asp Glu Phe Val Ile Asn Thr Pro His Ile Gly Ala Thr
            180                 185                 190

Lys Trp Trp Ile Gly Gly Ala Ala His Ser Ala Thr His Cys Ser Val
        195                 200                 205

Tyr Ala Arg Leu Ile Val Asp Gly Gln Asp Tyr Gly Val Lys Thr Phe
    210                 215                 220

Val Val Pro Leu Arg Asp Ser Asn His Asp Leu Met Pro Gly Val Thr
225                 230                 235                 240

Val Gly Asp Ile Gly Ala Lys Met Gly Arg Asp Gly Ile Asp Asn Gly
                245                 250                 255

Trp Ile Gln Phe Ser Asn Val Arg Ile Pro Arg Phe Phe Met Leu Gln
            260                 265                 270

Lys Phe Cys Lys Val Ser Ala Glu Gly Glu Val Thr Leu Pro Pro Leu
        275                 280                 285

Glu Gln Leu Ser Tyr Ser Ala Leu Leu Gly Gly Arg Val Met Met Val
    290                 295                 300

Leu Asp Ser Tyr Arg Met Leu Ala Arg Met Ser Thr Ile Ala Leu Arg
305                 310                 315                 320

Tyr Ala Ile Gly Arg Arg Gln Phe Lys Gly Asp Asn Val Asp Pro Lys
                325                 330                 335

Asp Pro Asn Ala Leu Glu Thr Gln Leu Ile Asp Tyr Pro Leu His Gln
            340                 345                 350
```


LCY Biotechnology Holding, Inc.
Ex. 1002
Page 145 of 3077

```
Lys Arg Leu Phe Pro Tyr Leu Ala Ala Ala Tyr Val Ile Ser Ala Gly
        355                 360                 365

Ala Leu Lys Val Glu Asp Thr Ile His Asn Thr Leu Ala Glu Leu Asp
    370                 375                 380

Ala Ala Val Glu Lys Asn Asp Thr Lys Ala Ile Phe Lys Ser Ile Asp
385                 390                 395                 400

Asp Met Lys Ser Leu Phe Val Asp Ser Gly Ser Leu Lys Ser Thr Ala
                405                 410                 415

Thr Trp Leu Gly Ala Glu Ala Ile Asp Gln Cys Arg Gln Ala Cys Gly
            420                 425                 430

Gly His Gly Tyr Ser Ser Tyr Asn Gly Phe Gly Lys Ala Tyr Asn Asp
        435                 440                 445

Trp Val Val Gln Cys Thr Trp Glu Gly Asp Asn Asn Val Leu Ala Met
    450                 455                 460

Ser Val Gly Lys Pro Ile Val Lys Gln Val Ile Ser Ile Glu Asp Ala
465                 470                 475                 480

Gly Lys Thr Val Arg Gly Ser Thr Ala Phe Leu Asn Gln Leu Lys Asp
                485                 490                 495

Tyr Thr Gly Ser Asn Ser Ser Lys Val Val Leu Asn Thr Val Ala Asp
            500                 505                 510

Leu Asp Asp Ile Lys Thr Val Ile Lys Ala Ile Glu Val Ala Ile Ile
        515                 520                 525

Arg Leu Ser Gln Glu Ala Ala Ser Ile Val Lys Lys Glu Ser Phe Asp
    530                 535                 540

Tyr Val Gly Ala Glu Leu Val Gln Leu Ser Lys Leu Lys Ala His His
545                 550                 555                 560

Tyr Leu Leu Thr Glu Tyr Ile Arg Arg Ile Asp Thr Phe Asp Gln Lys
                565                 570                 575

Asp Leu Val Pro Tyr Leu Ile Thr Leu Gly Lys Leu Tyr Ala Ala Thr
            580                 585                 590

Ile Val Leu Asp Arg Phe Ala Gly Val Phe Leu Thr Phe Asn Val Ala
        595                 600                 605
```

```
Ser Thr Glu Ala Ile Thr Ala Leu Ala Ser Val Gln Ile Pro Lys Leu
    610                 615                 620

Cys Ala Glu Val Arg Pro Asn Val Val Ala Tyr Thr Asp Ser Phe Gln
625                 630                 635                 640

Gln Ser Asp Met Ile Val Asn Ser Ala Ile Gly Arg Tyr Asp Gly Asp
                645                 650                 655

Ile Tyr Glu Asn Tyr Phe Asp Leu Val Lys Leu Gln Asn Pro Pro Ser
            660                 665                 670

Lys Thr Lys Ala Pro Tyr Ser Asp Ala Leu Glu Ala Met Leu Asn Arg
        675                 680                 685

Pro Thr Leu Asp Glu Arg Glu Arg Phe Glu Lys Ser Asp Glu Thr Ala
    690                 695                 700

Ala Ile Leu Ser Lys
705


<210> 40
<211> 735
<212> PRT
<213> Candida viswanathii

<400> 40
Met Val Asn Ile Ser Lys Leu Thr Gly Tyr Asn Lys Gln Asp Ile Arg
1               5                   10                  15

Asn Val Val Leu Leu Leu Gln Glu Phe Val Lys Thr Tyr Lys Asp Asn
            20                  25                  30

Lys Ile Lys Leu Asn Tyr Ser Ser Arg Pro Val Ile Leu Phe Leu Ser
        35                  40                  45

Thr Leu Val Ala Thr Ala Gly Ile Gly Val Phe Phe Thr Leu Arg Ser
    50                  55                  60

Ile Val Thr Lys Tyr Asn Glu Tyr Leu Leu Asn Lys Arg Leu Arg Arg
65                  70                  75                  80

Pro Ser Leu Ile Arg Gln Ser Ser Asn Ile Leu Lys Asn Gly Ser Arg
                85                  90                  95

Glu Ile Phe Ile Gln Lys Gly Asn Gly Lys Val Thr Arg Ile Ile Ile
            100                 105                 110
```


LCY Biotechnology Holding, Inc.
Ex. 1002
Page 147 of 3077

```
Pro Lys Ala Asn Asn Asp Gln Tyr Ala Ala Asp Lys Tyr Leu Tyr Lys
        115                 120                 125

Asp Phe Ala Arg Asn Glu Gln Ile Leu Gln Gln Gln Lys Gly Arg Leu
    130                 135                 140

Phe Asn Ser Arg Phe Leu Asn Gln Leu Thr Ile Ile Trp Lys Ile Leu
145                 150                 155                 160

Ile Pro Lys Phe Tyr Cys Gln Asn Thr Ser Leu Leu Leu Ser Gln Cys
                165                 170                 175

Phe Phe Leu Ile Phe Arg Thr Trp Leu Ser Leu Leu Ile Ala Lys Leu
            180                 185                 190

Asp Gly Gln Ile Val Lys Asn Leu Ile Ala Ala Asp Gly Arg Lys Phe
        195                 200                 205

Ala Arg Asp Leu Ile Tyr Phe Leu Leu Ile Ala Phe Pro Ala Ser Tyr
    210                 215                 220

Thr Asn Ala Ala Ile Lys Tyr Leu Glu Leu Arg Leu Ala Leu Gly Phe
225                 230                 235                 240

Arg Thr Asn Leu Thr Arg Tyr Ile His Asp Met Tyr Leu Asp Lys Thr
                245                 250                 255

Met Ser Tyr Tyr Lys Val Gly Leu Asn Gly Ala Asp Ile Gln Asn Ile
            260                 265                 270

Asp Gln Tyr Ile Thr Glu Asp Val Thr Lys Phe Cys Met Ser Leu Cys
        275                 280                 285

Ser Leu Phe Ser Ser Met Gly Lys Pro Phe Ile Asp Leu Ile Phe Phe
    290                 295                 300

Ser Val Tyr Leu Arg Asp Asn Leu Gly Thr Gly Ala Ile Ile Gly Ile
305                 310                 315                 320

Phe Ala Asn Tyr Phe Ala Thr Ala Ile Met Leu Lys Lys Ala Thr Pro
                325                 330                 335

Arg Phe Gly Lys Leu Ala Ala Lys Arg Thr His Leu Glu Gly Val Tyr
            340                 345                 350

Phe Asn Gln Gln Leu Asn Ile Met Thr Asn Ser Glu Glu Ile Gly Phe
        355                 360                 365
```

```
Tyr Lys Gly Ser Lys Ile Glu Lys Ser Lys Leu Ala Glu Asn Phe Asp
    370                 375                 380

Lys Leu Met Gly His Val Ser Arg Glu Ile Asn Leu Ser Ser Ser Tyr
385                 390                 395                 400

Ala Ala Leu Glu Asp Tyr Val Leu Lys Tyr Thr Trp Ser Ala Trp Gly
                405                 410                 415

Tyr Ile Phe Ser Gly Leu Pro Val Phe Leu Asp Val Leu Phe Pro Lys
            420                 425                 430

Glu Asp Pro Ser Ser Gly His Ile Ala Asp Ile Asp Asp Asp Asp His
        435                 440                 445

Ala His Gly His Gly His Thr Gly Glu Glu Thr Ser Ser Thr Thr Glu
    450                 455                 460

Asn Met Lys Thr Phe Val Thr Asn Lys Arg Leu Leu Leu Ser Leu Ala
465                 470                 475                 480

Asp Ala Gly Ser Arg Leu Met Val Ser Leu Lys Glu Val Thr Thr Leu
                485                 490                 495

Thr Gly Ile Thr Asn Arg Val Phe Asn Met Leu Thr Gln Leu His Arg
            500                 505                 510

Val His Asp Pro Lys Phe Asp Tyr Gly Asp Lys Tyr Gly Leu Pro Asp
        515                 520                 525

Ile His Gly Thr Tyr Gln Leu Asn Tyr Asp Gly Leu Arg Leu Glu His
    530                 535                 540

Val Pro Ile Thr Val Pro Thr Ala Glu Gly Ser Tyr Ser Thr Pro Leu
545                 550                 555                 560

Ile Pro Asp Leu Thr Phe Asp Ile Lys Gly Lys Asn Leu Leu Phe Val
                565                 570                 575

Gly Pro Asn Gly Ser Gly Lys Thr Ser Val Ala Arg Val Leu Ala Gly
            580                 585                 590

Leu Trp Pro Leu Tyr Ala Gly Leu Val Ser Lys Pro Ser Asp Leu Phe
        595                 600                 605

Phe Asn Pro Gln Lys Ser Tyr Phe Thr Thr Gly Ser Leu Arg Asp Gln
```

610 615 620

Val Val Tyr Pro Asn Arg Ser Glu Asn Thr Thr Asn Asp Gln Ile Phe
625 630 635 640

His Ile Leu His Cys Val His Leu Asp His Ile Val Lys Arg Tyr Gly
645 650 655

Leu Asn Gln Asn Leu Asp Phe Ala Lys Thr Leu Ser Gly Gly Glu Lys
660 665 670

Gln Arg Leu Ser Phe Ala Arg Val Leu Phe Asn Arg Pro Ser Ile Val
675 680 685

Ile Leu Asp Asp Ser Thr Ser Ala Leu Ser Pro Asp Met Glu Glu Leu
690 695 700

Met Tyr Gln Val Leu Gln Asp His Lys Ile Asn Tyr Val Thr Leu Ser
705 710 715 720

Asn Arg Pro Ser Leu Ser Lys Phe His Asp Lys Val Phe Glu Ile
725 730 735

<210> 41
<211> 665
<212> PRT
<213> Candida viswanathii

<400> 41
Met Thr Val Glu Asn Ala Lys Leu Gln Lys Asn Ser Leu Ala Val Ser
1 5 10 15

Leu Leu Lys Val Tyr Lys Ser Asn Arg Ser Leu Leu Leu Asn Thr Ser
20 25 30

Tyr Ile Ile Leu Ile Ile Ala Ala Phe Thr Gly Ala Thr Asn Thr Gly
35 40 45

Arg Gly Thr Ser Ser Arg Ser Ser Ala Lys Val Glu Thr Asp Glu Glu
50 55 60

Gln Ser Val Lys Lys Lys His Pro Lys Leu Ser Arg Glu Ser Phe His
65 70 75 80

Arg Leu Arg Lys Ala Ile Leu Pro Thr Phe Phe Asp Arg Thr Ile Val
85 90 95

Tyr Phe Phe Ala Asn Leu Thr Leu Leu Val Val Arg Ala Leu Leu Thr
100 105 110

```
Leu Arg Val Ala Thr Leu Asp Gly Gln Leu Val Gly Ala Leu Val Ser
        115                 120                 125

Arg Arg Ile Arg Val Phe Ala Lys Tyr Leu Leu Tyr Trp Met Leu Leu
    130                 135                 140

Gly Ile Pro Ala Ala Leu Thr Asn Ala Leu Leu Asn Trp Thr Lys Ser
145                 150                 155                 160

Asn Leu Ser Lys Ser Ile Arg Met Asn Leu Asn Asn Asn Ile Met Glu
                165                 170                 175

Glu Tyr Leu Pro Asp Asn Leu Asp Pro Asn Tyr Tyr Ser Leu Ile His
            180                 185                 190

Leu Thr Asp Asn Lys Ile Arg Asp Pro Asn Gln Arg Ile Thr Thr Asp
        195                 200                 205

Thr Ser Arg Leu Ser Asp Ala Leu Ala Ser Leu Pro Gly His Ile Leu
    210                 215                 220

Lys Pro Thr Leu Asp Ile Ile Leu Cys Ala Gln Gln Leu Ser Lys Ser
225                 230                 235                 240

Gly Val Gly Asn Gly Glu Gly Thr Leu Ala Leu Gly Ile Leu Ala His
                245                 250                 255

Phe Ser Thr Met Ile Ile Arg Phe Phe Ser Pro Pro Phe Ala Lys Leu
            260                 265                 270

Ala Ala Glu Arg Ala Asn Leu Glu Gly Gln Leu Arg Ser Ala His Ser
        275                 280                 285

Lys Ile Val Ala Asn Ser Glu Glu Ile Ala Phe Leu Gly Gly His Asp
    290                 295                 300

Arg Glu Leu Asp His Ile Asp His Cys Tyr Tyr Thr Leu Glu Arg Phe
305                 310                 315                 320

Ser Lys Gly Glu Tyr Trp Lys Arg Ala Ile His Glu Ile Thr Gln Thr
                325                 330                 335

Phe Ile Val Lys Tyr Phe Trp Gly Val Ala Gly Leu Val Leu Cys Ser
            340                 345                 350

Ala Pro Val Phe Ile Ala Lys Tyr Leu Gly Glu Pro Glu Asp Lys Asn
```

```
        355                 360                 365

Val Ala Gly Asn Phe Ile Thr Asn Arg Arg Leu Leu Met Ser Ala Ser
    370                 375                 380

Asp Ser Leu Asp Arg Leu Ile Tyr Ser Arg Arg Tyr Leu Leu Gln Val
385                 390                 395                 400

Val Gly His Ala Thr Arg Val Ser Asp Phe Leu Asp Thr Leu His Glu
                405                 410                 415

Val Glu Glu Lys Lys Lys Arg Ile Thr Ser Asn Val Gln Phe Asn Asn
            420                 425                 430

Asp Glu Ile Thr Phe Asp His Val Arg Leu Met Thr Pro Thr Glu Val
        435                 440                 445

Thr Leu Ile Pro Asp Leu Asn Phe Ser Ile Lys Pro Gly Asp His Leu
    450                 455                 460

Leu Ile Val Gly Pro Asn Gly Ser Gly Lys Ser Ser Leu Phe Arg Met
465                 470                 475                 480

Leu Gly Gly Leu Trp Pro Val Arg Phe Gly Thr Ile Arg Ile Pro Asn
                485                 490                 495

Thr Glu Asn Met Phe Tyr Leu Pro Gln Lys Ala Tyr Leu Val Glu Gly
            500                 505                 510

Ser Phe Arg Glu Gln Ile Ile Tyr Pro His Asn Val Thr Gln Gln Lys
        515                 520                 525

Lys Thr Asp Gln Gln Leu Lys Glu Ile Leu Lys Val Leu Lys Leu Glu
    530                 535                 540

Asp Tyr Ser Gly Gln Leu Asp Glu Val Lys Lys Trp Ser Glu Glu Leu
545                 550                 555                 560

Ser Ile Gly Ala Gln Gln Arg Leu Ala Met Ala Arg Leu Tyr Tyr His
                565                 570                 575

Glu Pro Lys Phe Ala Val Leu Asp Glu Cys Thr Ser Ala Val Ser Pro
            580                 585                 590

Asp Met Glu Gln Leu Met Tyr Gln His Ala Gln Gly Leu Gly Ile Thr
        595                 600                 605
```


LCY Biotechnology Holding, Inc.
Ex. 1002
Page 152 of 3077

```
Leu Leu Ser Val Ala His Arg Pro Ala Leu Trp His Phe His Lys Tyr
    610                 615                 620

Leu Leu Glu Phe Asp Gly Lys Gly Ser Tyr Tyr Phe Gly Thr Leu Asp
625                 630                 635                 640

Glu Lys His Lys Met Lys Leu Glu Glu Glu Glu Arg Leu Lys Lys Glu
                645                 650                 655

Asn Glu Lys Lys Ser Val Ala Lys Lys
            660                 665
```

<210> 42
<211> 650
<212> PRT
<213> Yarrowia lipolytica

<400> 42

```
Met Ser Ala Asn Glu Asn Ile Ser Arg Phe Asp Ala Pro Val Gly Lys
1               5                   10                  15

Glu His Pro Ala Tyr Glu Leu Phe His Asn His Thr Arg Ser Phe Val
            20                  25                  30

Tyr Gly Leu Gln Pro Arg Ala Cys Gln Gly Met Leu Asp Phe Asp Phe
        35                  40                  45

Ile Cys Lys Arg Glu Asn Pro Ser Val Ala Gly Val Ile Tyr Pro Phe
    50                  55                  60

Gly Gly Gln Phe Val Thr Lys Met Tyr Trp Gly Thr Lys Glu Thr Leu
65                  70                  75                  80

Leu Pro Val Tyr Gln Gln Val Glu Lys Ala Ala Ala Lys His Pro Glu
                85                  90                  95

Val Asp Val Val Val Asn Phe Ala Ser Ser Arg Ser Val Tyr Ser Ser
            100                 105                 110

Thr Met Glu Leu Leu Glu Tyr Pro Gln Phe Arg Thr Ile Ala Ile Ile
        115                 120                 125

Ala Glu Gly Val Pro Glu Arg Arg Ala Arg Glu Ile Leu His Lys Ala
    130                 135                 140

Gln Lys Lys Gly Val Thr Ile Ile Gly Pro Ala Thr Val Gly Gly Ile
145                 150                 155                 160

Lys Pro Gly Cys Phe Lys Val Gly Asn Thr Gly Gly Met Met Asp Asn
```

165 170 175

Ile Val Ala Ser Lys Leu Tyr Arg Pro Gly Ser Val Ala Tyr Val Ser
180 185 190

Lys Ser Gly Gly Met Ser Asn Glu Leu Asn Asn Ile Ile Ser His Thr
195 200 205

Thr Asp Gly Val Tyr Glu Gly Ile Ala Ile Gly Gly Asp Arg Tyr Pro
210 215 220

Gly Thr Thr Phe Ile Asp His Ile Leu Arg Tyr Glu Ala Asp Pro Lys
225 230 235 240

Cys Lys Ile Ile Val Leu Leu Gly Glu Val Gly Gly Val Glu Glu Tyr
245 250 255

Arg Val Ile Glu Ala Val Lys Asn Gly Gln Ile Lys Lys Pro Ile Val
260 265 270

Ala Trp Ala Ile Gly Thr Cys Ala Ser Met Phe Lys Thr Glu Val Gln
275 280 285

Phe Gly His Ala Gly Ser Met Ala Asn Ser Asp Leu Glu Thr Ala Lys
290 295 300

Ala Lys Asn Ala Ala Met Lys Ser Ala Gly Phe Tyr Val Pro Asp Thr
305 310 315 320

Phe Glu Asp Met Pro Glu Val Leu Ala Glu Leu Tyr Glu Lys Met Val
325 330 335

Ala Lys Gly Glu Leu Ser Arg Ile Ser Glu Pro Glu Val Pro Lys Ile
340 345 350

Pro Ile Asp Tyr Ser Trp Ala Gln Glu Leu Gly Leu Ile Arg Lys Pro
355 360 365

Ala Ala Phe Ile Ser Thr Ile Ser Asp Asp Arg Gly Gln Glu Leu Leu
370 375 380

Tyr Ala Gly Met Pro Ile Ser Glu Val Phe Lys Glu Asp Ile Gly Ile
385 390 395 400

Gly Gly Val Met Ser Leu Leu Trp Phe Arg Arg Arg Leu Pro Asp Tyr
405 410 415

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 154 of 3077

```
Ala Ser Lys Phe Leu Glu Met Val Leu Met Leu Thr Ala Asp His Gly
            420                 425                 430

Pro Ala Val Ser Gly Ala Met Asn Thr Ile Ile Thr Thr Arg Ala Gly
        435                 440                 445

Lys Asp Leu Ile Ser Ser Leu Val Ala Gly Leu Leu Thr Ile Gly Thr
    450                 455                 460

Arg Phe Gly Gly Ala Leu Asp Gly Ala Ala Thr Glu Phe Thr Thr Ala
465                 470                 475                 480

Tyr Asp Lys Gly Leu Ser Pro Arg Gln Phe Val Asp Thr Met Arg Lys
                485                 490                 495

Gln Asn Lys Leu Ile Pro Gly Ile Gly His Arg Val Lys Ser Arg Asn
            500                 505                 510

Asn Pro Asp Phe Arg Val Glu Leu Val Lys Asp Phe Val Lys Lys Asn
        515                 520                 525

Phe Pro Ser Thr Gln Leu Leu Asp Tyr Ala Leu Ala Val Glu Glu Val
    530                 535                 540

Thr Thr Ser Lys Lys Asp Asn Leu Ile Leu Asn Val Asp Gly Ala Ile
545                 550                 555                 560

Ala Val Ser Phe Val Asp Leu Met Arg Ser Cys Gly Ala Phe Thr Val
                565                 570                 575

Glu Glu Thr Glu Asp Tyr Leu Lys Asn Gly Val Leu Asn Gly Leu Phe
            580                 585                 590

Val Leu Gly Arg Ser Ile Gly Leu Ile Ala His His Leu Asp Gln Lys
        595                 600                 605

Arg Leu Lys Thr Gly Leu Tyr Arg His Pro Trp Asp Asp Ile Thr Tyr
    610                 615                 620

Leu Val Gly Gln Glu Ala Ile Gln Lys Lys Arg Val Glu Ile Ser Ala
625                 630                 635                 640

Gly Asp Val Ser Lys Ala Lys Thr Arg Ser
                645                 650
```

<210> 43
<211> 497
<212> PRT

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 155 of 3077

<213> Yarrowia lipolytica

<400> 43

```
Met Ser Ala Lys Ser Ile His Glu Ala Asp Gly Lys Ala Leu Leu Ala
1               5                  10                  15

His Phe Leu Ser Lys Ala Pro Val Trp Ala Glu Gln Gln Pro Ile Asn
            20                  25                  30

Thr Phe Glu Met Gly Thr Pro Lys Leu Ala Ser Leu Thr Phe Glu Asp
        35                  40                  45

Gly Val Ala Pro Glu Gln Ile Phe Ala Ala Ala Glu Lys Thr Tyr Pro
    50                  55                  60

Trp Leu Leu Glu Ser Gly Ala Lys Phe Val Ala Lys Pro Asp Gln Leu
65                  70                  75                  80

Ile Lys Arg Arg Gly Lys Ala Gly Leu Leu Val Leu Asn Lys Ser Trp
                85                  90                  95

Glu Glu Cys Lys Pro Trp Ile Ala Glu Arg Ala Ala Lys Pro Ile Asn
            100                 105                 110

Val Glu Gly Ile Asp Gly Val Leu Arg Thr Phe Leu Val Glu Pro Phe
        115                 120                 125

Val Pro His Asp Gln Lys His Glu Tyr Tyr Ile Asn Ile His Ser Val
    130                 135                 140

Arg Glu Gly Asp Trp Ile Leu Phe Tyr His Glu Gly Gly Val Asp Val
145                 150                 155                 160

Gly Asp Val Asp Ala Lys Ala Ala Lys Ile Leu Ile Pro Val Asp Ile
                165                 170                 175

Glu Asn Glu Tyr Pro Ser Asn Ala Thr Leu Thr Lys Glu Leu Leu Ala
            180                 185                 190

His Val Pro Glu Asp Gln His Gln Thr Leu Leu Asp Phe Ile Asn Arg
        195                 200                 205

Leu Tyr Ala Val Tyr Val Asp Leu Gln Phe Thr Tyr Leu Glu Ile Asn
    210                 215                 220

Pro Leu Val Val Ile Pro Thr Ala Gln Gly Val Glu Val His Tyr Leu
225                 230                 235                 240
```


LCY Biotechnology Holding, Inc.
Ex. 1002
Page 156 of 3077

```
Asp Leu Ala Gly Lys Leu Asp Gln Thr Ala Glu Phe Glu Cys Gly Pro
                245                 250                 255

Lys Trp Ala Ala Ala Arg Ser Pro Ala Ala Leu Gly Gln Val Val Thr
            260                 265                 270

Ile Asp Ala Gly Ser Thr Lys Val Ser Ile Asp Ala Gly Pro Ala Met
        275                 280                 285

Val Phe Pro Ala Pro Phe Gly Arg Glu Leu Ser Lys Glu Glu Ala Tyr
    290                 295                 300

Ile Ala Glu Leu Asp Ser Lys Thr Gly Ala Ser Leu Lys Leu Thr Val
305                 310                 315                 320

Leu Asn Ala Lys Gly Arg Ile Trp Thr Leu Val Ala Gly Gly Gly Ala
                325                 330                 335

Ser Val Val Tyr Ala Asp Ala Ile Ala Ser Ala Gly Phe Ala Asp Glu
            340                 345                 350

Leu Ala Asn Tyr Gly Glu Tyr Ser Gly Ala Pro Asn Glu Thr Gln Thr
        355                 360                 365

Tyr Glu Tyr Ala Lys Thr Val Leu Asp Leu Met Thr Arg Gly Asp Ala
    370                 375                 380

His Pro Glu Gly Lys Val Leu Phe Ile Gly Gly Gly Ile Ala Asn Phe
385                 390                 395                 400

Thr Gln Val Gly Ser Thr Phe Lys Gly Ile Ile Arg Ala Phe Arg Asp
                405                 410                 415

Tyr Gln Ser Ser Leu His Asn His Lys Val Lys Ile Tyr Val Arg Arg
            420                 425                 430

Gly Gly Pro Asn Trp Gln Glu Gly Leu Arg Leu Ile Lys Ser Ala Gly
        435                 440                 445

Asp Glu Leu Asn Leu Pro Met Glu Ile Tyr Gly Pro Asp Met His Val
    450                 455                 460

Ser Gly Ile Val Pro Leu Ala Leu Leu Gly Lys Arg Pro Lys Asn Val
465                 470                 475                 480

Lys Pro Phe Gly Thr Gly Pro Ser Thr Glu Ala Ser Thr Pro Leu Gly
                485                 490                 495
```


LCY Biotechnology Holding, Inc.
Ex. 1002
Page 157 of 3077

Val

<210> 44
<211> 650
<212> PRT
<213> Candida viswanathii

<400> 44
Met Ser Gly Leu Glu Ile Ala Ala Ala Ala Ile Leu Gly Ser Gln Leu
1 5 10 15

Leu Glu Ala Lys Tyr Leu Ile Ala Asp Asp Val Ser Leu Ala Lys Thr
20 25 30

Val Ala Val Asn Ala Leu Pro Tyr Leu Trp Lys Ala Ser Arg Gly Lys
35 40 45

Ala Ser Tyr Trp Tyr Phe Phe Glu Gln Ser Val Phe Lys Asn Pro Asn
50 55 60

Asn Lys Ala Leu Ala Phe Pro Arg Pro Arg Lys Asn Ala Pro Thr Pro
65 70 75 80

Lys Thr Asp Ala Glu Gly Phe Gln Ile Tyr Asp Asp Gln Phe Asp Leu
85 90 95

Glu Glu Tyr Thr Tyr Lys Glu Leu Tyr Asp Met Val Leu Lys Tyr Ser
100 105 110

Tyr Ile Leu Lys Asn Glu Tyr Gly Val Thr Ala Asn Asp Thr Ile Gly
115 120 125

Val Ser Cys Met Asn Lys Pro Leu Phe Ile Val Leu Trp Leu Ala Leu
130 135 140

Trp Asn Ile Gly Ala Leu Pro Ala Phe Leu Asn Phe Asn Thr Lys Asp
145 150 155 160

Lys Pro Leu Ile His Cys Leu Lys Ile Val Asn Ala Ser Gln Val Phe
165 170 175

Val Asp Pro Asp Cys Asp Ser Pro Ile Arg Asp Thr Glu Ala Gln Ile
180 185 190

Arg Glu Glu Leu Pro His Val Gln Ile Asn Tyr Ile Asp Glu Phe Ala
195 200 205

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 158 of 3077

```
Leu Phe Asp Arg Leu Arg Leu Lys Ser Thr Pro Lys His Arg Ala Glu
    210                 215                 220

Asp Lys Thr Arg Arg Pro Thr Asp Thr Asp Ser Ser Ala Cys Ala Leu
225                 230                 235                 240

Ile Tyr Thr Ser Gly Thr Thr Gly Leu Pro Lys Ala Gly Ile Met Ser
                245                 250                 255

Trp Arg Lys Ala Phe Met Ala Ser Val Phe Phe Gly His Ile Met Lys
            260                 265                 270

Ile Asp Ser Lys Ser Asn Val Leu Thr Ala Met Pro Leu Tyr His Ser
        275                 280                 285

Thr Ala Ala Met Leu Gly Leu Cys Pro Thr Leu Ile Val Gly Gly Cys
    290                 295                 300

Val Ser Val Ser Gln Lys Phe Ser Ala Thr Ser Phe Trp Thr Gln Ala
305                 310                 315                 320

Arg Leu Cys Gly Ala Thr His Val Gln Tyr Val Gly Glu Val Cys Arg
                325                 330                 335

Tyr Leu Leu Asn Ser Lys Pro His Pro Asp Gln Asp Arg His Asn Val
            340                 345                 350

Arg Ile Ala Tyr Gly Asn Gly Leu Arg Pro Asp Ile Trp Ser Glu Phe
        355                 360                 365

Lys Arg Arg Phe His Ile Glu Gly Ile Gly Glu Phe Tyr Ala Ala Thr
    370                 375                 380

Glu Ser Pro Ile Ala Thr Thr Asn Leu Gln Tyr Gly Glu Tyr Gly Val
385                 390                 395                 400

Gly Ala Cys Arg Lys Tyr Gly Ser Leu Ile Ser Leu Leu Leu Ser Thr
                405                 410                 415

Gln Gln Lys Leu Ala Lys Met Asp Pro Glu Asp Glu Ser Glu Ile Tyr
            420                 425                 430

Lys Asp Pro Lys Thr Gly Phe Cys Thr Glu Ala Ala Tyr Asn Glu Pro
        435                 440                 445

Gly Glu Leu Leu Met Arg Ile Leu Asn Pro Asn Asp Val Gln Lys Ser
    450                 455                 460
```


LCY Biotechnology Holding, Inc.
Ex. 1002
Page 159 of 3077

```
Phe Gln Gly Tyr Tyr Gly Asn Lys Ser Ala Thr Asn Ser Lys Ile Leu
465                 470                 475                 480

Thr Asn Val Phe Lys Lys Gly Asp Ala Trp Tyr Arg Ser Gly Asp Leu
                485                 490                 495

Leu Lys Met Asp Glu Asp Lys Leu Leu Tyr Phe Val Asp Arg Leu Gly
            500                 505                 510

Asp Thr Phe Arg Trp Lys Ser Glu Asn Val Ser Ala Thr Glu Val Glu
        515                 520                 525

Asn Glu Leu Met Gly Ser Lys Ala Leu Lys Gln Ser Val Val Val Gly
    530                 535                 540

Val Lys Val Pro Asn His Glu Gly Arg Ala Cys Phe Ala Val Cys Glu
545                 550                 555                 560

Ala Lys Asp Glu Leu Ser His Glu Glu Ile Leu Lys Leu Ile His Ser
                565                 570                 575

His Val Thr Lys Ser Leu Pro Val Tyr Ala Gln Pro Ala Phe Ile Lys
            580                 585                 590

Ile Gly Thr Ile Glu Ala Ser His Asn His Lys Val Pro Lys Asn Gln
        595                 600                 605

Phe Lys Asn Gln Lys Leu Pro Lys Gly Glu Asp Gly Lys Asp Leu Ile
    610                 615                 620

Tyr Trp Leu Asn Gly Asp Lys Tyr Gln Glu Leu Thr Glu Asp Asp Trp
625                 630                 635                 640

Ser Leu Ile Cys Thr Gly Lys Ala Lys Leu
                645                 650


<210> 45
<211> 517
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
      polypeptide

<400> 45
Met Ile Glu Gln Leu Leu Glu Tyr Trp Tyr Val Val Val Pro Val Leu
1               5                   10                  15
```

```
Tyr Ile Ile Lys Gln Leu Leu Ala Tyr Thr Lys Thr Arg Val Leu Met
            20                  25                  30

Lys Lys Leu Gly Ala Ala Pro Val Thr Asn Lys Leu Tyr Asp Asn Ala
        35                  40                  45

Phe Gly Ile Val Asn Gly Trp Lys Ala Leu Gln Phe Lys Lys Glu Gly
    50                  55                  60

Arg Ala Gln Glu Tyr Asn Asp Tyr Lys Phe Asp His Ser Lys Asn Pro
65                  70                  75                  80

Ser Val Gly Thr Tyr Val Ser Ile Leu Phe Gly Thr Arg Ile Val Val
                85                  90                  95

Thr Lys Asp Pro Glu Asn Ile Lys Ala Ile Leu Ala Thr Gln Phe Gly
            100                 105                 110

Asp Phe Ser Leu Gly Lys Arg His Thr Leu Phe Lys Pro Leu Leu Gly
        115                 120                 125

Asp Gly Ile Phe Thr Leu Asp Gly Glu Gly Trp Lys His Ser Arg Ala
    130                 135                 140

Met Leu Arg Pro Gln Phe Ala Arg Glu Gln Val Ala His Val Thr Ser
145                 150                 155                 160

Leu Glu Pro His Phe Gln Leu Leu Lys Lys His Ile Leu Lys His Lys
                165                 170                 175

Gly Glu Tyr Phe Asp Ile Gln Glu Leu Phe Phe Arg Phe Thr Val Asp
            180                 185                 190

Ser Ala Thr Glu Phe Leu Phe Gly Glu Ser Val His Ser Leu Lys Asp
        195                 200                 205

Glu Ser Ile Gly Ile Asn Gln Asp Asp Ile Asp Phe Ala Gly Arg Lys
    210                 215                 220

Asp Phe Ala Glu Ser Phe Asn Lys Ala Gln Glu Tyr Leu Ala Ile Arg
225                 230                 235                 240

Thr Leu Val Gln Thr Phe Tyr Trp Leu Val Asn Asn Lys Glu Phe Arg
                245                 250                 255

Asp Cys Thr Lys Ser Val His Lys Phe Thr Asn Tyr Tyr Val Gln Lys
            260                 265                 270
```

```
Ala Leu Asp Ala Ser Pro Glu Glu Leu Glu Lys Gln Ser Gly Tyr Val
        275                 280                 285

Phe Leu Tyr Glu Leu Val Lys Gln Thr Arg Asp Pro Asn Val Leu Arg
    290                 295                 300

Asp Gln Ser Leu Asn Ile Leu Leu Ala Gly Arg Asp Thr Thr Ala Gly
305                 310                 315                 320

Leu Leu Ser Phe Ala Val Phe Glu Leu Ala Arg His Pro Glu Ile Trp
                325                 330                 335

Ala Lys Leu Arg Glu Glu Ile Glu Gln Gln Phe Gly Leu Gly Glu Asp
            340                 345                 350

Ser Arg Val Glu Glu Ile Thr Phe Glu Ser Leu Lys Arg Cys Glu Tyr
        355                 360                 365

Leu Lys Ala Phe Leu Asn Glu Thr Leu Arg Ile Tyr Pro Ser Val Pro
    370                 375                 380

Arg Asn Phe Arg Ile Ala Thr Lys Asn Thr Thr Leu Pro Arg Gly Gly
385                 390                 395                 400

Gly Ser Asp Gly Thr Ser Pro Ile Leu Ile Gln Lys Gly Glu Ala Val
                405                 410                 415

Ser Tyr Gly Ile Asn Ser Thr His Leu Asp Pro Val Tyr Tyr Gly Pro
            420                 425                 430

Asp Ala Ala Glu Phe Arg Pro Glu Arg Trp Phe Glu Pro Ser Thr Lys
        435                 440                 445

Lys Leu Gly Trp Ala Tyr Leu Pro Phe Asn Gly Gly Pro Arg Ile Cys
    450                 455                 460

Leu Gly Gln Gln Phe Ala Leu Thr Glu Ala Gly Tyr Val Leu Val Arg
465                 470                 475                 480

Leu Val Gln Glu Phe Ser His Val Arg Ser Asp Pro Asp Glu Val Tyr
                485                 490                 495

Pro Pro Lys Arg Leu Thr Asn Leu Thr Met Cys Leu Gln Asp Gly Ala
            500                 505                 510

Ile Val Lys Phe Asp
        515
```


LCY Biotechnology Holding, Inc.
Ex. 1002
Page 162 of 3077

<210> 46
<211> 418
<212> PRT
<213> Candida viswanathii

<400> 46
Met His Ala Leu Phe Ser Lys Ser Val Phe Leu Lys Tyr Val Ser Ser
1 5 10 15

Pro Thr Thr Ser Ala Ile Pro His Ser Leu Glu Phe Ile Val Ser Arg
20 25 30

Ser Ser Tyr Leu Arg Arg Arg Ile Pro Pro Tyr Leu Pro Arg Cys Ser
35 40 45

His Phe Pro Ser Phe Tyr Tyr Ser Ser Ser Ser Val Tyr Thr Lys Lys
50 55 60

Ser Phe His Thr Met Ser Ala Asn Ile Pro Lys Thr Gln Lys Ala Val
65 70 75 80

Val Phe Glu Lys Asn Gly Gly Glu Leu Lys Tyr Lys Asp Ile Pro Val
85 90 95

Pro Thr Pro Lys Ala Asn Glu Leu Leu Ile Asn Val Lys Tyr Ser Gly
100 105 110

Val Cys His Thr Asp Leu His Ala Trp Lys Gly Asp Trp Pro Leu Asp
115 120 125

Thr Lys Leu Pro Leu Val Gly Gly His Glu Gly Ala Gly Val Val Val
130 135 140

Gly Met Gly Glu Asn Val Lys Gly Trp Lys Ile Gly Asp Phe Ala Gly
145 150 155 160

Ile Lys Trp Leu Asn Gly Ser Cys Met Ser Cys Glu Phe Cys Gln Gln
165 170 175

Gly Ala Glu Pro Asn Cys Gly Glu Ala Asp Leu Ser Gly Tyr Thr His
180 185 190

Asp Gly Ser Phe Glu Gln Tyr Ala Thr Ala Asp Ala Val Gln Ala Ala
195 200 205

Arg Ile Pro Ala Gly Thr Asp Leu Ala Glu Val Ala Pro Ile Leu Cys
210 215 220

```
Ala Gly Val Thr Val Tyr Lys Ala Leu Lys Thr Ala Asp Leu Ala Ala
225                 230                 235                 240

Gly Gln Trp Val Ala Ile Ser Gly Ala Gly Gly Gly Leu Gly Ser Leu
                245                 250                 255

Ala Val Gln Tyr Ala Val Ala Met Gly Leu Arg Val Val Ala Ile Asp
            260                 265                 270

Gly Gly Asp Glu Lys Gly Asp Phe Val Lys Ser Leu Gly Ala Glu Ala
        275                 280                 285

Tyr Ile Asp Phe Leu Lys Glu Lys Gly Ile Val Ala Ala Val Lys Lys
    290                 295                 300

Ala Thr Asp Gly Gly Pro His Gly Ala Ile Asn Val Ser Val Ser Glu
305                 310                 315                 320

Lys Ala Ile Asp Gln Ser Val Glu Tyr Val Arg Pro Leu Gly Lys Val
                325                 330                 335

Val Leu Val Gly Leu Pro Ala Gly Ser Lys Val Thr Ala Gly Val Phe
            340                 345                 350

Glu Ala Val Val Lys Ser Ile Glu Ile Lys Gly Ser Tyr Val Gly Asn
        355                 360                 365

Arg Lys Asp Thr Ala Glu Ala Val Asp Phe Phe Ser Arg Gly Leu Ile
    370                 375                 380

Lys Cys Pro Ile Lys Ile Val Gly Leu Ser Glu Leu Pro Gln Val Phe
385                 390                 395                 400

Lys Leu Met Glu Glu Gly Lys Ile Leu Gly Arg Tyr Val Leu Asp Thr
                405                 410                 415

Ser Lys
```

<210> 47
<211> 704
<212> PRT
<213> Candida viswanathii

<400> 47

```
Met Ala Pro Phe Leu Pro Asp Gln Val Asp Tyr Lys His Val Asp Thr
1               5                   10                  15
```


LCY Biotechnology Holding, Inc.
Ex. 1002
Page 164 of 3077

```
Leu Met Leu Leu Cys Asp Gly Ile Ile His Glu Thr Thr Val Asp Gln
            20                  25                  30

Ile Lys Asp Val Ile Ala Pro Asp Phe Pro Ala Asp Lys Tyr Glu Glu
        35                  40                  45

Tyr Val Arg Thr Phe Thr Lys Pro Ser Glu Thr Pro Gly Phe Arg Glu
    50                  55                  60

Thr Val Tyr Asn Thr Val Asn Ala Asn Thr Thr Asp Ala Ile His Gln
65                  70                  75                  80

Phe Ile Ile Leu Thr Asn Val Leu Ala Ser Arg Val Leu Ala Pro Ala
                85                  90                  95

Leu Thr Asn Ser Leu Thr Pro Ile Lys Asp Met Ser Leu Glu Asp Arg
            100                 105                 110

Glu Lys Leu Leu Ala Ser Trp Arg Asp Ser Pro Ile Ala Ala Lys Arg
        115                 120                 125

Lys Leu Phe Arg Leu Val Ser Thr Leu Thr Leu Val Thr Phe Thr Arg
    130                 135                 140

Leu Ala Asn Glu Leu His Leu Lys Ala Ile His Tyr Pro Gly Arg Glu
145                 150                 155                 160

Asp Arg Glu Lys Ala Tyr Glu Thr Gln Glu Ile Asp Pro Phe Lys Tyr
                165                 170                 175

Gln Phe Leu Glu Lys Pro Lys Phe Tyr Gly Ala Glu Leu Tyr Leu Pro
            180                 185                 190

Asp Ile Asp Val Ile Ile Ile Gly Ser Gly Ala Gly Ala Gly Val Val
        195                 200                 205

Ala His Thr Leu Ala Asn Asp Gly Phe Lys Ser Leu Val Leu Glu Lys
    210                 215                 220

Gly Lys Tyr Phe Ser Asn Ser Glu Leu Asn Phe Asp Asp Lys Asp Gly
225                 230                 235                 240

Val Gln Glu Leu Tyr Gln Ser Gly Gly Thr Leu Thr Thr Val Asn Gln
                245                 250                 255

Gln Leu Phe Val Leu Ala Gly Ser Thr Phe Gly Gly Gly Thr Thr Val
            260                 265                 270
```


LCY Biotechnology Holding, Inc.
Ex. 1002
Page 165 of 3077

```
Asn Trp Ser Ala Cys Leu Lys Thr Pro Phe Lys Val Arg Lys Glu Trp
        275                 280                 285

Tyr Asp Glu Phe Gly Val Asp Phe Ala Ala Asp Glu Ala Tyr Asp Lys
    290                 295                 300

Ala Gln Asp Tyr Val Trp Gln Gln Met Gly Ala Ser Thr Glu Gly Ile
305                 310                 315                 320

Thr His Ser Leu Ala Asn Glu Ile Ile Ile Glu Gly Gly Lys Lys Leu
                325                 330                 335

Gly Tyr Lys Ala Lys Val Leu Asp Gln Asn Ser Gly Gly His Pro Gln
            340                 345                 350

His Arg Cys Gly Phe Cys Tyr Leu Gly Cys Lys His Gly Ile Lys Gln
        355                 360                 365

Gly Ser Val Asn Asn Trp Phe Arg Asp Ala Ala Ala His Gly Ser Gln
    370                 375                 380

Phe Met Gln Gln Val Arg Val Leu Gln Ile Leu Asn Lys Lys Gly Ile
385                 390                 395                 400

Ala Tyr Gly Ile Leu Cys Glu Asp Val Val Thr Gly Ala Lys Phe Thr
                405                 410                 415

Ile Thr Gly Pro Lys Lys Phe Val Val Ala Ala Gly Ala Leu Asn Thr
            420                 425                 430

Pro Ser Val Leu Val Asn Ser Gly Phe Lys Asn Lys Asn Ile Gly Lys
        435                 440                 445

Asn Leu Thr Leu His Pro Val Ser Val Val Phe Gly Asp Phe Gly Lys
    450                 455                 460

Asp Val Gln Ala Asp His Phe His Asn Ser Ile Met Thr Ala Leu Cys
465                 470                 475                 480

Ser Glu Ala Ala Asp Leu Asp Gly Lys Gly His Gly Cys Arg Ile Glu
                485                 490                 495

Thr Ile Leu Asn Ala Pro Phe Ile Gln Ala Ser Phe Leu Pro Trp Arg
            500                 505                 510

Gly Ser Asn Glu Ala Arg Arg Asp Leu Leu Arg Tyr Asn Asn Met Val
        515                 520                 525
```

```
Ala Met Leu Leu Leu Ser Arg Asp Thr Thr Ser Gly Ser Val Ser Ala
    530                 535                 540

His Pro Thr Lys Pro Glu Ala Leu Val Val Glu Tyr Asp Val Asn Lys
545                 550                 555                 560

Phe Asp Arg Asn Ser Ile Leu Gln Ala Leu Leu Val Thr Ala Asp Leu
                565                 570                 575

Leu Tyr Ile Gln Gly Ala Lys Arg Ile Leu Ser Pro Gln Ala Trp Val
            580                 585                 590

Pro Ile Phe Glu Ser Asp Lys Pro Lys Asp Lys Arg Ser Ile Lys Asp
        595                 600                 605

Glu Asp Tyr Val Glu Trp Arg Ala Lys Val Ala Lys Ile Pro Phe Asp
    610                 615                 620

Thr Tyr Gly Ser Pro Tyr Gly Ser Ala His Gln Met Ser Ser Cys Arg
625                 630                 635                 640

Met Ser Gly Lys Gly Pro Lys Tyr Gly Ala Val Asp Thr Asp Gly Arg
                645                 650                 655

Leu Phe Glu Cys Ser Asn Val Tyr Val Ala Asp Ala Ser Leu Leu Pro
            660                 665                 670

Thr Ala Ser Gly Ala Asn Pro Met Val Thr Thr Met Thr Leu Ala Arg
        675                 680                 685

His Val Ala Leu Gly Leu Ala Asp Ser Leu Lys Thr Lys Ala Lys Leu
    690                 695                 700


<210> 48
<211> 342
<212> PRT
<213> Candida viswanathii

<400> 48
Met Leu Arg Ser Ser Val Arg Thr Phe Ser Thr Gln Ser Arg Val Leu
1               5                   10                  15

Ala Asn Tyr Gly Phe Val Gly Leu Gly Leu Met Gly Gln His Met Ala
            20                  25                  30

Arg His Val Tyr Asn Gln Leu Gln Pro Ala Asp Lys Leu Tyr Val His
        35                  40                  45
```

```
Asp Val Asn Pro Gln His Thr Thr Gln Phe Val Thr Asp Val Thr Thr
    50                  55                  60

Gln Lys Pro Gln Asn Ala Thr Gln Leu Thr Pro Leu Ser Ser Leu Lys
65                  70                  75                  80

Glu Phe Thr Thr Glu Pro Glu Ser Gln Leu Asp Phe Ile Val Thr Met
                85                  90                  95

Val Pro Glu Gly Lys His Val Lys Ala Val Val Ser Glu Leu Val Asp
            100                 105                 110

His Tyr Asn Ala Ser Gly Lys Tyr Asp Pro Ser Lys Lys Leu Thr Phe
        115                 120                 125

Val Asp Ser Ser Thr Ile Asp Ile Pro Thr Ser Arg Glu Val His Gln
    130                 135                 140

Leu Val Ala Asp Lys Leu Gln Gly Ala Thr Phe Ile Asp Ala Pro Val
145                 150                 155                 160

Ser Gly Gly Val Ala Gly Ala Arg Asn Gly Thr Leu Ser Phe Met Val
                165                 170                 175

Ser Arg Asp Thr Lys Glu Asp Val Asp Pro Asn Leu Val Thr Leu Leu
            180                 185                 190

Asn Tyr Met Gly Ser Asn Ile Phe Pro Cys Gly Gly Thr His Gly Thr
        195                 200                 205

Gly Leu Ala Ala Lys Leu Ala Asn Asn Tyr Leu Leu Ala Ile Thr Asn
    210                 215                 220

Ile Ala Val Ala Asp Ser Phe Gln Leu Ala Asn Ser Phe Gly Leu Asn
225                 230                 235                 240

Leu Gln Asn Tyr Ala Lys Leu Val Ser Thr Ser Thr Gly Lys Ser Trp
                245                 250                 255

Ala Ser Val Asp Asn Cys Pro Ile Pro Gly Val Tyr Pro Glu Lys Asn
            260                 265                 270

Leu Thr Cys Asp Asn Gly Tyr Lys Gly Gly Phe Val Thr Lys Leu Thr
        275                 280                 285

Arg Lys Asp Val Val Leu Ala Thr Glu Ser Ala Lys Ala Asn Asn Gln
    290                 295                 300
```


LCY Biotechnology Holding, Inc.
Ex. 1002
Page 168 of 3077

```
Phe Leu Met Leu Gly Glu Val Gly Arg Tyr Trp Tyr Asp Lys Ala Cys
305                 310                 315                 320

Glu Asp Glu Lys Tyr Ala Asn Arg Asp Leu Ser Val Leu Phe Glu Phe
                325                 330                 335

Leu Gly Asp Leu Lys Lys
            340
```

<210> 49
<211> 549
<212> PRT
<213> Candida viswanathii

<400> 49

```
Met Leu Ser Arg Val Leu Phe Lys Thr Lys Pro Arg Val Pro Thr Lys
1               5                   10                  15

Ser Ile Thr Ala Met Ala Ile Arg Asn Lys Ser Ile Val Thr Leu Ser
            20                  25                  30

Ser Thr Thr Ser Thr Tyr Pro Thr Asp His Thr Thr Pro Ser Thr Glu
        35                  40                  45

Pro Tyr Ile Thr Pro Ser Phe Val Asn Asn Glu Phe Ile Lys Ser Asp
    50                  55                  60

Ser Asn Thr Trp Phe Asp Val His Asp Pro Ala Thr Asn Tyr Val Val
65                  70                  75                  80

Ser Lys Val Pro Gln Ser Thr Pro Glu Glu Leu Glu Glu Ala Ile Ala
                85                  90                  95

Ser Ala His Ala Ala Phe Pro Lys Trp Arg Asp Thr Ser Ile Ile Lys
            100                 105                 110

Arg Gln Gly Ile Ala Phe Lys Phe Val Gln Leu Leu Arg Glu Asn Met
        115                 120                 125

Asp Arg Ile Ala Ser Val Ile Val Leu Glu Gln Gly Lys Thr Phe Val
    130                 135                 140

Asp Ala Gln Gly Asp Val Thr Arg Gly Leu Gln Val Ala Glu Ala Ala
145                 150                 155                 160

Cys Asn Ile Thr Asn Asp Leu Lys Gly Glu Ser Leu Glu Val Ser Thr
                165                 170                 175
```


LCY Biotechnology Holding, Inc.
Ex. 1002
Page 169 of 3077

```
Asp Met Glu Thr Lys Met Ile Arg Glu Pro Leu Gly Val Val Gly Ser
            180                 185                 190

Ile Cys Pro Phe Asn Phe Pro Ala Met Val Pro Leu Trp Ser Leu Pro
        195                 200                 205

Leu Val Leu Val Thr Gly Asn Thr Ala Val Ile Lys Pro Ser Glu Arg
    210                 215                 220

Val Pro Gly Ala Ser Met Ile Ile Cys Glu Leu Ala Ala Lys Ala Gly
225                 230                 235                 240

Val Pro Pro Gly Val Leu Asn Ile Val His Gly Lys His Asp Thr Val
                245                 250                 255

Asn Lys Leu Ile Glu Asp Pro Arg Ile Lys Ala Leu Thr Phe Val Gly
            260                 265                 270

Gly Asp Lys Ala Gly Lys Tyr Ile Tyr Glu Lys Gly Ser Ser Leu Gly
        275                 280                 285

Lys Arg Val Gln Ala Asn Leu Gly Ala Lys Asn His Leu Val Val Leu
    290                 295                 300

Pro Asp Ala His Lys Gln Ser Phe Val Asn Ala Val Asn Gly Ala Ala
305                 310                 315                 320

Phe Gly Ala Ala Gly Gln Arg Cys Met Ala Ile Ser Val Leu Val Thr
                325                 330                 335

Val Gly Lys Thr Lys Glu Trp Val Gln Asp Val Ile Lys Asp Ala Lys
            340                 345                 350

Leu Leu Asn Thr Gly Ser Gly Phe Asp Pro Lys Ser Asp Leu Gly Pro
        355                 360                 365

Val Ile Asn Pro Glu Ser Leu Thr Arg Ala Glu Glu Ile Ile Ala Asp
    370                 375                 380

Ser Val Ala Asn Gly Ala Val Leu Glu Leu Asp Gly Arg Gly Tyr Arg
385                 390                 395                 400

Pro Glu Asp Ala Arg Phe Ala Lys Gly Asn Phe Leu Gly Pro Thr Ile
                405                 410                 415

Leu Thr Asn Val Lys Pro Gly Leu Arg Ala Tyr Asp Glu Glu Ile Phe
            420                 425                 430
```


LCY Biotechnology Holding, Inc.
Ex. 1002
Page 170 of 3077

Ala Pro Val Leu Ser Val Val Asn Val Asp Thr Ile Asp Glu Ala Ile
435 440 445

Glu Leu Ile Asn Asn Asn Lys Tyr Gly Asn Gly Val Ser Leu Phe Thr
450 455 460

Ser Ser Gly Gly Ser Ala Gln Tyr Phe Thr Lys Arg Ile Asp Val Gly
465 470 475 480

Gln Val Gly Ile Asn Val Pro Ile Pro Val Pro Leu Pro Met Phe Ser
485 490 495

Phe Thr Gly Ser Arg Gly Ser Phe Leu Gly Asp Leu Asn Phe Tyr Gly
500 505 510

Lys Ala Gly Ile Thr Phe Leu Thr Lys Pro Lys Thr Ile Thr Ser Ala
515 520 525

Trp Lys Thr Asn Leu Ile Asp Asp Glu Ile Leu Lys Pro Ser Thr Ser
530 535 540

Met Pro Val Gln Gln
545

<210> 50
<211> 241
<212> PRT
<213> Candida viswanathii

<400> 50
Met Ser Asp Glu Glu Ser Asp Ile Leu Tyr Glu Val Arg Asp Arg Thr
1 5 10 15

Ala Ile Ile Thr Leu Asn Ile Pro Lys Arg Leu Asn Ala Leu Asn Gly
20 25 30

Ala Gln Tyr Leu Lys Leu Gly Lys Phe Leu Glu Arg Ala Asn Asn Glu
35 40 45

Glu Asp Thr Val Leu Thr Leu Ile Gln Ala Ser Gly Arg Phe Phe Ser
50 55 60

Ala Gly Ala Asn Phe Ala Asp Asn Asp Met Ala Lys Val Glu Met Ser
65 70 75 80

Lys Leu Phe Ser His Glu Tyr Trp Leu Glu Arg Phe Val Ala Arg Asn
85 90 95

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 171 of 3077

```
Ile Trp Leu Thr Asn Leu Phe Asn Asp His Lys Lys Ile Leu Ala Ala
            100                 105                 110

Ala Val Asn Gly Pro Val Ile Gly Leu Ser Thr Gly Leu Leu Leu Leu
        115                 120                 125

Val Asp Leu Val Tyr Val His Asp Leu Asn Lys Phe Tyr Leu Leu Ala
    130                 135                 140

Pro Phe Ala Asn Leu Gly Leu Val Ala Glu Gly Ala Ser Ser Ala Thr
145                 150                 155                 160

Leu Phe Asn Arg Leu Gly Trp Ser Lys Ala Ser Glu Ala Leu Leu Leu
                165                 170                 175

Ala Lys Pro Ile Gly Gly Gln Asp Cys Tyr Asn Ala Gly Phe Ile Asn
            180                 185                 190

Lys His Tyr Asp Gly Lys Phe Ser Ser Thr Glu Glu Phe Asn Glu His
        195                 200                 205

Val Tyr Lys Glu Leu Thr Glu Ala Phe Glu Asn Leu His Asp Asp Ser
    210                 215                 220

Ile Leu Gln Asn Lys Gln Leu Leu Lys Leu Ser Arg Asp Gln Ala Ile
225                 230                 235                 240

Asn
```

<210> 51
<211> 268
<212> PRT
<213> Candida viswanathii

<400> 51

```
Met Ser Asp Asp Leu Ile Thr Tyr Glu Val Lys Asp Arg Ala Ala Val
1               5                   10                  15

Ile Thr Leu Asn Asn Pro Lys Lys Leu Asn Ala Leu Ser Ile Pro Gln
            20                  25                  30

Tyr Asp Thr Ile Cys Lys Leu Leu Glu Arg Ala Asn Ala Glu Glu Asp
        35                  40                  45

Thr Val Ile Thr Leu Leu Gln Ser Thr Gly Arg Val Phe Ser Ala Gly
    50                  55                  60
```

```
Ala Asn Ala Asp Ser Ile Val Gly Gln Asp Ala Glu Leu Glu Thr Trp
65                  70                  75                  80

Leu Asn Met Ser Val Ala Lys Gln Thr Phe Leu Val Gln Thr Phe Leu
                85                  90                  95

Ala His Lys Lys Ile Leu Ala Val Ala Leu Asn Gly Pro Val Ile Gly
            100                 105                 110

Leu Ser Ala Ala Phe Val Ala Leu Cys Asp Leu Val Tyr Val His Asn
        115                 120                 125

Ala Ala Lys Thr Phe Phe Leu Thr Pro Phe Ala Asn Ile Gly Ile Leu
    130                 135                 140

Ala Glu Gly Gly Thr Ser Ala Thr Leu Pro Met Arg Val Gly Trp Ser
145                 150                 155                 160

Arg Ala Ala Glu Ala Leu Leu Leu Ser Lys Arg Ile Ser Gly Asp Asp
                165                 170                 175

Leu Gln Arg Ala Gly Phe Phe Asn Lys Asp Tyr Lys Gly Gln Phe Lys
            180                 185                 190

Ser Ala Glu Glu Phe Asn Glu Val Val Leu Lys Glu Leu Leu Asp Ala
        195                 200                 205

Thr Glu Asn Leu His Glu Asp Ser Ile Ile Gln Asn Lys Glu Leu Leu
    210                 215                 220

Lys Ala Ile Phe Lys Pro Lys Ile Ser Glu Val Asn Ser Gln Glu Val
225                 230                 235                 240

Ser Arg Gly Val Tyr Lys Trp Thr Ser Gly Val Pro Met Asp Arg Phe
                245                 250                 255

Lys Lys Leu Leu Asn Gly Glu Leu Lys His Lys Leu
            260                 265


<210> 52
<211> 291
<212> PRT
<213> Candida viswanathii

<400> 52
Met Pro Asn Thr Leu Asp His Asn Tyr Leu Lys Lys Ser Val Trp Lys
1               5                   10                  15

Ser Asp Ile Phe Ala Gly Lys Val Ile Phe Ile Thr Gly Gly Ala Gly
```

20 25 30

Thr Ile Cys Arg Val Gln Ala Glu Ala Met Val Leu Leu Gly Ala Asn
35 40 45

Ala Ala Ile Ile Gly Arg Asn Val Glu Lys Thr Glu Glu Ala Ala Lys
50 55 60

Glu Ile Ala Ser Leu Arg Pro Gly Ala Lys Val Ile Gly Ile Gly Ala
65 70 75 80

Val Asp Val Arg Lys Ile Gln Thr Ile Lys Asp Ala Val Asp Arg Thr
85 90 95

Val Ala Glu Leu Gly Arg Ile Asp Tyr Val Ile Ala Gly Ala Ala Gly
100 105 110

Asn Phe Leu Cys Asp Phe Asn His Leu Ser Ala Asn Ala Phe Lys Ser
115 120 125

Val Ile Asp Ile Asp Leu Leu Gly Ser Phe Asn Thr Val Lys Val Thr
130 135 140

Phe Asp Gln Leu Arg Lys Asn Lys Gly Ala Val Leu Phe Val Ser Ala
145 150 155 160

Thr Leu His Tyr Thr Gly Val Pro Met Gln Ser His Val Gly Ala Ala
165 170 175

Lys Ser Gly Val Asp Ala Leu Ser Asn Ala Leu Ala Val Glu Leu Gly
180 185 190

Pro Leu Gly Ile Arg Phe Asn Cys Ile Ala Pro Gly Ala Ile Ala Gly
195 200 205

Thr Glu Gly Met Ser Arg Leu Ala Pro Pro Thr Asp Thr Pro Leu Glu
210 215 220

Thr Lys Ile Pro Leu Gln Arg Gln Gly Thr Thr Glu Asp Ile Ala Asp
225 230 235 240

Ala Thr Val Phe Leu Phe Ser Pro Ala Ser Ser Tyr Ile Thr Gly Asp
245 250 255

Val Leu Val Val Asp Gly Ala Met Trp Gln Thr Gly Gly Gly Ile Leu
260 265 270

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 174 of 3077

Asn Asp Phe Tyr Pro Asp Ile Ile Ile His Gln Asn Ala Asp Pro Glu
275 280 285

Gly Lys Leu
290

<210> 53
<211> 292
<212> PRT
<213> Candida viswanathii

<400> 53
Met Pro Asn Thr Leu Thr Glu Ala Tyr Lys Gln Glu Ser Ser Trp Lys
1 5 10 15

Pro Asp Leu Phe Lys Gly Lys Val Val Phe Ile Thr Gly Gly Ala Gly
20 25 30

Ser Ile Cys Arg Val Gln Ala Glu Ala Met Val Leu Leu Gly Ala Asn
35 40 45

Ala Ala Ile Ile Gly Arg Asn Val Glu Lys Thr Glu Ser Ala Ala Lys
50 55 60

Glu Ile Ala Ser Leu Arg Ser Gly Ala Lys Val Leu Gly Ile Gly Gly
65 70 75 80

Ile Asp Val Arg Lys Val Asp Ser Leu Lys Ser Ala Val Asp Lys Thr
85 90 95

Val Ala Glu Leu Gly Arg Ile Asp Phe Val Ile Ala Gly Ala Ala Gly
100 105 110

Asn Phe Leu Cys Asp Phe Asn His Leu Ser Ser Asn Ala Phe Lys Ser
115 120 125

Ile Ile Asp Ile Asp Leu Leu Gly Ser Phe Asn Thr Val Lys Val Thr
130 135 140

Phe Asp Gln Leu Arg Lys Asn Lys Gly Ala Ile Leu Phe Val Ser Ala
145 150 155 160

Thr Leu His Tyr Tyr Gly Val Pro Phe Gln Ile Gly Val Gly Ala Ala
165 170 175

Lys Ala Gly Val Asp Ala Leu Ser Asn Ala Leu Ala Val Glu Leu Gly
180 185 190

Pro Leu Gly Ile Arg Ser Asn Cys Ile Ala Pro Gly Pro Ile Asp Gly

195 200 205

Thr Glu Gly Val Glu Arg Leu Val Arg Ala Ser Lys Ala Glu Ala Ala
210 215 220

Lys Lys Val Pro Leu Gln Arg Leu Gly Thr Thr Gln Asp Ile Ala Asp
225 230 235 240

Gly Thr Val Tyr Leu Phe Ser Pro Ala Ala Ser Phe Val Thr Gly Asp
245 250 255

Val Leu Val Val Asp Gly Ala Ser Trp Gln Ile Ser Ser Gly Val Gly
260 265 270

Ala Thr Asn Tyr Pro Val Ser Ile Leu Asn Ala Ile Asp Ala Pro Lys
275 280 285

Gly Gly Lys Leu
290

<210> 54
<211> 804
<212> DNA
<213> Candida viswanathii

<400> 54
atggttagca caaaaacata cacagaaagg gcatcagcac acccctccaa ggttgcccaa 60

cgtttattcc gcttaatgga gtccaaaaag accaacctct gcgcctcgat cgacgtgacc 120

acaaccgccg agttcctttc gctcatcgac aagctcggtc cccacatctg tctcgtgaag 180

acgcacatcg atatcatctc agacttcagc tacgagggca cgattgagcc gttgcttgtg 240

cttgcagagc gccacgggtt cttgatattc gaggacagga agtttgctga tatcggaaac 300

accgtgatgt tgcagtacac ctcgggggta taccggatcg cggcgtggag tgacatcacg 360

aacgcgcacg gagtgactgg gaagggcgtc gttgaagggt tgaaacgcgg tgcggagggg 420

gtagaaaagg aaaggggcgt gttgatgttg gcggagttgt cgagtaaagg ctcgttggcg 480

catggtgaat atacccgtga gacgatcgag attgcgaaga gtgatcggga gttcgtgatt 540

gggttcatcg cgcagcggga catggggggt agagaagaag ggtttgattg gatcatcatg 600

acgcctggtg tggggttgga tgataaaggc gatgcgttgg gccagcagta taggactgtt 660

gatgaggtgg ttctgactgg taccgatgtg attattgtcg ggagagggtt gtttggaaaa 720

ggaagagacc ctgaggtgga gggaaagaga tacagggatg ctggatggaa ggcatacttg 780

aagagaactg gtcagttaga ataa 804

<210> 55

<211> 415
<212> DNA
<213> Candida viswanathii

<400> 55
cgacgggtac aacgagaatt gtattgaatt gatcaagaac atgatcttgg tgttacagaa 60

catcaagttc ttggaccaga ctgagaatgc acagatatac aaggcgtcat gtgataaaat 120

ggatgagatt tatccacaat tgaagaaaga gtttatggaa agtggtcaac cagaagctaa 180

acaggaagaa gcaaacgaag aggtgaaaca agaagaagaa ggtaaataag tattttgtat 240

tatataacaa acaaagtaag gaatacagat ttatacaata aattgccata ctagtcacgt 300

gagatatctc atccattccc caactcccaa gaaaaaaaaa aagtgaaaaa aaaaatcaaa 360

cccaaagatc aacctcccca tcatcatcgt catcaaaccc ccagctcaat tcgca 415

<210> 56
<211> 267
<212> DNA
<213> Candida viswanathii

<400> 56
taaatattgt aataaatagg tctatataca tacactaagc ttctaggacg tcattgtagt 60

cttcgaagtt gtctgctagt ttagttctca tgatttcgaa aaccaataac gcaatggatg 120

tagcagggat ggtggttagt gcgttcctga caaacccaga gtacgccgcc tcaaaccacg 180

tcacattcgc cctttgcttc atccgcatca cttgcttgaa ggtatccacg tacgagttgt 240

aatacacctt gaagaacggc ttcgtct 267

<210> 57
<211> 1961
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 57
ttaaaggtat ctacggttgt ttcggtatga ccccccgggga tctgacgggt acaacgagaa 60

ttgtattgaa ttgatcaaga acatgatctt ggtgttacag aacatcaagt tcttggacca 120

gactgagaat gcacagatat acaaggcgtc atgtgataaa atggatgaga tttatccaca 180

attgaagaaa gagtttatgg aaagtggtca accagaagct aaacaggaag aagcaaacga 240

agaggtgaaa caagaagaag aaggtaaata agtattttgt attatataac aaacaaagta 300

aggaatacag atttatacaa taaattgcca tactagtcac gtgagatatc tcatccattc 360

cccaactccc aagaaaaaaa aaaagtgaaa aaaaaaatca aacccaaaga tcaacctccc 420

catcatcatc gtcatcaaac ccccagctca attcgcaatg gttagcacaa aaacatacac 480

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 177 of 3077

agaaagggca tcagcacacc cctccaaggt tgcccaacgt ttattccgct taatggagtc 540

caaaaagacc aacctctgcg cctcgatcga cgtgaccaca accgccgagt tcctttcgct 600

catcgacaag ctcggtcccc acatctgtct cgtgaagacg cacatcgata tcatctcaga 660

cttcagctac gagggcacga ttgagccgtt gcttgtgctt gcagagcgcc acgggttctt 720

gatattcgag gacaggaagt ttgctgatat cggaaacacc gtgatgttgc agtacacctc 780

gggggtatac cggatcgcgg cgtggagtga catcacgaac gcgcacggag tgactgggaa 840

gggcgtcgtt gaagggttga aacgcggtgc ggagggggta gaaaaggaaa ggggcgtgtt 900

gatgttggcg gagttgtcga gtaaaggctc gttggcgcat ggtgaatata cccgtgagac 960

gatcgagatt gcgaagagtg atcgggagtt cgtgattggg ttcatcgcgc agcgggacat 1020

gggggggtaga gaagaagggt ttgattggat catcatgacg cctggtgtgg ggttggatga 1080

taaaggcgat gcgttgggcc agcagtatag gactgttgat gaggtggttc tgactggtac 1140

cgatgtgatt attgtcggga gagggttgtt tggaaaagga agagaccctg aggtggaggg 1200

aaagagatac agggatgctg gatggaaggc atacttgaag agaactggtc agttagaata 1260

aatattgtaa taaataggtc tatatacata cactaagctt ctaggacgtc attgtagtct 1320

tcgaagttgt ctgctagttt agttctcatg atttcgaaaa ccaataacgc aatggatgta 1380

gcagggatgg tggttagtgc gttcctgaca aacccagagt acgccgcctc aaaccacgtc 1440

acattcgccc tttgcttcat ccgcatcact tgcttgaagg tatccacgta cgagttgtaa 1500

tacaccttga agaacggctt cgtctacggt cgacgacggg tacaacgaga attgtattga 1560

attgatcaag aacatgatct tggtgttaca gaacatcaag ttcttggacc agactgagaa 1620

tgcacagata tacaaggcgt catgtgataa aatggatgag atttatccac aattgaagaa 1680

agagtttatg gaaagtggtc aaccagaagc taaacaggaa gaagcaaacg aagaggtgaa 1740

acaagaagaa gaaggtaaat aagtattttg tattatataa caaacaaagt aaggaataca 1800

gatttataca ataaattgcc atactagtca cgtgagatat ctcatccatt ccccaactcc 1860

caagaaaaaa aaaaagtgaa aaaaaaaatc aaacccaaag atcaacctcc ccatcatcat 1920

cgtcatcaaa cccccagctc aattcgcaga gctcggtacc c 1961

<210> 58
<211> 1731
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 58
taaatattgt aataaatagg tctatataca tacactaagc ttctaggacg tcattgtagt 60

cttcgaagtt gtctgctagt ttagttctca tgatttcgaa aaccaataac gcaatggatg 120

tagcagggat ggtggttagt gcgttcctga caaacccaga gtacgccgcc tcaaaccacg 180

tcacattcgc cctttgcttc atccgcatca cttgcttgaa ggtatccacg tacgagttgt 240

aatacacctt gaagaacggc ttcgtctcga cgggtacaac gagaattgta ttgaattgat 300

caagaacatg atcttggtgt tacagaacat caagttcttg gaccagactg agaatgcaca 360

gatatacaag gcgtcatgtg ataaaatgga tgagatttat ccacaattga agaaagagtt 420

tatggaaagt ggtcaaccag aagctaaaca ggaagaagca aacgaagagg tgaaacaaga 480

agaagaaggt aaataagtat tttgtattat ataacaaaca aagtaaggaa tacagattta 540

tacaataaat tgccatacta gtcacgtgag atatctcatc cattccccaa ctcccaagaa 600

aaaaaaaaag tgaaaaaaaa aatcaaaccc aaagatcaac ctccccatca tcatcgtcat 660

caaaccccca gctcaattcg caatggttag cacaaaaaca tacacagaaa gggcatcagc 720

acacccctcc aaggttgccc aacgtttatt ccgcttaatg gagtccaaaa agaccaacct 780

ctgcgcctcg atcgacgtga ccacaaccgc cgagttcctt tcgctcatcg acaagctcgg 840

tccccacatc tgtctcgtga agacgcacat cgatatcatc tcagacttca gctacgaggg 900

cacgattgag ccgttgcttg tgcttgcaga gcgccacggg ttcttgatat tcgaggacag 960

gaagtttgct gatatcggaa acaccgtgat gttgcagtac acctcggggg tataccggat 1020

cgcggcgtgg agtgacatca cgaacgcgca cggagtgact gggaagggcg tcgttgaagg 1080

gttgaaacgc ggtgcggagg gggtagaaaa ggaaaggggc gtgttgatgt tggcggagtt 1140

gtcgagtaaa ggctcgttgg cgcatggtga atatacccgt gagacgatcg agattgcgaa 1200

gagtgatcgg gagttcgtga ttgggttcat cgcgcagcgg gacatggggg gtagagaaga 1260

agggtttgat tggatcatca tgacgcctgg tgtggggttg gatgataaag gcgatgcgtt 1320

gggccagcag tataggactg ttgatgaggt ggttctgact ggtaccgatg tgattattgt 1380

cgggagaggg ttgtttggaa aaggaagaga ccctgaggtg gagggaaaga gatacaggga 1440

tgctggatgg aaggcatact tgaagagaac tggtcagtta gaataaatat tgtaataaat 1500

aggtctatat acatacacta agcttctagg acgtcattgt agtcttcgaa gttgtctgct 1560

agtttagttc tcatgatttc gaaaaccaat aacgcaatgg atgtagcagg gatggtggtt 1620

agtgcgttcc tgacaaaccc agagtacgcc gcctcaaacc acgtcacatt cgccctttgc 1680

ttcatccgca tcacttgctt gaaggtatcc acgtacgagt tgtaatacac a 1731

<210> 59
<211> 1884
<212> DNA
<213> Candida viswanathii

<400> 59

```
atgtttaact ttaagttgtc gcaacaagta ttaaagaatt ccaccaaatc cattatgcca       60
attttgaaaa aaccattctc caccagccac gcaaagggtg acttgttcaa ataccagtca      120
caattaccca agttgcctgt tcctactttg gaagaaaccg catccaagta cctcaagacc      180
gttgagccat tcttgaacca agagcaattg gaatccacca aggccaaagt cgctgagttt      240
gttagaccag gtggtgccgg tgaagccttg caagccagat tgaacaactt tgccgccgac      300
aaggacaact ggttggctga attttgggac gactatgcat acatgtctta tagagatcct      360
gttgttccat atgtttctta ctttttcagt cacaaggatg tcaagaacat cattggccaa      420
gaccaattgt tgaaggccac tttgattgct tactacacta ttgagttcca agaaaaggtt      480
ttggacgaaa gtttggaccc agaagtcatc aagggtaacc cattctgtat gaacgccttc      540
aagtacatgt tcaacaactc gagagttcca gctgaaggct ccgacatcac ccaacactac      600
aacggtgaag aaaaccaatt tttcgttgtc atctacaaga acaacttcta caaggttcca      660
acccacaaga acggccaaag attgaccaag ggtgaaatct acagctactt gcaagaaatc      720
aagaacgatg ccactccaaa gggtctcggt ttgggtgctt tgacctcatt gaacagagac      780
gaatggttga gtgcctacaa caacttgttg aagtccccaa tcaacgaagc ttccttggga      840
tccatctttg cttccagctt tgtcattgcc ttggactcca acaacccagt caccattgaa      900
gaaaaatcca agaactgctg gcacggggac ggtcaaaaca gattctttga caagcctttg      960
gaattcttcg tcagtgctaa cggtaactct ggtttccttg gtgaacactc cagaatggac     1020
gctaccccaa ccgtgcaatt gaacaacacc atctacaagc aaatcttgga aaccaatcca     1080
aacgacttga ttgttgaaat tggttcttct gctccaagat tcggcaatgc tgaaatcttg     1140
cctttcgaca tcaacccaac caccagagcc aacatcaaag acgctattgc caagtttgac     1200
gccaccattg ctgcccacga cgaagaaatc ttccaacact acggttacgg taagggattg     1260
atcaagaagt tcaaggtctc cccagatgcc tacgtgcaat tgttgatgca attggcatac     1320
ttcaagtaca ccggcaagat cagaccaact tatgaatccg ccgccaccag aaagttcttg     1380
aagggtagaa ccgaaaccgg tagaactgtt tccaacgaat ccaagaagtt tgttgagacc     1440
tggtccgatc caaacgctag cagcgccgac aaggttgcca ctttccaagc tgccgctaag     1500
caacacgttg cttacttgtc tgctgccgcc gatggtaagg gtgtcgaccg tcacttgttt     1560
ggtttgaagc agatgattca accaggcgaa ccaatccctg aaatcttcac tgacccaatc     1620
tttagctatt ctcaaacctg gtacatttct tcttcccaag tcccatctga attcttccaa     1680
tcttggggtt ggtcgcaagt cattgacgac ggtttcggtt tggcttactt gatcaacaac     1740
gactggatcc acgttcacat ttcttgtaag agaggtaacg gcttgcaatc cgaccacttg     1800
aaatggtact tggttgatag tgctaacgaa atgaaggatg tcttgactaa gggattattg     1860
actgatgcta agcctaagtt gtaa                                           1884
```

<210> 60
<211> 1830
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 60
atgccaattt tgaaaaaacc attctccacc agccacgcaa agggtgactt gttcaaatac 60

cagtcacaat tacccaagtt gcctgttcct actttggaag aaaccgcatc caagtacctc 120

aagaccgttg agccattctt gaaccaagag caattggaat ccaccaaggc caaagtcgct 180

gagtttgtta gaccaggtgg tgccggtgaa gccttgcaag ccagattgaa caactttgcc 240

gccgacaagg acaactggtt ggctgaattt tgggacgact atgcatacat gtcttataga 300

gatcctgttg ttccatatgt ttcttacttt ttcagtcaca aggatgtcaa gaacatcatt 360

ggccaagacc aattgttgaa ggccactttg attgcttact acactattga gttccaagaa 420

aaggttttgg acgaaagttt ggacccagaa gtcatcaagg gtaacccatt ctgtatgaac 480

gccttcaagt acatgttcaa caactcgaga gttccagctg aaggctccga catcacccaa 540

cactacaacg gtgaagaaaa ccaatttttc gttgtcatct acaagaacaa cttctacaag 600

gttccaaccc acaagaacgg ccaaagattg accaagggtg aaatctacag ctacttgcaa 660

gaaatcaaga acgatgccac tccaaagggt ctcggtttgg gtgctttgac ctcattgaac 720

agagacgaat ggttgagtgc ctacaacaac ttgttgaagt ccccaatcaa cgaagcttcc 780

ttgggatcca tctttgcttc cagctttgtc attgccttgg actccaacaa cccagtcacc 840

attgaagaaa aatccaagaa ctgctggcac ggggacggtc aaaacagatt ctttgacaag 900

cctttggaat tcttcgtcag tgctaacggt aactctggtt tccttggtga acactccaga 960

atggacgcta ccccaaccgt gcaattgaac aacaccatct acaagcaaat cttggaaacc 1020

aatccaaacg acttgattgt tgaaattggt tcttctgctc caagattcgg caatgctgaa 1080

atcttgcctt tcgacatcaa cccaaccacc agagccaaca tcaaagacgc tattgccaag 1140

tttgacgcca ccattgctgc ccacgacgaa gaaatcttcc aacactacgg ttacggtaag 1200

ggattgatca agaagttcaa ggtctcccca gatgcctacg tgcaattgtt gatgcaattg 1260

gcatacttca agtacaccgg caagatcaga ccaacttatg aatccgccgc caccagaaag 1320

ttcttgaagg gtagaaccga aaccggtaga actgtttcca acgaatccaa gaagtttgtt 1380

gagacctggt ccgatccaaa cgctagcagc gccgacaagg ttgccacttt ccaagctgcc 1440

gctaagcaac acgttgctta cttgtctgct gccgccgatg gtaagggtgt cgaccgtcac 1500

ttgtttggtt tgaagcagat gattcaacca ggcgaaccaa tccctgaaat cttcactgac 1560

ccaatcttta gctattctca aacctggtac atttcttctt cccaagtccc atctgaattc 1620

ttccaatctt ggggttggtc gcaagtcatt gacgacggtt tcggtttggc ttacttgatc 1680

aacaacgact ggatccacgt tcacatttct tgtaagagag gtaacggctt gcaatccgac 1740

cacttgaaat ggtacttggt tgatagtgct aacgaaatga aggatgtctt gactaaggga 1800

ttattgactg atgctaagcc taagttgtaa 1830

<210> 61
<211> 1821
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 61
atgccaattt tgaaaaaacc attctccacc agccacgcaa agggtgactt gttcaaatac 60

cagtcacaat tacccaagtt gcctgttcct actttggaag aaaccgcatc caagtacctc 120

aagaccgttg agccattctt gaaccaagag caattggaat ccaccaaggc caaagtcgct 180

gagtttgtta gaccaggtgg tgccggtgaa gccttgcaag ccagattgaa caactttgcc 240

gccgacaagg acaactggtt ggctgaattt tgggacgact atgcatacat gtcttataga 300

gatcctgttg ttccatatgt ttcttacttt ttcagtcaca aggatgtcaa gaacatcatt 360

ggccaagacc aattgttgaa ggccactttg attgcttact acactattga gttccaagaa 420

aaggttttgg acgaaagttt ggacccagaa gtcatcaagg gtaacccatt ctgtatgaac 480

gccttcaagt acatgttcaa caactcgaga gttccagctg aaggctccga catcacccaa 540

cactacaacg gtgaagaaaa ccaatttttc gttgtcatct acaagaacaa cttctacaag 600

gttccaaccc acaagaacgg ccaaagattg accaagggtg aaatctacag ctacttgcaa 660

gaaatcaaga acgatgccac tccaaagggt ctcggtttgg gtgctttgac ctcattgaac 720

agagacgaat ggttgagtgc ctacaacaac ttgttgaagt ccccaatcaa cgaagcttcc 780

ttgggatcca tctttgcttc agctttgtc attgccttgg actccaacaa cccagtcacc 840

attgaagaaa aatccaagaa ctgctggcac ggggacggtc aaaacagatt ctttgacaag 900

cctttggaat tcttcgtcag tgctaacggt aactctggtt tccttggtga acactccaga 960

atggacgcta ccccaaccgt gcaattgaac aacaccatct acaagcaaat cttggaaacc 1020

aatccaaacg acttgattgt tgaaattggt tcttctgctc caagattcgg caatgctgaa 1080

atcttgcctt tcgacatcaa cccaaccacc agagccaaca tcaaagacgc tattgccaag 1140

tttgacgcca ccattgctgc ccacgacgaa gaaatcttcc aacactacgg ttacggtaag 1200

ggattgatca agaagttcaa ggtctcccca gatgcctacg tgcaattgtt gatgcaattg 1260

gcatacttca agtacaccgg caagatcaga ccaacttatg aatccgccgc caccagaaag 1320

ttcttgaagg gtagaaccga aaccggtaga actgtttcca acgaatccaa gaagtttgtt 1380

gagacctggt ccgatccaaa cgctagcagc gccgacaagg ttgccacttt ccaagctgcc 1440

gctaagcaac acgttgctta cttgtctgct gccgccgatg gtaagggtgt cgaccgtcac 1500

ttgtttggtt tgaagcagat gattcaacca ggcgaaccaa tccctgaaat cttcactgac 1560

ccaatcttta gctattctca aacctggtac atttcttctt cccaagtccc atctgaattc 1620

ttccaatctt ggggttggtc gcaagtcatt gacgacggtt tcggtttggc ttacttgatc 1680

aacaacgact ggatccacgt tcacatttct tgtaagagag gtaacggctt gcaatccgac 1740

cacttgaaat ggtacttggt tgatagtgct aacgaaatga aggatgtctt gactaaggga 1800

ttattgactg atgctaagta a 1821

<210> 62
<211> 1875
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 62
atgtttaact ttaagttgtc gcaacaagta ttaaagaatt ccaccaaatc cattatgcca 60

attttgaaaa aaccattctc caccagccac gcaaagggtg acttgttcaa ataccagtca 120

caattaccca agttgcctgt tcctactttg gaagaaaccg catccaagta cctcaagacc 180

gttgagccat tcttgaacca agagcaattg gaatccacca aggccaaagt cgctgagttt 240

gttagaccag gtggtgccgg tgaagccttg caagccagat tgaacaactt tgccgccgac 300

aaggacaact ggttggctga attttgggac gactatgcat acatgtctta tagagatcct 360

gttgttccat atgtttctta ctttttcagt cacaaggatg tcaagaacat cattggccaa 420

gaccaattgt tgaaggccac tttgattgct tactacacta ttgagttcca agaaaaggtt 480

ttggacgaaa gtttggaccc agaagtcatc aagggtaacc cattctgtat gaacgccttc 540

aagtacatgt tcaacaactc gagagttcca gctgaaggct ccgacatcac ccaacactac 600

aacggtgaag aaaaccaatt tttcgttgtc atctacaaga acaacttcta caaggttcca 660

acccacaaga acggccaaag attgaccaag ggtgaaatct acagctactt gcaagaaatc 720

aagaacgatg ccactccaaa gggtctcggt ttgggtgctt tgacctcatt gaacagagac 780

gaatggttga gtgcctacaa caacttgttg aagtccccaa tcaacgaagc ttccttggga 840

tccatctttg cttccagctt tgtcattgcc ttggactcca acaacccagt caccattgaa 900

gaaaaatcca agaactgctg gcacggggac ggtcaaaaca gattctttga caagcctttg 960

gaattcttcg tcagtgctaa cggtaactct ggtttccttg gtgaacactc cagaatggac 1020

gctaccccaa ccgtgcaatt gaacaacacc atctacaagc aaatcttgga aaccaatcca 1080

aacgacttga ttgttgaaat tggttcttct gctccaagat tcggcaatgc tgaaatcttg 1140

cctttcgaca tcaacccaac caccagagcc aacatcaaag acgctattgc caagtttgac 1200

gccaccattg ctgcccacga cgaagaaatc ttccaacact acggttacgg taagggattg 1260

atcaagaagt tcaaggtctc cccagatgcc tacgtgcaat tgttgatgca attggcatac 1320

ttcaagtaca ccggcaagat cagaccaact tatgaatccg ccgccaccag aaagttcttg 1380

aagggtagaa ccgaaaccgg tagaactgtt tccaacgaat ccaagaagtt tgttgagacc 1440

tggtccgatc caaacgctag cagcgccgac aaggttgcca ctttccaagc tgccgctaag 1500

caacacgttg cttacttgtc tgctgccgcc gatggtaagg gtgtcgaccg tcacttgttt 1560

ggtttgaagc agatgattca accaggcgaa ccaatccctg aaatcttcac tgacccaatc 1620

tttagctatt ctcaaacctg gtacatttct tcttcccaag tcccatctga attcttccaa 1680

tcttggggtt ggtcgcaagt cattgacgac ggtttcggtt tggcttactt gatcaacaac 1740

gactggatcc acgttcacat ttcttgtaag agaggtaacg gcttgcaatc cgaccacttg 1800

aaatggtact tggttgatag tgctaacgaa atgaaggatg tcttgactaa gggattattg 1860

actgatgcta agtaa 1875

<210> 63
<211> 2301
<212> DNA
<213> Candida viswanathii

<400> 63
atgtcaactt accagttcca agaaacttta gaaaagcttc caatcccaga cttaaaccaa 60

acatgcgcta actacctcaa cgtcttgcgt ccgctccaaa cagagcaaga acacatcaag 120

accaagactg ccgttgagaa tttcctcaag aatggcaccg gccagtacct cgacgctgcc 180

ttgcgcgaat acgcccagac ccgccccagc tacattgaac agttctggta cgacgcatac 240

ttaaactacg actcccccgt tgttttaaac ttgaatccgt ttttcttact cgaagacgac 300

ccattcacca acgagtccag ctccgtcaat ccccaggtta agcgtgccac cagcttggtg 360

atgtcctcgt tgaagttcat ccaggcgttg aagaacgaga cgttgagcgt cgacaccttg 420

aaaggaggca agccgttgtg catgtaccag tacaccaaat tgtttggggc gctgcgtatc 480

ccctccgagg acgggtgtgt gatgcagagc gacccggcaa gtaaccacat tgtcgtgatg 540

tctaagtccc agttgtattg gtttgacgtg ttggatagca agaacaactt gattttgagc 600

gaagcggagt tgaatgtgaa cttccagtcg atcattcatg actctttgag aactccctcg 660

gatgagattg caaagagctc gtttggggtg ttgacgacgg agaatagaag aatttgggct 720

aatgttagac ataatttgat gtcgactacc aataataagg ttaaccatga ggtgttgtcg 780

attattgact cagcgttgtt tgtgttatgc tttgatgata tcgtgattaa tgacttgtcg 840

gagttgtcca agaatatgct ttgtgggttg tccattttgg ataatggtat ccaagtgggt 900

acttgtacaa acagatggta tgacaagttg cagattatcg ttaccaagaa tgccaaagcc 960

ggtattaatt tcgaacatac tggtgttgac ggtcatacgg tgttgagatt tgtcagtgat 1020

atctacaccg actccatctt gtcctttgcc aactccatca atagcaacgc tccatcttta 1080

tggaacacca ctacaaacta caaagaactt gatggggaag atttaatcac cgtgccacgt 1140

aagttggaat gggagttgac ccctgactta tccttggctt tgagatttgg agaaaccaga 1200

ttgtcggact tgatcaacca gaacgagttc cgccacttgg agttcaagaa ctatgggtct 1260

acacaaatca agaagatgaa gttctcacca gatgcgtttg ttcaaatggc gttccaagca 1320

acctactatg cgctttacgg taaggttgaa tgtacttatg agccagccat gacgaagcag 1380

ttctaccacg gaagaacaga agctataaga accgtctcgc aagaatcgaa cttgtttgtg 1440

cgtaagttct ttgactccac tgtgtccatc cagaagaagt tagagtactt gacccaagct 1500

tgcaccaagc actcccagca aaccagaatg agttctgccg gtcaaggtgt cgatcgtcac 1560

ttgtatgctt tattctgtat ctggaagaga tacttacacg atgccgaaga cgacgatgac 1620

gatgcctcca aagaccaacc atcatcgacg gaattatccc gtgatgacac catcattgct 1680

caaaccaacg gtgacagcac cgatgatgac gactccacca ccgttgtcgg acacggcgcc 1740

aaccaccatc acaacttgaa aagcgccgat ttgttgaaaa ccatccctga gatattcgca 1800

gacaatggct gggacaagtt gaacaacacg atcatatcca cctccaactg tggtaatcct 1860

tctttgagat tgtttggatt tggtcctgtt tcagctaacg ggtttggtat cgggtatatc 1920

ttgaaggacg actcgatctc gatctgtgct tcgtcgaagc atcgtcagac gcagagattc 1980

ttggtcactt tgaactcgta cttgttggag atccagaaca tctggaaaca agcgcaaaag 2040

atggagaagt tgaaggctga gttggctaaa gccgtcgagg atgcaaaacc taagggacag 2100

aaatcaggtg aagcagaggg gatcatcgag cctaagccta ataacttgag cactcttttg 2160

ggaggttatg gttatttcga tatgggcgat gaggacatca agagcagagg acaatctcca 2220

gaacctccgt tcttgcatcg tgctggatcc ggtttcagca ttcgtgaaat aggcaagaag 2280

ttgagattaa gcgagtatta g 2301

<210> 64
<211> 90
<212> DNA
<213> Candida viswanathii

<400> 64
atgctttccc gtaccacttt aagagttgcc agacaacaaa ccagattatt gtctacttcc 60

agaatcttgt tcaacagcaa gaccgaccaa 90

<210> 65
<211> 75
<212> DNA
<213> Candida viswanathii

<400> 65
atgtctgcat taagatcatt ccaacgctct tctaacgttg ctaagtccac cttaaagaac 60

tcagtcagaa cctat 75

<210> 66
<211> 99
<212> DNA
<213> Candida viswanathii

<400> 66
atgtttaact ttaagttgtc gcaacaagta ttaaagaatt ccaccaaatc cattatgcca 60

attttgaaaa aaccattctc caccagccac gcaaagggt 99

<210> 67
<211> 2388
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 67
atgctttccc gtaccacttt aagagttgcc agacaacaaa ccagattatt gtctacttcc 60

agaatcttgt tcaacagcaa gaccgaccaa tcaacttacc agttccaaga aactttagaa 120

aagcttccaa tcccagactt aaaccaaaca tgcgctaact acctcaacgt cttgcgtccg 180

ctccaaacag agcaagaaca catcaagacc aagactgccg ttgagaattt cctcaagaat 240

ggcaccggcc agtacctcga cgctgccttg cgcgaatacg cccagacccg ccccagctac 300

attgaacagt tctggtacga cgcatactta aactacgact cccccgttgt tttaaacttg 360

aatccgtttt tcttactcga agacgaccca ttcaccaacg agtccagctc cgtcaatccc 420

caggttaagc gtgccaccag cttggtgatg tcctcgttga agttcatcca ggcgttgaag 480

aacgagacgt tgagcgtcga caccttgaaa ggaggcaagc cgttgtgcat gtaccagtac 540

accaaattgt ttggggcgct gcgtatcccc tccgaggacg ggtgtgtgat gcagagcgac 600

ccggcaagta accacattgt cgtgatgtct aagtcccagt tgtattggtt tgacgtgttg 660

gatagcaaga acaacttgat tttgagcgaa gcggagttga atgtgaactt ccagtcgatc 720

attcatgact ctttgagaac tccctcggat gagattgcaa agagctcgtt tggggtgttg 780

acgacggaga atagaagaat ttgggctaat gttagacata atttgatgtc gactaccaat 840

aataaggtta accatgaggt gttgtcgatt attgactcag cgttgtttgt gttatgcttt 900

gatgatatcg tgattaatga cttgtcggag ttgtccaaga atatgctttg tgggttgtcc 960

attttggata atggtatcca agtgggtact tgtacaaaca gatggtatga caagttgcag 1020

attatcgtta ccaagaatgc caaagccggt attaatttcg aacatactgg tgttgacggt 1080

catacggtgt tgagatttgt cagtgatatc tacaccgact ccatcttgtc ctttgccaac 1140

tccatcaata gcaacgctcc atctttatgg aacaccacta caaactacaa agaacttgat 1200

ggggaagatt taatcaccgt gccacgtaag ttggaatggg agttgacccc tgacttatcc 1260

ttggctttga gatttggaga aaccagattg tcggacttga tcaaccagaa cgagttccgc 1320

cacttggagt tcaagaacta tgggtctaca caaatcaaga agatgaagtt ctcaccagat 1380

gcgtttgttc aaatggcgtt ccaagcaacc tactatgcgc tttacggtaa ggttgaatgt 1440

acttatgagc cagccatgac gaagcagttc taccacggaa gaacagaagc tataagaacc 1500

gtctcgcaag aatcgaactt gtttgtgcgt aagttctttg actccactgt gtccatccag 1560

aagaagttag agtacttgac ccaagcttgc accaagcact cccagcaaac cagaatgagt 1620

tctgccggtc aaggtgtcga tcgtcacttg tatgctttat tctgtatctg gaagagatac 1680

ttacacgatg ccgaagacga cgatgacgat gcctccaaag accaaccatc atcgacggaa 1740

ttatcccgtg atgacaccat cattgctcaa accaacggtg acagcaccga tgatgacgac 1800

tccaccaccg ttgtcggaca cggcgccaac caccatcaca acttgaaaag cgccgatttg 1860

ttgaaaacca tccctgagat attcgcagac aatggctggg acaagttgaa caacacgatc 1920

atatccacct ccaactgtgg taatccttct ttgagattgt ttggatttgg tcctgtttca 1980

gctaacgggt ttggtatcgg gtatatcttg aaggacgact cgatctcgat ctgtgcttcg 2040

tcgaagcatc gtcagacgca gagattcttg gtcactttga actcgtactt gttggagatc 2100

cagaacatct ggaaacaagc gcaaaagatg gagaagttga aggctgagtt ggctaaagcc 2160

gtcgaggatg caaaacctaa gggacagaaa tcaggtgaag cagaggggat catcgagcct 2220

aagcctaata acttgagcac tcttttggga ggttatggtt atttcgatat gggcgatgag 2280

gacatcaaga gcagaggaca atctccagaa cctccgttct tgcatcgtgc tggatccggt 2340

ttcagcattc gtgaaatagg caagaagttg agattaagcg agtattag 2388

<210> 68
<211> 2373
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 68

atgtctgcat taagatcatt ccaacgctct tctaacgttg ctaagtccac cttaaagaac 60

tcagtcagaa cctattcaac ttaccagttc caagaaactt tagaaaagct tccaatccca 120

gacttaaacc aaacatgcgc taactacctc aacgtcttgc gtccgctcca aacagagcaa 180

gaacacatca agaccaagac tgccgttgag aatttcctca agaatggcac cggccagtac 240

ctcgacgctg ccttgcgcga atacgcccag acccgcccca gctacattga acagttctgg 300

tacgacgcat acttaaacta cgactccccc gttgttttaa acttgaatcc gtttttctta 360

ctcgaagacg acccattcac caacgagtcc agctccgtca atccccaggt taagcgtgcc 420

accagcttgg tgatgtcctc gttgaagttc atccaggcgt tgaagaacga gacgttgagc 480

gtcgacacct tgaaaggagg caagccgttg tgcatgtacc agtacaccaa attgtttggg 540

gcgctgcgta tcccctccga ggacgggtgt gtgatgcaga gcgacccggc aagtaaccac 600

attgtcgtga tgtctaagtc ccagttgtat tggtttgacg tgttggatag caagaacaac 660

ttgattttga gcgaagcgga gttgaatgtg aacttccagt cgatcattca tgactctttg 720

agaactccct cggatgagat tgcaaagagc tcgtttgggg tgttgacgac ggagaataga 780

agaatttggg ctaatgttag acataatttg atgtcgacta ccaataataa ggttaaccat 840

gaggtgttgt cgattattga ctcagcgttg tttgtgttat gctttgatga tatcgtgatt 900

aatgacttgt cggagttgtc caagaatatg ctttgtgggt tgtccatttt ggataatggt 960

atccaagtgg gtacttgtac aaacagatgg tatgacaagt tgcagattat cgttaccaag 1020

aatgccaaag ccggtattaa tttcgaacat actggtgttg acggtcatac ggtgttgaga 1080

tttgtcagtg atatctacac cgactccatc ttgtcctttg ccaactccat caatagcaac 1140

gctccatctt tatggaacac cactacaaac tacaaagaac ttgatgggga agatttaatc 1200

accgtgccac gtaagttgga atgggagttg acccctgact tatccttggc tttgagattt 1260

ggagaaacca gattgtcgga cttgatcaac cagaacgagt tccgccactt ggagttcaag 1320

aactatgggt ctacacaaat caagaagatg aagttctcac cagatgcgtt tgttcaaatg 1380

gcgttccaag caacctacta tgcgctttac ggtaaggttg aatgtactta tgagccagcc 1440

atgacgaagc agttctacca cggaagaaca gaagctataa gaaccgtctc gcaagaatcg 1500

aacttgtttg tgcgtaagtt ctttgactcc actgtgtcca tccagaagaa gttagagtac 1560

ttgacccaag cttgcaccaa gcactcccag caaaccagaa tgagttctgc cggtcaaggt 1620

gtcgatcgtc acttgtatgc tttattctgt atctggaaga gatacttaca cgatgccgaa 1680

gacgacgatg acgatgcctc caaagaccaa ccatcatcga cggaattatc ccgtgatgac 1740

accatcattg ctcaaaccaa cggtgacagc accgatgatg acgactccac caccgttgtc 1800

ggacacggcg ccaaccacca tcacaacttg aaaagcgccg atttgttgaa aaccatccct 1860

gagatattcg cagacaatgg ctgggacaag ttgaacaaca cgatcatatc cacctccaac 1920

tgtggtaatc cttctttgag attgtttgga tttggtcctg tttcagctaa cgggtttggt 1980

atcgggtata tcttgaagga cgactcgatc tcgatctgtg cttcgtcgaa gcatcgtcag 2040

acgcagagat tcttggtcac tttgaactcg tacttgttgg agatccagaa catctggaaa 2100

caagcgcaaa agatggagaa gttgaaggct gagttggcta aagccgtcga ggatgcaaaa 2160

cctaagggac agaaatcagg tgaagcagag gggatcatcg agcctaagcc taataacttg 2220

agcactcttt tgggaggtta tggttatttc gatatgggcg atgaggacat caagagcaga 2280

ggacaatctc cagaacctcc gttcttgcat cgtgctggat ccggtttcag cattcgtgaa 2340

ataggcaaga agttgagatt aagcgagtat tag 2373

<210> 69
<211> 2397
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 69
atgtttaact ttaagttgtc gcaacaagta ttaaagaatt ccaccaaatc cattatgcca 60

attttgaaaa aaccattctc caccagccac gcaaagggtt caacttacca gttccaagaa 120

actttagaaa agcttccaat cccagactta aaccaaacat gcgctaacta cctcaacgtc 180

ttgcgtccgc tccaaacaga gcaagaacac atcaagacca agactgccgt tgagaatttc 240

ctcaagaatg gcaccggcca gtacctcgac gctgccttgc gcgaatacgc ccagacccgc 300

cccagctaca ttgaacagtt ctggtacgac gcatacttaa actacgactc ccccgttgtt 360

ttaaacttga atccgttttt cttactcgaa gacgacccat tcaccaacga gtccagctcc 420

gtcaatcccc aggttaagcg tgccaccagc ttggtgatgt cctcgttgaa gttcatccag 480

gcgttgaaga acgagacgtt gagcgtcgac accttgaaag gaggcaagcc gttgtgcatg 540

taccagtaca ccaaattgtt tggggcgctg cgtatcccct ccgaggacgg gtgtgtgatg 600

cagagcgacc cggcaagtaa ccacattgtc gtgatgtcta agtcccagtt gtattggttt 660

gacgtgttgg atagcaagaa caacttgatt ttgagcgaag cggagttgaa tgtgaacttc 720

cagtcgatca ttcatgactc tttgagaact ccctcggatg agattgcaaa gagctcgttt 780

ggggtgttga cgacggagaa tagaagaatt tgggctaatg ttagacataa tttgatgtcg 840

actaccaata ataaggttaa ccatgaggtg ttgtcgatta ttgactcagc gttgtttgtg 900

ttatgctttg atgatatcgt gattaatgac ttgtcggagt tgtccaagaa tatgctttgt 960

gggttgtcca ttttggataa tggtatccaa gtgggtactt gtacaaacag atggtatgac 1020

aagttgcaga ttatcgttac caagaatgcc aaagccggta ttaatttcga acatactggt 1080

gttgacggtc atacggtgtt gagatttgtc agtgatatct acaccgactc catcttgtcc 1140

tttgccaact ccatcaatag caacgctcca tctttatgga acaccactac aaactacaaa 1200

gaacttgatg gggaagattt aatcaccgtg ccacgtaagt tggaatggga gttgacccct 1260

gacttatcct tggctttgag atttggagaa accagattgt cggacttgat caaccagaac 1320

gagttccgcc acttggagtt caagaactat gggtctacac aaatcaagaa gatgaagttc 1380

tcaccagatg cgtttgttca aatggcgttc caagcaacct actatgcgct ttacggtaag 1440

gttgaatgta cttatgagcc agccatgacg aagcagttct accacggaag aacagaagct 1500

ataagaaccg tctcgcaaga atcgaacttg tttgtgcgta agttctttga ctccactgtg 1560

tccatccaga agaagttaga gtacttgacc caagcttgca ccaagcactc ccagcaaacc 1620

agaatgagtt ctgccggtca aggtgtcgat cgtcacttgt atgctttatt ctgtatctgg 1680

aagagatact tacacgatgc cgaagacgac gatgacgatg cctccaaaga ccaaccatca 1740

tcgacggaat tatcccgtga tgacaccatc attgctcaaa ccaacggtga cagcaccgat 1800

gatgacgact ccaccaccgt tgtcggacac ggcgccaacc accatcacaa cttgaaaagc 1860

gccgatttgt tgaaaaccat ccctgagata ttcgcagaca atggctggga caagttgaac 1920

aacacgatca tatccacctc caactgtggt aatccttctt tgagattgtt tggatttggt 1980

cctgtttcag ctaacgggtt tggtatcggg tatatcttga aggacgactc gatctcgatc 2040

tgtgcttcgt cgaagcatcg tcagacgcag agattcttgg tcactttgaa ctcgtacttg 2100

ttggagatcc agaacatctg gaaacaagcg caaaagatgg agaagttgaa ggctgagttg 2160

gctaaagccg tcgaggatgc aaaacctaag ggacagaaat caggtgaagc agaggggatc 2220

atcgagccta agcctaataa cttgagcact cttttgggag gttatggtta tttcgatatg 2280

ggcgatgagg acatcaagag cagaggacaa tctccagaac ctccgttctt gcatcgtgct 2340

ggatccggtt tcagcattcg tgaaataggc aagaagttga gattaagcga gtattag 2397

<210> 70
<211> 2310
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 70
atgtcaactt accagttcca agaaacttta gaaaagcttc caatcccaga cttaaaccaa 60

acatgcgcta actacctcaa cgtcttgcgt ccgctccaaa cagagcaaga acacatcaag 120

accaagactg ccgttgagaa tttcctcaag aatggcaccg gccagtacct cgacgctgcc 180

```
ttgcgcgaat acgcccagac ccgccccagc tacattgaac agttctggta cgacgcatac      240
ttaaactacg actccccccgt tgttttaaac ttgaatccgt ttttcttact cgaagacgac      300
ccattcacca acgagtccag ctccgtcaat ccccaggtta agcgtgccac cagcttggtg      360
atgtcctcgt tgaagttcat ccaggcgttg aagaacgaga cgttgagcgt cgacaccttg      420
aaaggaggca agccgttgtg catgtaccag tacaccaaat tgtttggggc gctgcgtatc      480
ccctccgagg acgggtgtgt gatgcagagc gacccggcaa gtaaccacat tgtcgtgatg      540
tctaagtccc agttgtattg gtttgacgtg ttggatagca agaacaactt gattttgagc      600
gaagcggagt tgaatgtgaa cttccagtcg atcattcatg actctttgag aactccctcg      660
gatgagattg caaagagctc gtttggggtg ttgacgacgg agaatagaag aatttgggct      720
aatgttagac ataatttgat gtcgactacc aataataagg ttaaccatga ggtgttgtcg      780
attattgact cagcgttgtt tgtgttatgc tttgatgata tcgtgattaa tgacttgtcg      840
gagttgtcca agaatatgct ttgtgggttg tccattttgg ataatggtat ccaagtgggt      900
acttgtacaa acagatggta tgacaagttg cagattatcg ttaccaagaa tgccaaagcc      960
ggtattaatt tcgaacatac tggtgttgac ggtcatacgg tgttgagatt tgtcagtgat     1020
atctacaccg actccatctt gtcctttgcc aactccatca atagcaacgc tccatcttta     1080
tggaacacca ctacaaacta caaagaactt gatggggaag atttaatcac cgtgccacgt     1140
aagttggaat gggagttgac ccctgactta tccttggctt tgagatttgg agaaaccaga     1200
ttgtcggact tgatcaacca gaacgagttc cgccacttgg agttcaagaa ctatgggtct     1260
acacaaatca agaagatgaa gttctcacca gatgcgtttg ttcaaatggc gttccaagca     1320
acctactatg cgctttacgg taaggttgaa tgtacttatg agccagccat gacgaagcag     1380
ttctaccacg gaagaacaga agctataaga accgtctcgc aagaatcgaa cttgtttgtg     1440
cgtaagttct ttgactccac tgtgtccatc cagaagaagt tagagtactt gacccaagct     1500
tgcaccaagc actcccagca aaccagaatg agttctgccg gtcaaggtgt cgatcgtcac     1560
ttgtatgctt tattctgtat ctggaagaga tacttacacg atgccgaaga cgacgatgac     1620
gatgcctcca aagaccaacc atcatcgacg gaattatccc gtgatgacac catcattgct     1680
caaaccaacg gtgacagcac cgatgatgac gactccacca ccgttgtcgg acacggcgcc     1740
aaccaccatc acaacttgaa aagcgccgat ttgttgaaaa ccatccctga gatattcgca     1800
gacaatggct gggacaagtt gaacaacacg atcatatcca cctccaactg tggtaatcct     1860
tctttgagat tgtttggatt tggtcctgtt tcagctaacg ggtttggtat cgggtatatc     1920
ttgaaggacg actcgatctc gatctgtgct tcgtcgaagc atcgtcagac gcagagattc     1980
ttggtcactt tgaactcgta cttgttggag atccagaaca tctggaaaca agcgcaaaag     2040
atggagaagt tgaaggctga gttggctaaa gccgtcgagg atgcaaaacc taagggacag     2100
```

aaatcaggtg aagcagaggg gatcatcgag cctaagccta ataacttgag cactcttttg 2160

ggaggttatg gttatttcga tatgggcgat gaggacatca agagcagagg acaatctcca 2220

gaacctccgt tcttgcatcg tgctggatcc ggtttcagca ttcgtgaaat aggcaagaag 2280

ttgagattaa gcgagtatcc taagttgtag 2310

<210> 71
<211> 855
<212> DNA
<213> Candida viswanathii

<400> 71
atggacgacg ttgattctgc tttagccgat aatgttaaat ccttcgctgc cggtggtttt 60

ggtggtattt gtgccgtgtt gaccggtcat ccattcgact tggtcaaggt cagattgcaa 120

actgggttgt acaaatcgtc agtgcaatgt gtcaaagaaa caatagccaa agacgggttg 180

tttgggttat acagaggtgt cttgcctcca ttgttgggtg tcaccccaat gtttgctgtt 240

tccttctggg gttacgacgt tggtaagaaa ttggtttctt ccttcactgg caaatcagtc 300

gacaagtttg aaatcaaaga catttcgact gccggtttca tcagtgccat cccaaccacc 360

ttggttgccg ctccattcga aagagtcaag gtcatgatgc agatccaaga aggtgccaag 420

tccaagtcca tgggtgccgt tgttgctgaa atgtacagaa ccggtggtat cagatccatc 480

ttcaagggta ctgtggctac tttggccaga gatggtccag gttctgcctt gtatttcgct 540

acttatgaat gggtcaagaa ggaattgact gcccctggtg aagacttgtc cttgtttgct 600

attaccactg ctggtggttt cgctggtatt gccatgtggt tgggtgtttt cccaattgac 660

accatcaagt ctacccagca gtcttccaac gtcaaggtct ccattgtgca ggccactaag 720

aacatctacg ctaaaggtgg tatcaaggct ttcttccctg gtgttggtcc tgccttggcc 780

agagctttcc cagccaacgc tgctaccttc ctaggggttg aattggccag aaagttcttg 840

gacaaagtta tataa 855

<210> 72
<211> 1560
<212> DNA
<213> Candida viswanathii

<400> 72
atgtctgcca ttttgaaaca aagagttaga tatgctccat acctcaaaaa attgagaaca 60

gcagaacaat gtgtcgactt gttcaagaac ggccaatact tgggttggtc gggtttcacc 120

ggtgtcggtg ccccaaaggc tgtcccagcc gctttggttg accacgttga aaagaacaaa 180

ttgcagggcc aaatgggctt ccacttgttt gtcggtgctt ctgccggtcc agaagaaagc 240

agatgggctg aaaacaacat gcttttatct agatctccac atcaagtcgg taagccaatt 300

gctgctgcca tcaatgacgg cagaactcag ttcttcgaca agcacttgtc catgttccca 360

caagatttga cctacggctt ctacaccaag aacaagccaa acggctccaa cttggattac 420

accatcattg aagctaccgc cataacagag gatggtgcca ttatcccagg tcctgccgtc 480

ggtgcctccc cagaaatgct ttcggtttcc gacaacatca tcattgaagt caacacaaaa 540

acaccatcct tcgaaggtat acacgacatt gacttgccag tcaacccacc attcagacaa 600

ccatacccac acaccactgc cgactacaga attggcagaa ccgccatccc agtcgaccca 660

gaaaaggtta tcgccattgt cgaaaccaca gagggtgaca aggttccacc aaacacccca 720

agtgacgctc aatcccaagc cattgctggt cacttgattg aattcttgga aaacgaagtc 780

aaacaaggtc gtctcccaga aaacttgcac ccattacaat ccggtatcgg taacattgcc 840

aacgctgtcg ttgaaggttt ggcccaatcc aacttcaaga acttgactgt ctggaccgaa 900

gtcttgcaag attccttctt ggacttcttc gaatccggaa gtttggactt tgcaaccgcc 960

acctccatca gattgaccga ggaaggtttc aagaagttct acgaaaactg ggacgtctac 1020

tccaagaagt tgtgcttgag atcgcaagtt gtctccaact cgccagaaat catcagaaga 1080

ttgggtgtct tggctttcaa caccccagtt gaagttgaca tctacgccca cgccaactcc 1140

accaacgtca tgggttccag aatgttgaac ggtttgggag gctcgggtga tttcttgaga 1200

aacgccaagt tgtccgtcat gcacacccca tctgctcgtc caaccaaggt tgacccaacc 1260

ggtgtctctt gtattgtccc aatggccacc cacgttgatc aaaccgaaca cgatctcgat 1320

gttgtcgtta ccgaacaagg tttggctgat ctcagaggtc tcagtccaaa ggaaagagct 1380

caagttataa tcaacaactg tgcccaccca gactaccagg ctcaattgca agactacttt 1440

gacagagctg ttttctacgc caccaagaag aagatgttgc acgaaccaca catgttgaac 1500

gaagctttcg ctatgcactt gaacttgcaa gaaaatggta ccatgaagtt gaacaaatag 1560

<210> 73
<211> 1548
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 73
atggctccat acctcaaaaa attgagaaca gcagaacaat gtgtcgactt gttcaagaac 60

ggccaatact tgggttggtc gggtttcacc ggtgtcggtg ccccaaaggc tgtcccagcc 120

gctttggttg accacgttga aaagaacaaa ttgcagggcc aaatgggctt ccacttgttt 180

gtcggtgctt ctgccggtcc agaagaaagc agatgggctg aaaacaacat gcttttatct 240

agatctccac atcaagtcgg taagccaatt gctgctgcca tcaatgacgg cagaactcag 300

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 193 of 3077

ttcttcgaca agcacttgtc catgttccca caagatttga cctacggctt ctacaccaag 360

aacaagccaa acggctccaa cttggattac accatcattg aagctaccgc cataacagag 420

gatggtgcca ttatcccagg tcctgccgtc ggtgcctccc cagaaatgct ttcggtttcc 480

gacaacatca tcattgaagt caacacaaaa acaccatcct tcgaaggtat acacgacatt 540

gacttgccag tcaacccacc attcagacaa ccatacccac acaccactgc cgactacaga 600

attggcagaa ccgccatccc agtcgaccca gaaaaggtta tcgccattgt cgaaaccaca 660

gagggtgaca aggttccacc aaacacccca agtgacgctc aatcccaagc cattgctggt 720

cacttgattg aattcttgga aaacgaagtc aaacaaggtc gtctcccaga aaacttgcac 780

ccattacaat ccggtatcgg taacattgcc aacgctgtcg ttgaaggttt ggcccaatcc 840

aacttcaaga acttgactgt ctggaccgaa gtcttgcaag attccttctt ggacttcttc 900

gaatccggaa gtttggactt tgcaaccgcc acctccatca gattgaccga ggaaggtttc 960

aagaagttct acgaaaactg ggacgtctac tccaagaagt tgtgcttgag atcgcaagtt 1020

gtctccaact cgccagaaat catcagaaga ttgggtgtct tggctttcaa caccccagtt 1080

gaagttgaca tctacgccca cgccaactcc accaacgtca tgggttccag aatgttgaac 1140

ggtttgggag gctcgggtga tttcttgaga aacgccaagt tgtccgtcat gcacacccca 1200

tctgctcgtc caaccaaggt tgacccaacc ggtgtctctt gtattgtccc aatggccacc 1260

cacgttgatc aaaccgaaca cgatctcgat gttgtcgtta ccgaacaagg tttggctgat 1320

ctcagaggtc tcagtccaaa ggaaagagct caagttataa tcaacaactg tgcccaccca 1380

gactaccagg ctcaattgca agactacttt gacagagctg ttttctacgc caccaagaag 1440

aagatgttgc acgaaccaca catgttgaac gaagctttcg ctatgcactt gaacttgcaa 1500

gaaaatggta ccatgaagtt gaacaaaggt agacgggcca agttgtag 1548

<210> 74
<211> 6672
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
polynucleotide

<220>
<221> modified_base
<222> (4079)..(4079)
<223> a, c, t, g, unknown or other

<400> 74
atgacagatc tttccccaag tccaacagac tcccttaatt acacacagtt gcactcatcc 60

ttgccatcac atttcttagg tgggaactcg gtgctcaccg ctgagccttc tgccgtgaca 120

gatttcgtca aaacacacca aggtcacact gttatcacca aagtcttgat tgccaacaac 180

ggtattggtg ccgtcaaaga aataagatcc gtcagaaaat gggcctacga aacttttggt 240

gacgaaagag ctatacagtt tgtcgccatg gccactcccg aagatatgga ggctaacgcc 300

gagtacattc gaatggccga ccagtttgtc gaggtcccag gtggtaccaa taacaacaac 360

tacgcgaatg ttgacttgat tgtcgaaatc gctgaaagaa ccgatgtcca cgccgtttgg 420

gctggttggg gtcatgcctc cgaaaaccct ttgttgccag raaggttggc agcttcccct 480

aagaagatcg tgtttattgg tcctccaggg tctgccatga gatctttggg tgacaagatt 540

tcttccacca ttgttgcaca acacgccaaa gtgccatgta tcccatggtc tggtactggt 600

gtcgaagagg tccacgtcga cccagaaacc aagttggtgt ctgttgacga ccacgtctac 660

gccaaaggtt gctgtacctc gccagaagac ggtttggaaa aagccaaacg tatcggattc 720

ccagttatgg ttaaggcatc cgaaggtggt ggtggtaaag gtatcagaaa agtcgaccac 780

gaaaaggact tcatcagttt gtacaaccag gcggctaacg aaataccagg gtcaccaatt 840

ttcatcatga agttggccgg tgacgccaga cacttggaag tgcaattgtt tgccgatcag 900

tacggtacca acatttcgct tttcggtaga gattgttctg tgcaaagaag acatcaaaag 960

atcattgaag aagctccagt cacaattgcc aacaaagaca cttttgttga gatggagaaa 1020

gctgccgtca gattgggtaa gttggttggt tacgtgtctg ccggtaccgt tgaatacctt 1080

tactcctacg ccgaagacaa gttctacttt ttggaattga acccaagatt gcaagttgaa 1140

catccaacta ccgaaatggt ttccggtgtc aacttaccag ccgctcagtt gcaaattgct 1200

atgggtctcc caatgcacag aatcagagac atcagattgt tgtacggtgt tgatccacac 1260

tctgccactg agattgattt cgagttcaag tccccaaact cattgatcac gcaaagaaag 1320

ccagctccaa agggtcactg taccgcttgt cgtatcactt ctgaagatcc aggtgaaggg 1380

ttcaagccaa gtggtggtac tcttcacgag ttgaacttcc gttcttcgtc caatgtctgg 1440

ggttacttct cggtcgccaa ccaatcttct atccactcct ttgctgattc ccagtttggt 1500

cacattttcg cctttggtga aaatcgtcaa gcctctagaa agcacatgat tgttgccttg 1560

aaggaattga gtatcagagg tgactttaga accactgttg aatacttgat caagttgttg 1620

gagactccag atttcgccga caacaccatc actaccggtt ggttggatga gttgatcacc 1680

aagaagttga ctgccgaaag accagatcct atcgttgctg ttgtctgtgg tgccgtcacc 1740

aaagcccaca tccaagccga agaagacaag aaggagtaca ttgagtcttt ggaaaagggt 1800

caagttccaa acaagtcctt gttgaaaact atcttcccag ttgagtttat ctacgaaggt 1860

gaaagataca agtttactgc caccaagtcc tccgaagaca agtacacttt gttcctcaac 1920

ggttctagat gtgtcattgg tgctcgctca ttgtctgatg gtggcttgtt gtgtgctttg 1980

gacggtaagt cccactctgt ctactggaag gaagaagcag cggccactag attgtctgtt 2040

gacggtaaga cttgcttgtt ggaagttgaa aacgacccaa cccaattgag aactccgtct 2100
ccaggtaagt tggtcaagta cttggttgag agtggtgaac acgttgatgc cggccaatct 2160
tatgccgaag ttgaagtcat gaagatgtgt atgcctttga ttgcacaaga aaacggtact 2220
gttcaattgc tcaaacaacc aggttccact cttaacgctg gtgacatctt ggcaatcttg 2280
gcattggacg atccatctaa agttaaacac gccaagccat atgaaggcac tttgccagag 2340
atgggtgatc caactgttac cggttccaaa ccagctcact tgttccaaca ttacgacacc 2400
atcttgaaga acatcttggc tggttacgat aaccaagtca ttttgaactc cactttgaag 2460
aacatgatgg agatcttgaa gaacaaggag ttgccttatt ctgaatggag attgcaaatc 2520
tccgccttgc attcaagaat cccaccaaag ttggatgagg ctttgacgtc cttgattgaa 2580
agaaccgaaa gcagaggcgc cgaattccca gctcgtcaga ttttgaagct cgtcaacaag 2640
actcttggtg aaccaggcaa cgaattgttg ggcgatgttg ttgctcctct tgtctccatt 2700
gccaaccgct accagaacgg cttggttgaa cacgagtacg actactttgc ttcattggtt 2760
aacgagtact gcaatgttga acacttcttt agtggtgaaa acgtgagaga agaagatgtt 2820
atcttgagat tgagagacga gaacaagtct gatttgaaga aggttatcag catttgcttg 2880
tcccactccc gtgtcagtgc taagaacaac ttgattttgg ccatcttgga agcttatgaa 2940
ccattgttgc aatccaactc ttcaactgcc gttgccatta gagattcttt gaagaagata 3000
gtccagttgg attctcgtgc ttgtgccaag gttggtttga aagctagaga acttttgatt 3060
caatgttctt tgccatccat caaggaaaga tctgaccaat tggaacacat tttgagaagt 3120
gcagtcgttg agacttctta tggtgaagtt ttcgccaagc acagagaacc taaattggaa 3180
atcatccaag aagttgtcga atccaagcac gttgttttcg atgtcttgtc gcaatttttg 3240
gtccaccaag actgctgggt tgccattgct gctgccgaag tctatgttag acgttcctac 3300
agagcttatg atttgggtaa gatcgattac cacattcatg acagattgcc aattgttgaa 3360
tggaagttca agttggctca aatcgcaggt tccagataca acgccgtcca atctgccagt 3420
gttggtgacg actcgaccac tatgaagcat gctgcatctg tttctgactt gtcgtttgtt 3480
gttgattcca agagcgaatc cacttccaga actggtgttt tggttccagc tagacatttg 3540
gacgatgttg atgagattct ttctgctgca ttggagtact tccaaccatc tgatgcactc 3600
tctttccaag ctaagggaga aagaccagag ttgttgaatg ttttgaacat tgtcatcacc 3660
gacattgacg gttactctga cgaagatgaa tgcttgaaga gaattcatga aatcttgaac 3720
gagtacgaag acgatttggt ctttgctggt gttcgtcgtg ttacttttgt tttcgcccac 3780
caagttggtt cttatccaaa gtactacacc ttcactggtc cagtgtatga agaaaacaag 3840
gttatcagac acatcgaacc agctttggct ttccaattgg aattgggaag attagccaac 3900

tttgacatca agccaatttt caccaataac agaaacattc atgtttacga agctattggt 3960

aagaatgctc cttcggataa gagattcttc actagaggta ttattagagg tggtgtcctc 4020

aaagatgaaa tcagtcttac tgagtacttg attgctgaat cgaacagatt gatcagtgnt 4080

atcttggata ccttggaagt tattgacact tccaactccg atttgaacca cattttcatc 4140

aacttctcca acgttttcaa cgtccaacca gctgatgttg aagctgcttt tgcttcattc 4200

ttggaaagat ttggtagaag attgtggaga ttgagagtta ctggtgctga aatcagaatt 4260

gtctgtaccg acccacaggg caactcattc ccattgcgtg ccattatcaa taacgtttca 4320

ggttatgttg tcaagtcgga attgtacttg gaagtgaaga accctaaggg tgattgggtc 4380

ttcaaatcca ttggccaccc tggctcaatg cacttgcaac caatctcgac tccataccca 4440

gtcaaggaat ccttgcagcc aaaacgttac agagctcaca acatgggaac cacttttgtt 4500

tacgatttcc cagagttgtt ccgtcaagcc accatttccc aatggaagaa gcacggcaag 4560

aaggctccta aagatgtctt cacttctttg gagttgatca ccgatgaaaa cgatgctttg 4620

gttgccgttg aaagagatcc aggtgccaac aagattggta tggttggttt caaggtcact 4680

gccaagaccc ctgaataccc acgcggacgt tcattcatca ttgttgccaa tgatatcacc 4740

cacaagattg gttcctttgg tccagatgaa gatgaatact tcaacaagtg taccgacttg 4800

gccagaaagt tgggtgttcc aagaatttac ctttctgcca actccggtgc cagaattggt 4860

gttgctgaag agttgattcc attgtaccaa gttgcttgga acgaagaagg taacccagat 4920

aaaggtttca gatacttgta cttgaaccca gacgccaaag aagctttgga aaaagacggc 4980

aagggtgaca ctattgttac tgaacgtatt gtcgaagatg gtcaagaacg tcacgttatc 5040

aaggccatta ttggtgctga gaacggcttg ggtgttgaat gtttgaaagg ttccggtttg 5100

attgctggtg ccacttcaag agcctacaga gacatcttca ccattacctt ggtcacttgt 5160

agatctgttg gtattggtgc ctatttggtc agattgggtc aaagagctat ccaaattgaa 5220

ggtcaaccaa tcattttgac tggtgcacca gctatcaaca agttgttggg tagagaagtt 5280

tactcgtcga acttgcaatt gggtggtacc cagatcatgt acaacaatgg tgtttcccac 5340

ttaactgcca gtgacgattt ggctggtgtt gagaagatca tggaatggtt gtcctacgtt 5400

ccagctaagc gtggtatgcc agtaccaatc ttggaaagtg aagatacctg ggacagagac 5460

attgactact acccaccaaa gcaagaagct ttcgacatca gatggatgat cgaaggtaag 5520

caagttgaag gtgaagagtt tgaatctggt ttgtttgaca aaggttcatt ccaggaaact 5580

ttatcaggat gggctaaagg tgttgtcgtt ggtagagctc gtctcggtgg tatcccaatt 5640

ggtgtcattg gtgttgagac cagaactatt gaaaacatga tcccagctga cccagccaac 5700

ccaagttcca ctgaagcctt gatccaagaa gccggtcaag tctggtatcc aaactctgcg 5760

ttcaagaccg cacaagccat taacgacttc aacaacggtg aacaattgcc attgatgatc 5820

ttggccaact ggagaggttt ctctggtggt cagagagata tgtacaacga ggtcttgaag 5880

tacggttcct tcattgttga cgctttagtt gatttcaagc agccaatctt cacttacatc 5940

ccaccaaatg gtgaattgag aggtggctct tgggtcgttg ttgatccaac catcaactcc 6000

gacatgatgg aaatgtatgc cgacgttgac tccagagctg gtgttttgga accagaaggt 6060

atggttggta tcaaatacag acgggacaag ttgttggcta ccatgcaaag attggatcca 6120

acttatgccc aattgaagga gaagttgaac gactcgagct tgtcgccaga agaacatgcc 6180

caagtcagca ccaagattgt caagcgtgaa aaggcattgt tgccaatcta tgcccaaatt 6240

tctgtccagt ttgccgactt gcacgacaga tccggacgta tgatggctaa aggtgtcatt 6300

agaaaagaaa tcaagtgggt tgacgccaga cgtttcttct tctggagatt gagaagaaga 6360

ttgaacgaag agtacgtttt gaagttgatt ggtgaacagg tcaagaatgc caacaagttg 6420

gaaaaggttg ccaggttgaa gagttggatg ccaactgttg actacgacga tgaccaagct 6480

gtcagtactt ggattgaaga gaaccacgcc aaattgcaaa agagagttga agaattgaga 6540

caggagaaga acaagtccga cattgtcaaa ttgttgcaag aagacccatc aaacgctgcc 6600

tctgttatga gggatttcgt tgatagattg tccgatgaag aaaaggaaaa gttccttaaa 6660

tcattgaact ag 6672

<210> 75
<211> 6825
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 75
atgagatgcc aagtatctca accatcacga tttactaact tgcttgtaca tagactccca 60

cgaacactac ttaattatcc agttgtaaat accctattta ttcctagacg tcattattcc 120

cttaattttt cattcaagaa cctactaaag aaaatgacag atctttcccc aagtccaaca 180

gactccctta attacacaca gttgcactca tccttgccat cacatttctt aggtgggaac 240

tcggtgctca ccgctgagcc ttctgccgtg acagatttcg tcaaaacaca ccaaggtcac 300

actgttatca ccaaagtctt gattgccaac aacggtattg gtgccgtcaa agaaataaga 360

tccgtcagaa aatgggccta cgaaactttt ggtgacgaaa gagctataca gtttgtcgcc 420

atggccactc ccgaagatat ggaggctaac gccgagtaca ttcgaatggc cgaccagttt 480

gtcgaggtcc caggtggtac caataacaac aactacgcga atgttgactt gattgtcgaa 540

atcgctgaaa gaaccgatgt ccacgccgtt tgggctggtt ggggtcatgc ctccgaaaac 600

cctttgttgc cagaaaggtt ggcagcttcc cctaagaaga tcgtgtttat tggtcctcca 660

gggtctgcca tgagatcttt gggtgacaag atttcttcca ccattgttgc acaacacgcc 720
aaagtgccat gtatcccatg gtctggtact ggtgtcgaag aggtccacgt cgacccagaa 780
accaagttgg tgtctgttga cgaccacgtc tacgccaaag gttgctgtac ctcgccagaa 840
gacggtttgg aaaaagccaa acgtatcgga ttcccagtta tggttaaggc atccgaaggt 900
ggtggtggta aaggtatcag aaaagtcgac cacgaaaagg acttcatcag tttgtacaac 960
caggcggcta acgaaatacc agggtcacca attttcatca tgaagttggc cggtgacgcc 1020
agacacttgg aagtgcaatt gtttgccgat cagtacggta ccaacatttc gcttttcggt 1080
agagattgtt ctgtgcaaag aagacatcaa aagatcattg aagaagctcc agtcacaatt 1140
gccaacaaag acacttttgt tgagatggag aaagctgccg tcagattggg taagttggtt 1200
ggttacgtgt ctgccggtac cgttgaatac ctttactcct acgccgaaga caagttctac 1260
tttttggaat tgaacccaag attgcaagtt gaacatccaa ctaccgaaat ggtttccggt 1320
gtcaacttac cagccgctca gttgcaaatt gctatgggtc tcccaatgca cagaatcaga 1380
gacatcagat tgttgtacgg tgttgatcca cactctgcca ctgagattga tttcgagttc 1440
aagtccccaa actcattgat cacgcaaaga aagccagctc caaagggtca ctgtaccgct 1500
tgtcgtatca cttctgaaga tccaggtgaa gggttcaagc caagtggtgg tactcttcac 1560
gagttgaact tccgttcttc gtccaatgtc tggggttact tctcggtcgc caaccaatct 1620
tctatccact cctttgctga ttcccagttt ggtcacattt tcgcctttgg tgaaaatcgt 1680
caagcctcta gaaagcacat gattgttgcc ttgaaggaat tgagtatcag aggtgacttt 1740
agaaccactg ttgaatactt gatcaagttg ttggagactc cagatttcgc cgacaacacc 1800
atcactaccg gttggttgga tgagttgatc accaagaagt tgactgccga aagaccagat 1860
cctatcgttg ctgttgtctg tggtgccgtc accaaagccc acatccaagc cgaagaagac 1920
aagaaggagt acattgagtc tttggaaaag ggtcaagttc caaacaagtc cttgttgaaa 1980
actatcttcc cagttgagtt tatctacgaa ggtgaaagat acaagtttac tgccaccaag 2040
tcctccgaag acaagtacac tttgttcctc aacggttcta gatgtgtcat tggtgctcgc 2100
tcattgtctg atggtggctt gttgtgtgct ttggacggta agtcccactc tgtctactgg 2160
aaggaagaag cagcggccac tagattgtct gttgacggta agacttgctt gttggaagtt 2220
gaaaacgacc caacccaatt gagaactccg tctccaggta agttggtcaa gtacttggtt 2280
gagagtggtg aacacgttga tgccggccaa tcttatgccg aagttgaagt catgaagatg 2340
tgtatgcctt tgattgcaca agaaaacggt actgttcaat tgctcaaaca accaggttcc 2400
actcttaacg ctggtgacat cttggcaatc ttggcattgg acgatccatc taaagttaaa 2460
cacgccaagc catatgaagg cactttgcca gagatgggtg atccaactgt taccggttcc 2520

aaaccagctc acttgttcca acattacgac accatcttga agaacatctt ggctggttac 2580

gataaccaag tcattttgaa ctccactttg aagaacatga tggagatctt gaagaacaag 2640

gagttgcctt attctgaatg gagattgcaa atctccgcct tgcattcaag aatcccacca 2700

aagttggatg aggctttgac gtccttgatt gaaagaaccg aaagcagagg cgccgaattc 2760

ccagctcgtc agattttgaa gctcgtcaac aagactcttg gtgaaccagg caacgaattg 2820

ttgggcgatg ttgttgctcc tcttgtctcc attgccaacc gctaccagaa cggcttggtt 2880

gaacacgagt acgactactt tgcttcattg gttaacgagt actgcaatgt tgaacacttc 2940

tttagtggtg aaaacgtgag agaagaagat gttatcttga gattgagaga cgagaacaag 3000

tctgatttga agaaggttat cagcatttgc ttgtcccact cccgtgtcag tgctaagaac 3060

aacttgattt tggccatctt ggaagcttat gaaccattgt tgcaatccaa ctcttcaact 3120

gccgttgcca ttagagattc tttgaagaag atagtccagt tggattctcg tgcttgtgcc 3180

aaggttggtt tgaaagctag agaacttttg attcaatgtt ctttgccatc catcaaggaa 3240

agatctgacc aattggaaca cattttgaga agtgcagtcg ttgagacttc ttatggtgaa 3300

gttttcgcca agcacagaga acctaaattg gaaatcatcc aagaagttgt cgaatccaag 3360

cacgttgttt tcgatgtctt gtcgcaattt ttggtccacc aagactgctg ggttgccatt 3420

gctgctgccg aagtctatgt tagacgttcc tacagagctt atgatttggg taagatcgat 3480

taccacattc atgacagatt gccaattgtt gaatggaagt tcaagttggc tcaaatcgca 3540

ggttccagat acaacgccgt ccaatctgcc agtgttggtg acgactcgac cactatgaag 3600

catgctgcat ctgtttctga cttgtcgttt gttgttgatt ccaagagcga atccacttcc 3660

agaactggtg ttttggttcc agctagacat ttggacgatg ttgatgagat tctttctgct 3720

gcattggagt acttccaacc atctgatgca ctctctttcc aagctaaggg agaaagacca 3780

gagttgttga atgttttgaa cattgtcatc accgacattg acggttactc tgacgaagat 3840

gaatgcttga agagaattca tgaaatcttg aacgagtacg aagacgattt ggtctttgct 3900

ggtgttcgtc gtgttacttt tgttttcgcc caccaagttg gttcttatcc aaagtactac 3960

accttcactg gtccagtgta tgaagaaaac aaggttatca gacacatcga accagctttg 4020

gctttccaat tggaattggg aagattagcc aactttgaca tcaagccaat tttcaccaat 4080

aacagaaaca ttcatgttta cgaagctatt ggtaagaatg ctccttcgga taagagattc 4140

ttcactagag gtattattag aggtggtgtc ctcaaagatg aaatcagtct tactgagtac 4200

ttgattgctg aatcgaacag attgatcagt gatatcttgg ataccttgga agttattgac 4260

acttccaact ccgatttgaa ccacattttc atcaacttct ccaacgtttt caacgtccaa 4320

ccagctgatg ttgaagctgc ttttgcttca ttcttggaaa gatttggtag aagattgtgg 4380

agattgagag ttactggtgc tgaaatcaga attgtctgta ccgacccaca gggcaactca 4440

ttcccattgc gtgccattat caataacgtt tcaggttatg ttgtcaagtc ggaattgtac 4500
ttggaagtga agaaccctaa gggtgattgg gtcttcaaat ccattggcca ccctggctca 4560
atgcacttgc aaccaatctc gactccatac ccagtcaagg aatccttgca gccaaaacgt 4620
tacagagctc acaacatggg aaccactttt gtttacgatt tcccagagtt gttccgtcaa 4680
gccaccattt cccaatggaa gaagcacggc aagaaggctc ctaaagatgt cttcacttct 4740
ttggagttga tcaccgatga aaacgatgct ttggttgccg ttgaaagaga tccaggtgcc 4800
aacaagattg gtatggttgg tttcaaggtc actgccaaga cccctgaata cccacgcgga 4860
cgttcattca tcattgttgc caatgatatc acccacaaga ttggttcctt tggtccagat 4920
gaagatgaat acttcaacaa gtgtaccgac ttggccagaa agttgggtgt tccaagaatt 4980
taccttttctg ccaactccgg tgccagaatt ggtgttgctg aagagttgat tccattgtac 5040
caagttgctt ggaacgaaga aggtaaccca gataaaggtt tcagatactt gtacttgaac 5100
ccagacgcca aagaagcttt ggaaaaagac ggcaagggtg acactattgt tactgaacgt 5160
attgtcgaag atggtcaaga acgtcacgtt atcaaggcca ttattggtgc tgagaacggc 5220
ttgggtgttg aatgtttgaa aggttccggt ttgattgctg gtgccacttc aagagcctac 5280
agagacatct tcaccattac cttggtcact tgtagatctg ttggtattgg tgcctatttg 5340
gtcagattgg gtcaaagagc tatccaaatt gaaggtcaac caatcatttt gactggtgca 5400
ccagctatca acaagttgtt gggtagagaa gtttactcgt cgaacttgca attgggtggt 5460
acccagatca tgtacaacaa tggtgtttcc cacttaactg ccagtgacga tttggctggt 5520
gttgagaaga tcatggaatg gttgtcctac gttccagcta agcgtggtat gccagtacca 5580
atcttggaaa gtgaagatac ctgggacaga gacattgact actacccacc aaagcaagaa 5640
gctttcgaca tcagatggat gatcgaaggt aagcaagttg aaggtgaaga gtttgaatct 5700
ggtttgtttg acaaaggttc attccaggaa actttatcag gatgggctaa aggtgttgtc 5760
gttggtagag ctcgtctcgg tggtatccca attggtgtca ttggtgttga gaccagaact 5820
attgaaaaca tgatcccagc tgacccagcc aacccaagtt ccactgaagc cttgatccaa 5880
gaagccggtc aagtctggta tccaaactct gcgttcaaga ccgcacaagc cattaacgac 5940
ttcaacaacg gtgaacaatt gccattgatg atcttggcca actggagagg tttctctggt 6000
ggtcagagag atatgtacaa cgaggtcttg aagtacggtt ccttcattgt tgacgcttta 6060
gttgatttca agcagccaat cttcacttac atcccaccaa atggtgaatt gagaggtggc 6120
tcttgggtcg ttgttgatcc aaccatcaac tccgacatga tggaaatgta tgccgacgtt 6180
gactccagag ctggtgtttt ggaaccagaa ggtatggttg gtatcaaata cagacgggac 6240
aagttgttgg ctaccatgca aagattggat ccaacttatg cccaattgaa ggagaagttg 6300

aacgactcga gcttgtcgcc agaagaacat gcccaagtca gcaccaagat tgtcaagcgt 6360

gaaaaggcat tgttgccaat ctatgcccaa atttctgtcc agtttgccga cttgcacgac 6420

agatccggac gtatgatggc taaaggtgtc attagaaaag aaatcaagtg ggttgacgcc 6480

agacgtttct tcttctggag attgagaaga agattgaacg aagagtacgt tttgaagttg 6540

attggtgaac aggtcaagaa tgccaacaag ttggaaaagg ttgccaggtt gaagagttgg 6600

atgccaactg ttgactacga cgatgaccaa gctgtcagta cttggattga agagaaccac 6660

gccaaattgc aaaagagagt tgaagaattg agacaggaga agaacaagtc cgacattgtc 6720

aaattgttgc aagaagaccc atcaaacgct gcctctgtta tgagggattt cgttgataga 6780

ttgtccgatg aagaaaagga aaagttcctt aaatcattga actag 6825

<210> 76
<211> 2028
<212> DNA
<213> Candida viswanathii

<400> 76
atgccagaat caactcaaca atctcacatt tctttggacc atgaaaagat gcagcaacct 60

ccaactggct tcaccgaacg ctccaccgca aagccaaact tggccgactt tgacacctac 120

aagaagttgt acaagcagtc ggtagaaaac ccaaacgagt ttttcacaga acaagcaaac 180

aaaaacttgg actggttcaa acctttcgac ttggccagat tccccgtcga tccaaaggac 240

gacttcaaga atggcgactt gcctgcctgg ttcatcaacg gtcagttgaa cgcttcttat 300

aacgcggttg atagatgggc catcaagaac ccagacaagc cagctatcat ttatgaaggt 360

gacgaaccag acctgggcag aatcatcact tatggcgaat tgttgaaaga agtcagtaag 420

ttggcccaga ctttgaccaa gttgggtgtc aaaaaaggtg actctgttgc tgtctacttg 480

ccaatgatcc ctgaagctat cgtcaccttg ttggccattg tcagaatcgg tgccgtgcac 540

tcggttgtct ttgctggttt ctcctccgct tccttgagag acagaatctt ggatgctgac 600

tccagaatcg tcatcactgc cgacgagtcc aagagaggtg gtaagaccat tgaaaccaag 660

aagattgtcg atgacgcctt gaaggaatgt cctaaggtta gaaacgttat tgttttcaag 720

agaaccggca actcccacgt tcctttctct gctggtagag acttgtggtg gcacgaagaa 780

atggccaaat acggtccata cttcccacca gtgccagtca actccgaaga tccattgttc 840

ttgttgtaca cttctggttc tactggtaag ccaaagggtg tccaacacaa caccgccggt 900

tacttgttgg gcgctatctt gaccaccaag tacacttttg atgtccacga agaagacgtc 960

ttgttcaccg ctggtgacgt cggttggatc acgggtcaca cctactgtgt ctacggccca 1020

ttgttgtgtg gtgccaccac cgttgtcttt gaaggtaccc cagcttatcc aaactactcc 1080

agatactggg agattgtcga ccaatacaag gtcaaccaat tctacgttgc cccaaccgct 1140

ttgagattgt tgaagagagc cggtaccaag tacgttgaaa aatacgactt gtcctccttg 1200

agagtcttgg gctctgtcgg tgaaccaatt gcggccgagg tctggcactg gtacaatgac 1260

aacattggta gaggcaaggc ccacattgtc gatacctact ggcaaaccga atctggttcc 1320

cacttgttga ccccattggc cggtgtcact ccaaccaagc caggttccgc ttccttgcca 1380

ttctttggtg ttgttccaaa gatcttggac ccaaccaccg gtgaagaatt agaaggtaac 1440

gatgttgaag gtgtccttgc cattaaatcc gcttggccat ccattactag aggtatttac 1500

aacgattaca atagatttat cgaaacctac ttggccccat accacgacca ctacttttcc 1560

ggtgatggtg ctgctagaga taacgacggc ttctactgga tcttgggtag agtcgatgat 1620

gtcgtcaatg tttccggtca cagattgtcg actgctgaaa ttgaagctgc cttgattgag 1680

cacccattgg ttgctgaatc cgctgttgtt ggttatgctg acgacttgac tggtcaagcc 1740

gtcgctgcct atgtgtcctt gaagaaagat aaagttattg gtgacgacat ggaggcagtc 1800

aagaaggagt tgatcttgac tgtcagaaaa gaaattggtc catttgctgc accaaagttg 1860

atcttgttgg ttgacgactt gccaaagacc agatccggta agatcatgag aagaatcttg 1920

agaaaggtct tggctggtga agaagatcaa ttgggtgata tctccacctt gtccaaccca 1980

ggcgttgtcc aacaaatcat tgaaattgtc catgcctgca agaaataa 2028

<210> 77
<211> 2016
<212> DNA
<213> Candida viswanathii

<400> 77
atgactacag actcaaacac ccacaaggtt gttcacgaag ccaatggcgt caagttgaga 60

caaaccccaa aggaattctt ccaaaggcaa ccaaacaagg gacacatcca agacacggaa 120

gagtacaaga aattgtatga gcaatccata aaagaccccc aaggtttctt tggcccattg 180

gctaaagaat tgttgtcctg ggacagcgat ttccacaccg tcaagtcagg tacccttaag 240

aacggtgacg ctgcctggtt cttgggcggt aagttgaacg cttcttacaa ctgtgttgcc 300

agacacgctt tggccaaccc aaacaagcca gccatcatct acgaagcaga tgaagaaagg 360

gattcttaca tcttgaccta cggtgatttg ttgagagaag tttccaaagt cgctggtgtc 420

ttgcactcct ggggcatcag aaagggtgac accgttgctg tttacttacc aatgaatgcc 480

caagccctca ttgccatgtt ggccattgct agattgggtg ctgcccactc cgtcatcttt 540

gctgggttct cctccggttc tatcaaagac agagtcaacg acgccagctg taaagcgctt 600

attacctgtg atgaaggtag aagaggtggc agaaccacca acatcaagaa gttgtgcgac 660

gaagctttga agcaatgccc aacggtcaag cacgtcttgg tccacagaag aaccggcaac 720

ccagaaatca aattgaccga aggcagagat tactactggg acgaagaaac cgccaagttt 780

tccggatact tcccacccgt gccagttgac tctgaagacc cattgttctt gttgtacact 840

tctggttcca ctggtactcc aaagggtgtt gtccacacca ccgccggtta cttgttgggt 900

gctgctcttt ccaccaaata catctttgac gtccacccag aagatgtctt cttcactgct 960

ggtgatgtcg gttggatcac tggtcacacc tacgctttat atggtccatt gttgttgggt 1020

gttccttctg ttgttttcga aggtactcca gcttacccag actacggaag attctggcaa 1080

attgttgaaa aacacaaggc tacccacttc tacgttgctc caactgcttt gagattgttg 1140

cgtaaagctg gtgaacaaga agttggcaag tacgacttga gttcattgag aaccttgggt 1200

tctgttggtg aaccaatctc cccagatatt tgggaatggt acaatgaatt tgtcggtaag 1260

gaccaatgtc atatctcgga tacctactgg caaaccgaat ccggttctca cttgattgct 1320

ccattggctg gtgccattgc caacaagcca ggttctgctt cttacccatt ctttggtatt 1380

gaagctgctt tgatcgaccc agtcactggt gttgaaatca agggcaacga tgttgaaggt 1440

gtgttggttg tcaaggacca ctggccatcc atggcaagaa ctgttttcaa caaccacgtc 1500

aagtacatgg atacctacat gaacccatac ccaggctact atttcactgg tgatggtgct 1560

gccagagata acgacggtta ctactggatc agaggtagag ttgatgatgt cgtcaatgtt 1620

tctggccaca gattgtccac cgccgaaatt gaatctgcat tgattgaaga cccaagagtt 1680

ggtgaatccg ctgttgttgg tatcaacgat gaccttactg gtcaagctgt tattgcttac 1740

gttgccttga aatccggcga cgttgctgat gaagacgctt tgagaaagga attgatcttg 1800

ttggttagaa aagaaattgg tccatttgcg gctccaaaat ctgtcatctt ggttcaagac 1860

ttgcctaaga ccagatctgg taagatcatg agaagaatct tgagaaaagt cagttctaac 1920

gaagccgatc aattgggtga catcaccact ttgcaaaacc caggttctgt tgaaggtatg 1980

atttctgcct tcggtgctca attcggtaga aagtaa 2016

<210> 78
<211> 1152
<212> DNA
<213> Candida viswanathii

<400> 78
atgcttacac taacttctgg tcccaacccc ttgcctgatt tcgaagaagc cttgagagtc 60

atcaaagtgg acgatacaca ctatgttggt gcccattcgt tgagattgcc tgtgaaagga 120

ggacgaggcg tttatggtgg ccatatgatt gcccagtctt tgctagtggg tattgaatca 180

acaagagatg acaagactaa taaagtgttc attcctgact cgtaccacct gtactttatt 240

ggagccggaa atgccaaaat ccccatgaac tacacggtag agaaactata cgacgatgaa 300

aacgtcagca agcgtttcat cattgcggaa caaaagggaa gacatcgttt gacttgtttg 360

gttacattga ggagacccgg taccaagcct ttccacgaca gcgacaactt ggacatcagc 420

atccccgtgc ccaagatcca gttaaaacac cctgatcccg acaagttgca ccaagttcag 480

cacacggact ttattaggaa tgcattcggt aaagagctta tggattacag agagtgccct 540

gaagagaacg aactttatgc cgcagaaaga tggttgactg tgttcaccgg tattcgaaac 600

cagcctaaac cgggcgccag tttggaaacg gtcgtcgaag aattgccaga cgcccagggc 660

cagatgcaca cagtggagaa gagtatttta agacccaaag attcccagtc gttcaaggac 720

cctatctaca actttgtggg tttagccgac ttgtctgatt cggcattcct caccaccatg 780

gcgagaatct tgcacatccc atgggcgcct tccatcgaaa tcgacgacac atacgatcca 840

gcaagagatg caacgtatat catgcgttca actttgaatg cagcccatat cttccattac 900

aatgcgatgt cattggatca ccatatttat ttccacaatg aagattacac ttccgacgac 960

ggttcaggtt tcgacatttg caaagattgg ttggctttca cttatcagat gaaaagactc 1020

tctaataacc gtaccctcgt ccgtgggttc ttgttcaacg aaaaacacaa atgtatcgcc 1080

acagtggtgc aagaagggtt gactattatg caaaatggag ttggtcgaac tgctgataaa 1140

tcacgtctat ag 1152

<210> 79
<211> 981
<212> DNA
<213> Candida viswanathii

<400> 79
atgattgaga acatttcagg aaacggaaac tacccacaga atcacgaggt tgacttggag 60

aaggaattcg gagttgaaaa aataggtatc aacttgtacc gtggcaagag cccaataccg 120

aagccagaca gacgttccag aggtgcttac ggtggttact tggccggtca agctttgtta 180

gttgccatga agtctacccc gccagagtac agaccacacc tgttccattc ctactttatc 240

aaagctgtca atgacaagga gacattggag tggagagttg aagaaacatc caacggcaga 300

aactacgcca acagatcctt gcaagccttc caagccggta acttggtcta cacagccaac 360

gtctccttga ccaagaaaaa ctctgccaag aaagctgagg aagcaacggg tgtcaagccc 420

tttgagttcc agggaaaacc acacgaatgg tttgaaaaac acaagagaga tgaccttcct 480

cttgctacac cttcttcatc ccttttgatc taccacaaat tcttcccaga ggttgtctcg 540

ttggaagctt ccaaagaaga agaatcaaag ccagcagccg acagagaatt gagctggtac 600

ttcaaatggg gtatcaacaa cgaagagggc caccaccagc cattagtcaa cttgaactcc 660

gattaccaat atgttggaat ggctgcgttg accgatgctg tctacttgaa cagattattg 720

agaattttga gagttgaaga cgctgaccac acccagcttg tccactactt ttccgtctcc 780

ttggaccata ctatgtactt ccatgacgac gactttgacg taaccaaatg gatgggattc 840

actttcaagg ttaccagatt cagtcacaac cgtgcattgt gccagggtga agtttataac 900

gacaaggggg tacacgtgtg tacaattgtc caggaagggc tcatgatgct taacggtctt 960

gaacaaggtg ctaaattata a 981

<210> 80
<211> 981
<212> DNA
<213> Candida viswanathii

<400> 80
atgattgaga gcatttcaga aaacggaaac tacccacaga accacgaagt tgacttggag 60

aaggaattcg gagttgaaaa aataggtatc aacttgtacc gtggcaagag cccaataccg 120

aagccagaca gacgttccag aggtgcttac ggtggttact tggctggtca agctttgtta 180

gttgccatga aatctacccc gccagattac agaccacatc tgttccattc ctactttatc 240

aaagctgtca atgacaaggt gacattggag tggagagttg aagaaacatc caacggccga 300

aactacgcca acagagcctt gcaagccttc caagccggtg acttggtcta cacagctaac 360

gtctccttga ccaagaaaaa ctccgccaag aaaactgagg aagcaacggg tgtcaagccc 420

tttgagttcc aggggaagcc acacgaatgg tttgataagc acaagataga tgaccttcct 480

cttgctacgc cttcttcatc ccttttgatc taccacaaat tcttcccaga ggttgtgtcg 540

ttggaagctt ccaaagaaga ggaatcaaag ccagcagctg acagagaatt gagctggtat 600

ttcaaatggg gtatcaacaa cgaagagggc caccaccagc cattagtcaa cttgaactct 660

gattaccaat atgttggtat ggctgcgttg accgatgccg tctacttgaa cagattattg 720

agaatcttga gagttcaaga cgctgaccac acccaacttg tccactactt ttccgtctcc 780

ttggaccaca ccatgtactt ccatgacgac gactttgatg ttaccaaatg gatgggattc 840

actttcaagg ttaccagatt cagtcacaac cgtgcattgt gccaaggtga agtttacaac 900

gacaaggggg tacacgtgtg tacaattgtt caagaaggac tcatgatgct taacggtctt 960

gaagaaggtg ctaaattata a 981

<210> 81
<211> 840
<212> DNA
<213> Candida viswanathii

<400> 81
atgagtcacc caacgccaga agaagtctac ggagtcacca aagttgccga aaacaaatac 60

gtggggaaca gacccttgaa caagcccacg ccgaaaacac gcggagtgta tggcggaaac 120

ttctgtgccc aggccatctt ggtagccatc gagtcagccc cggaagggtt cacaccacac 180

tccatccaca gcaacttcat tcgtggcggc gaccccgaag tccctgttga gtgggaggtc 240

gaggtcatct cgaacggcaa gtctttcgcc aaccgcatag tcaagggcgt gcagcacggc 300

atcgtggtct acgtcgccac ggtctcgttg accaacaaga actcaaccac ccgcaacgag 360

agcttcacgt acgacacgcc cccggacgag accgtcaaga cgtacggcaa cgccgagcta 420

gacacatact accagggctg gctctacttg gaggtcaaga actaccccaa gcagttgcac 480

tcgcaccaga tctcgtactc cgtcaagtgg gggccggaga acgacgcgtg gaaggacgcc 540

agccagacgt accagtttgt tgggctcgct gcgatctccg acgtgcttga cctcggccag 600

atcttgcgca acttggacat ccacttgagc acgcccaagt tcaacgtgtc gctcgaccac 660

agtgtgtact tccacggcgc ggactttgac gtgaccaggt ggtcgacgac cacgatccgc 720

atgacgaagc tcgcgcacgg ccgcgcgttg atagagggcg agatgtactc ggaccagggc 780

aggcacatcg ccagcatagt gcaagagagg ttgtacatag ctgagtcgcc gaagctctga 840

<210> 82
<211> 897
<212> DNA
<213> Candida viswanathii

<400> 82
atgacgctga ttgcacccgc cacagacata cccgccgtgg acatctccac catatacgac 60

gtcaaacaga tcgacgccaa ccgctaccgc ggcgttcggc cgctccagaa acaggcacgc 120

gagttccgcg gagtctttgg aggcaacttg gttgcccagt ccgtggtggt ggccttgcgg 180

tcggtgcccg caggattcca ccccaactcc gtccacgcct actacgtgcg tgccgtgacc 240

gacgagacgc caattgagtg ggaggtcgaa gagacaacca cgggccgcac cttcgccaac 300

cggctgatac gcggattgca gaacaagaag gtggtgttca cggcgagcat ctcgttgacc 360

aggaagaact ccaatgcgga agtgatccaa aagaccgggc acccgtcgtt gcagttccag 420

cgggacaccg agccatacta cgagcagatg cgagcccgcc ctggcgagtg caaggtgctc 480

tacattaacg ccaacacgca cctcacggtc agacagttcc ctgaagtcag ctcccgcgac 540

gtgttctcgt tcttggtgcg gtttggcaac gacgggcggg agaggatcgt ggggatgact 600

cccgagtacc agtacgccgc gttggctgcg ttgtcggact gggtgaggct ccgtttctac 660

tttgacaata tgggggtcga cgtgcagacg tcgttcgacg tgtcgttgga ccacagcatc 720

tacttccacg acgacgggtt cgacgctacc gagtacttgg ttttctccgt caaggtgtcg 780

cggatcagcc acagccgggt gctttacgac gggcagatct tcaacgaccg tggagcccac 840

gttgctagca tcagacagga gcgcttatat gtcgtcagca acaagccaaa gttttag 897

<210> 83
<211> 1062
<212> DNA
<213> Candida viswanathii

<400> 83
atgcccacat tcaactacaa ggacggcgaa acaatagacg tgcagaaaga gtttggcgtg 60

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 207 of 3077

gtcgagacag cgcccaacaa gtacgtcggc gtgaagccac tcgtcaagcc gatgcctcat 120

gtaaagggag tctttggcgg taaccttgca ggacaagcat tgctagtcgc catgaagtcg 180

gttggcccag actttagccc acactccttg cattcgtact ttatacgcgc cggcagcgac 240

cagacacccg tcgagtggac tgtgcaggcg atatctgatg gtaactcgtt ttgtaacagg 300

tttatcaagg gagtccaaaa cggacaggtc atatacattg cgaatgtgtc gttgactaaa 360

cgcaactctg ctgctgatgc gatgaagaag tacgaggagt accatgcgca gatccgccag 420

aagggcaagg acggcgatgc tgatgaagag gacgaggatg acgatgacga ggatgataat 480

gccccagcta agccatttgg gttccagact ccgagccata agtggatcaa ggaccgcgat 540

ctcgacaagt tgccagtttc cgacatggag tcgaatcttt tgttgtacta caagttgcct 600

cccgagttcg tgagcttgaa gtcttcaaca gaggaagagc tgttgccagt ctcagagcgc 660

aggatgggtg ctttggcgaa gtggggtatt gagaacgagc aagggttcaa ccagcctttg 720

actaatctcg acaagtcgtt ccagtatgtt ggtttggcca atatcactga tgggttgtac 780

ttgggcacat tgaataggat attgcgcatt gatgacttga cacttgatga gcgggcgact 840

aactacttta gcgtatcatt ggaccacgtc atctatttcc acgatgatga ttttgatgtg 900

accaagtgga tgggctttac gttcaggtgt tcgagatatt cacataatcg ggttattttt 960

gaaggtgaga tctacagtga caagggagtc caggttgcat ctattattca agagggctta 1020

gttagattta aagacggcta cttgaaaaac gccaagttgt ga 1062

<210> 84
<211> 1050
<212> DNA
<213> Candida viswanathii

<400> 84
atgtcaacca ccggaatcta caacgccaaa cccgcggctg ccatattgga aacaaaaacc 60

aaggtcaagt tagtttatac tgaaggaccg aaattggtct acgaaggatt gcaccccgtt 120

gagaccataa aagaaactgt tagaggtact tatggtggtg attttattgc tcaaggggtg 180

aacgcagctt gggaatcaat tggcaacagg atggattttc aacctcattc attgcacgca 240

tactatataa aaagcggttc ccaggagtca gtattacgct gggaagtctt caaggtgagt 300

gatagcagaa gttttgccaa ccggatgtta actgcatacc agacgcacac caaccaattg 360

gtgttcacga tgcagatatc gttcaccaaa gacaacaacg aggagatcaa gcgtgcggaa 420

tacgaaaagt tgttggccag cggtggcaaa atcagaagca tcccatttgt catcaaaaag 480

gctccgaatg aaaagtattt caagttacga gacagaatca atgagcttca ttattctgag 540

tataccaatg ggatcattgc catggcttcc ccccttgact tccatgagta tgccactgag 600

atgaactatg acaccttggg aaacaaagag tacgggacgt ttatgaaaat cttggataac 660

tattccttag gcaaggatta ttctaaacaa tccttcttgg ggttgtcatt cttatccgat 720

ggtgtttggg tattttgctt tactagagct ttggggttgc cacttggtag ctacgaaaag 780

gagttctttc gtgtcagttt ggaccacacg gtatatttcc acgacgctaa tttcgactcc 840

agtgagtggc tcttcttaga ttttaggttt gtgaacatga agaacaacag attcttgact 900

gtgttgaact actatacttt gcaaggcaaa cttattgcga ctgtgctcca agaagtctac 960

acttctttac atcagggtat cattgacaag tcgcaagaaa tagctgtcaa aagcggaaat 1020

aagaagcagg tcacaactcc aaagttgtag 1050

<210> 85
<211> 1056
<212> DNA
<213> Candida viswanathii

<400> 85
atggacaaac tacaagctga agtatacgag gccgaacctc ccgtggccaa attggaagcc 60

aaaacagcag ccaaattgat atcttctgat ggggcaaaat tgacctacga aggagtgtac 120

cctgtggagc tagtcagaaa aggtctcagg ggaacctacg ggggagattt catagcacag 180

ggaatcaatg ttgcctggga gtcaattggc aataagactg actttcaacc ccattcatta 240

catgcatatt ttgtcaaggc aggatccgac ctgtctgtcc tacgttggga agtgctcaag 300

gtgagtgata gcagaaattt tgccaaccgt ctaatgttag cgtaccagac tcacaccaac 360

gaattggtgt tcacaatgca gatatcgttc actaaggata acaacgagga gatcaagcgt 420

gccgagtaca agcaattatt gcagagtggc ggtaaaataa gaagtatccc gtttgcaatc 480

aagaaaccac ccaatgaaaa gtactttaaa ttgaaagaca aggtggacga tcttccatat 540

tttgagcaca ccaatgggaa tatggcgact gcgattccac cagacttcct tgaatacgcc 600

acagagatga atcacgacac ggtggggaac aaagagtttg gtattttcat gaaagtattg 660

gatgattact ccttgggcaa aaactatgag agacagtcgt tcttggggtt ggccttcttg 720

tctgatgccg tgtggttatc gagtttcacc ccagccttag ggttgccatt gggcacttta 780

gaaaggaaat tctttcgtgt tagcttagac cacaccatgt atttccatga cgccaacttt 840

gattctagtg agtggatctt tgtggatttc aggtttgtga atttgaacaa caacaggctc 900

ttgggtgttg tgaactttta cacgttacag ggtaagcttg ttgctactgt gattcaagaa 960

gcgtacatgt ttttgcacca ggcgatcata gacaaatcac aagagatagc tgaaaaaagc 1020

gggcataaga agcaggtgat caccccgaaa ttatag 1056

<210> 86
<211> 1056
<212> DNA
<213> Candida viswanathii

<400> 86
atggacaaat tacaagctga agtatacgac gccgaacctc ctgtggccaa attggaagcc 60

aaaacagcgg ccaaattgat atctgctgat ggggcgaaat tgatctacga aggagtgtac 120

cctgtggagc tagtcagaaa aggtcttagg ggaacctacg ggggagattt catagcacaa 180

ggaatcaatg ttgcctggga gtcgattggc aataagactg actttcaacc ccattcatta 240

catgcatatt ttgtcaaggc agggtccgac cagtctgttt tacgttggga agtgctcaag 300

gtgagtgata gcagaagctt tgccaaccgt ctaatgttag catatcagac tcacaccaac 360

caattggtct tcacgatgca gatatcgttc actaaggata acaacgaaga aattaaacgt 420

gctgagtaca agcagctatt ggagagtggc ggtaaaataa gaagtatccc gtttgcaatc 480

aagaaaccac ccaatgagaa gtacttcaaa ttgaaagaca aagtggacga ccttccatat 540

tttgaacata ccaatgggaa tatggcaact gcaataccac ctgatttcct tgaatacgcc 600

acagagatga atcacgacac ggtggggaac aaagagtttg gtattttcat gaaagtatta 660

gatgattact ccttggataa aaactatgag agacagtcgt tcttggggtt ggccttcttg 720

tctgacgccg tgtggttgtc gagtttcacg ccagccttag ggttgccatt gggcactttt 780

gaaaggaaat tctttcgtgt tagtttagac cacaccatgt atttccatga cgccaacttt 840

gattctagtg agtggatctt tgtggatttc aggtttgtga atttgaacaa caacaggctc 900

ttgggtgtgg tgaactttta cactttacag ggtaagcttg ttgctactgt gattcaagaa 960

gcgtacatgt ttttgcacca ggggatcata gaccaatcac aagagatagc tgaaaagagc 1020

gggcataaga agcaggtgat caccccgaaa ttatag 1056

<210> 87
<211> 1056
<212> DNA
<213> Candida viswanathii

<400> 87
atggagagat tacaggccga agtctacgac gctaaaccaa cagttgctaa attggaaacc 60

aagactaccg tcaaattagt tgctcaggac ggaaagagat tggtttatga agctatttat 120

ccagttgaac atatcaaaga agggattcca ggagcttatg gaggtgatac cttggcccaa 180

gggatgaatg ctgcgtggga ctcccttggc gacaagaaag atttccagcc tcattcagtt 240

cattcgtact ttgttaagcc tgcaacaaac aaatcaattt tgagatggga agtgatcaag 300

gtcagtgatg gcagaagttt tgctaatcgt ttcgtgtctg gctatcaaac ccacaacaat 360

gcgttggtct tcactatgca gatctccttt accaagtaca acgatgaagt ggtcaagatt 420

gctgaataca acaagttgtt ggaaagtggt ggcaaaatca gaagtattcc atttgtcatc 480

aagaaagctc cgaatgagaa gtacttcaaa tttaaggaca atcttggaga tctcagatac 540

attgagcaca ccaatggcaa cattgccaca gccatgctgg aagatttgtt tgaatacgcc 600

actgagatga accatgatac tctcggaaat caagagtttg gtattttcat gaaggtcttg 660

gataattact cgttaggtag tgactacact aaacaatctt acttggggtt ggcattctta 720

tcggatgcta tttggatgtc tgtttgttca agagccttgg ggttgccatt tggtagctac 780

catagacaat tttttggagt tagcatggac cactcgatgt atttccacga cgctaatttc 840

gactccaccg agtggatctt tttagacttt aggtttgtca atttgaaaaa tgacaggctc 900

ttgggcgttg ctaattttta cactttagag ggtaagctta tttccacttt gattcaagaa 960

gcttacctct ttttacaccc aggaatcatt gccaagtccc aagagatagc tgtcaagagt 1020

ggaaacaaga ggcaagttaa aatgccaaag ttgtag 1056

<210> 88
<211> 831
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 88
atgagtcacc caacgccaga agaagtctac ggagtcacca aagttgccga aaacaaatac 60

gtggggaaca gacccttgaa caagcccacg ccgaaaacac gcggagtgta tggcggaaac 120

ttctgtgccc aggccatctt ggtagccatc gagtcagccc cggaagggtt cacaccacac 180

tccatccaca gcaacttcat tcgtggcggc gaccccgaag tccctgttga gtgggaggtc 240

gaggtcatct cgaacggcaa gtctttcgcc aaccgcatag tcaagggcgt gcagcacggc 300

atcgtggtct acgtcgccac ggtctcgttg accaacaaga actcaaccac ccgcaacgag 360

agcttcacgt acgacacgcc cccggacgag accgtcaaga cgtacggcaa cgccgagcta 420

gacacatact accagggctg gctctacttg gaggtcaaga actaccccaa gcagttgcac 480

tcgcaccaga tctcgtactc cgtcaagtgg gggccggaga acgacgcgtg gaaggacgcc 540

agccagacgt accagtttgt tgggctcgct gcgatctccg acgtgcttga cctcggccag 600

atcttgcgca acttggacat ccacttgagc acgcccaagt tcaacgtgtc gctcgaccac 660

agtgtgtact tccacggcgc ggactttgac gtgaccaggt ggtcgacgac cacgatccgc 720

atgacgaagc tcgcgcacgg ccgcgcgttg atagagggcg agatgtactc ggaccagggc 780

aggcacatcg ccagcatagt gcaagagagg ttgtacatag ctgagtcgta a 831

<210> 89
<211> 696
<212> DNA
<213> Candida viswanathii

<400> 89

atggtcgccg attctttagt ctaccaccca accgtctcca aattagtcaa gttcttggac 60

acaaccccaa agagggaaaa ggtcttcaga ttattgtcct acttgtccag attcttgggc 120

tactacgcct acagaaaggg ctactccaag gaaaccatcg ccctttttcgc caacttgaaa 180

ggaaacttca cattcatcag aaaggccatg agattcttga agccaataaa tcacttgcaa 240

ttggcctcca aggcatacga caacaagttg ttggacccag tcttgcagat caccaccatc 300

atcagaaact tggcctacgc cggctacttg accatcgacg gtgtcatatt cttcaagttg 360

ttgggtctca ttgacgccaa gaagttccct aacttggcta catacgcctc cagattctgg 420

ttgatcgggt tgattgccgg tttgatcaac tccttgagaa tcatctactc cttgaaggac 480

tacgagcacc aggagggcga caaggagaag gagaccgacg ctaaggctat ccacactaag 540

ttgtacgccg ctaagagaaa attggtctgg gacttgttgg atacttttat tgctttgaac 600

tccttggaca tcttgcattt caccgagggt gacgtcgggt tcgctggtac tatcacctcc 660

ctcttgggat tggaagactt gtggaaggcc acttaa 696

<210> 90
<211> 2040
<212> DNA
<213> Candida viswanathii

<400> 90
atggctttag acaagttaga tttgtatgtc atcataacat tggtggtcgc tgtggccgcc 60

tattttgcta agaaccagtt ccttgatcag ccccaggaca ccgggttcct caacacggac 120

agcggaagca actccagaga cgtcttgctg acattgaaga agaataataa aaacacgttg 180

ttgttgtttg ggtcccagac cggtacggca gaagattacg ccaacaaatt gtcaagagaa 240

ttgcactcca gatttggctt gaaaaccatg gttgcagatt tcgctgatta cgattgggat 300

aacttcggag atatcaccga agatatcttg gtgtttttca tcgttgccac ctacggtgag 360

ggtgaaccta ccgacaatgc cgacgagttc cacacctggt tgactgaaga agctgacact 420

ttgagtactt tgagatatac cgtgttcggg ttgggtaact ccacctacga gttcttcaat 480

gctattggta gaaagtttga cagattgttg agtgagaaag gtggtgacag atttgctgaa 540

tatgctgaag gtgacgacgg cactggcacc ttggacgaag atttcatggc ctggaaggat 600

aatgtctttg acgccttgaa gaatgacttg aactttgaag aaaaggaatt gaagtacgaa 660

ccaaacgtga aattgactga gagagatgac ttgtctgctg ccgactccca agtttccttg 720

ggtgagccaa acaagaagta catcaactcc gagggcatcg acttgaccaa gggtccattc 780

gaccacaccc acccatactt ggccaggatc accgagacca gagagttgtt cagctccaag 840

gaaagacact gtattcacgt tgaatttgac atttctgaat cgaacttgaa atacaccacc 900

ggtgaccatc tagccatctg gccatccaac tccgacgaaa acatcaagca atttgccaag 960

tgtttcggat tggaagataa actcgacact gttattgaat tgaaggcatt ggactccact 1020

tacaccattc cattcccaac tccaattact tacggtgctg tcattagaca ccatttagaa 1080

atctccggtc cagtctcgag acaattcttt ttgtcgattg ctgggtttgc tcctgatgaa 1140

gaaacaaaga agactttcac cagacttggt ggtgacaaac aagaattcgc caccaaggtt 1200

acccgcagaa agttcaacat tgccgatgcc ttgttatatt cctccaacaa cactccatgg 1260

tccgatgttc cttttgagtt ccttattgaa aacatccaac acttgactcc acgttactac 1320

tccatttctt cttcgtcgtt gagtgaaaaa caactcatca atgttactgc agtcgttgag 1380

gccgaagaag aagccgatgg cagaccagtc actggtgttg ttaccaactt gttgaagaac 1440

attgaaattg cgcaaaacaa gactggcgaa aagccacttg ttcactacga tttgagcggc 1500

ccaagaggca agttcaacaa gttcaagttg ccagtgcacg tgagaagatc caactttaag 1560

ttgccaaaga actccaccac cccagttatc ttgattggtc caggtactgg tgttgcccca 1620

ttgagaggtt tcgttagaga aagagttcaa caagtcaaga atggtgtcaa tgttggcaag 1680

actttgttgt tttatggttg cagaaactcc aacgaggact ttttgtacaa gcaagaatgg 1740

gccgagtacg cttctgtttt gggtgaaaac tttgagatgt tcaatgcctt ctctagacaa 1800

gacccatcca agaaggttta cgtccaggat aagattttag aaaacagcca acttgtgcac 1860

gaattgttga ccgaaggtgc cattatctac gtctgtggtg acgccagtag aatggccaga 1920

gacgtccaga ccacgatctc caagattgtt gccaaaagca gagaaatcag tgaagacaag 1980

gccgctgaat tggtcaagtc ctggaaagtc caaaatagat accaagaaga tgtttggtag 2040

<210> 91
<211> 2091
<212> DNA
<213> Candida viswanathii

<400> 91
atgggtgccc ctttaacagt cgccgttggc gaagcaaaac caggcgaaac cgctccaaga 60

agaaaagccg ctcaaaaaat ggcctctgtc gaacgcccaa cagactcaaa ggcaaccact 120

ttgccagact tcattgaaga gtgttttgcc agaaacggca ccagagatgc catggcctgg 180

agagacttgg tcgaaatcca cgtcgaaacc aaacaggtta ccaaaatcat tgacggcgaa 240

cagaaaaagg tcgataagga ctggatctac tacgaaatgg gtccttacaa ctacatatcc 300

taccccaagt tgttgacgtt ggtcaagaac tactccaagg gtttgttgga gttgggcttg 360

gccccagatc aagaatccaa gttgatgatc tttgccagta cctcccacaa gtggatgcag 420

accttcttag cctccagttt ccaaggtatc cccgttgtca ccgcctacga caccttgggt 480

gagtcgggct tgacccactc cttggtgcaa accgaatccg atgccgtgtt caccgacaac 540

caattgttgt cctccttgat tcgtcctttg gagaaggcca cctccgtcaa gtatgtcatc 600

cacggggaaa agattgaccc taacgacaag agacagggcg gcaaaatcta ccaggatgcg 660

gaaaaggcca aggagaagat tttacaaatt agaccagata ttaaatttat ttctttcgac 720

gaggttgttg cattgggtga acaatcgtcc aaagaattgc atttcccaaa accagaagac 780

ccaatctgta tcatgtacac ctcgggttcc accggtgctc caaagggtgt ggttatcacc 840

aatgccaaca ttgttgccgc cgtgggtggt atctccacca atgctactag agacttggtt 900

agaactgtcg acagagtgat tgcatttttg ccattggccc acattttcga gttggccttt 960

gagttggtta ccttctggtg gggggctcca ttgggttacg ccaatgtcaa gactttgacc 1020

gaagcctcct gcagaaactg tcagccagac ttgattgaat tcaaaccaac catcatggtt 1080

ggtgttgctg ccgtttggga atcggtcaga aagggtgtct tgtctaaatt gaaacaggct 1140

tctccaatcc aacaaaagat cttctgggct gcattcaatg ccaagtctac tttgaaccgt 1200

tatggcttgc caggcggtgg gttgtttgac gctgtcttca agaaggttaa agccgccact 1260

ggtggccaat tgcgttatgt gttgaatggt gggtccccaa tctctgttga tgcccaagtg 1320

tttatctcca ccttgcttgc gccaatgttg ttgggttacg gtttgactga aacctgtgcc 1380

aataccacca ttgtcgaaca cacgcgcttc cagattggta ctttgggtac cttggttgga 1440

tctgtcactg ccaagttggt tgatgttgct gatgctggat actacgccaa gaacaaccag 1500

ggtgaaatct ggttgaaagg cggtccagtt gtcaaggaat actacaagaa cgaagaagaa 1560

accaaggctg cattcaccga agatggctgg ttcaagactg gtgatattgg tgaatggacc 1620

gccgacggtg gtttgaacat cattgaccgt aagaagaact tggtcaagac tttgaatggt 1680

gaatacattg ctttggagaa attggaaagt atttacagat ccaaccactt gattttgaac 1740

ttgtgtgttt acgctgacca aaccaaggtc aagccaattg ctattgtctt gccaattgaa 1800

gccaacttga agtctatgtt gaaggacgaa aagattatcc cagatgctga ttcacaagaa 1860

ttgagcagct tggttcacaa caagaaggtt gcccaagctg tcttgagaca cttgctccaa 1920

accggtaaac aacaaggttt gaaaggtatt gaattgttgc agaatgttgt cttgttggat 1980

gacgagtgga ccccacagaa tggttttgtt acttctgccc aaaagttgca gagaaagaag 2040

attttagaaa gttgtaaaaa agaagttgaa gaggcataca agtcgtctta g 2091

<210> 92
<211> 1989
<212> DNA
<213> Candida viswanathii

<400> 92
atgcctaccg aacttcaaaa agaaagagaa ctcaccaagt tcaacccaaa ggagttgaac 60

tacttcttgg aaggttccca agaaagatcc gagatcatca gcaacatggt cgaacaaatg 120

caaaaagacc ctatcttgaa ggtcgacgct tcatactaca acttgaccaa agaccaacaa 180

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 214 of 3077

agagaagtca ccgccaagaa gattgccaga ctctccagat actttgagca cgagtaccca 240

gaccaacagg cccagagatt gtcgatcctc ggtgtctttg acccacaagt cttcaccaga 300

atcggtgtca acttgggttt gtttgtttcc tgtgtccgtg gtaacggtac caactcccag 360

ttcttctact ggaccataaa taagggtatc gacaagttga gaggtatcta tggttgtttt 420

ggtatgactg agttggccca cggttccaac gtccaaggta ttgaaaccac cgccactttt 480

gacgaagaca ctgacgagtt tgtcatcaac accccacaca ttggtgccac caagtggtgg 540

atcggtggtg ctgcgcactc cgccacccac tgctccgtct acgccagatt gaaggtcaaa 600

ggaaaggact acggtgtcaa gacctttgtt gtcccattga gagactccaa ccacgacctc 660

gagccaggtg tgactgttgg tgacattggt gccaagatgg gtagagacgg tatcgataac 720

ggttggatcc agttctccaa cgtcagaatc ccaagattct ttatgttgca aaagtactgt 780

aaggtttccc gtctgggtga agtcaccatg ccaccatctg aacaattgtc ttactcggct 840

ttgattggtg gtagagtcac catgatgatg gactcctaca gaatgaccag tagattcatc 900

accattgcct tgagatacgc catccacaga agacaattca agaagaagga caccgatacc 960

attgaaacca agttgattga ctacccattg catcaaaaga gattgttccc attcttggct 1020

gccgcttact tgttctccca aggtgccttg tacttagaac aaaccatgaa cgcaaccaac 1080

gacaagttgg acgaagctgt cagtgctggt gaaaaggaag ccattgacgc tgccattgtc 1140

gaatccaaga aattgttcgt cgcttccggt tgtttgaagt ccacctgtac ctggttgact 1200

gctgaagcca ttgacgaagc tcgtcaagct tgtggtggtc acggttactc gtcttacaac 1260

ggtttcggta aagcctactc cgactgggtt gtccaatgta cctgggaagg tgacaacaac 1320

atcttggcca tgaacgttgc caagccaatg gttagagact tgttgaagga gccagaacaa 1380

aagggattgg ttctctccag cgttgccgac ttggacgacc cagccaagtt ggttaaggct 1440

ttcgaccacg ccctttccgg cttggccaga gacattggtg ctgttgctga agacaagggt 1500

ttcgacatta ccggtccaag tttggttttg gtttccaagt tgaacgctca cagattcttg 1560

attgacggtt tcttcaagcg tatcacccca gaatggtctg aagtcttgag acctttgggt 1620

ttcttgtatg ccgactggat cttgaccaac tttggtgcca ccttcttgca gtacggtatc 1680

attaccccag atgtcagcag aaagatttcc tccgagcact tcccagcctt gtgtgccaag 1740

gttagaccaa acgttgttgg tttgactgat ggtttcaact tgactgacat gatgaccaat 1800

gctgctattg gtagatatga tggtaacgtc tacgaacact acttcgaaac tgtcaaggct 1860

ttgaacccac cagaaaacac caaggctcca tactccaagg ctttggaaga catgttgaac 1920

cgtccagacc ttgaagtcag agaaagaggt gaaaagtccg aagaagctgc tgaaatcttg 1980

tccagttaa 1989

<210> 93
<211> 2130
<212> DNA
<213> Candida viswanathii

<400> 93
atgactttta caaagaaaaa cgttagtgta tcacaaggtc ctgaccctag atcatccatc 60

caaaaggaaa gagacagctc caaatggaac cctcaacaaa tgaactactt cttggaaggc 120

tccgtcgaaa gaagtgagtt gatgaaggct ttggcccaac aaatggaaag agacccaatc 180

ttgttcacag acggctccta ctacgacttg accaaggacc aacaaagaga attgaccgcc 240

gtcaagatca acagaatcgc cagatacaga gaacaagaat ccatcgacac tttcaacaag 300

agattgtcct tgattggtat ctttgaccca caggtcggta ccagaattgg tgtcaacctc 360

ggtttgttcc tttcttgtat cagaggtaac ggtaccactt cccaattgaa ctactgggct 420

aacgaaaagg aaaccgctga cgttaaaggt atctacggtt gtttcggtat gaccgaattg 480

gcccacggtt ccaacgttgc tggtttggaa accaccgcca catttgacaa ggaatctgac 540

gagtttgtca tcaacacccc acacattggt gccaccaagt ggtggattgg tggtgctgct 600

cactccgcca cccactgttc tgtctacgcc agattgattg ttgacggtca agattacggt 660

gtcaagactt ttgttgtccc attgagagac tccaaccacg acctcatgcc aggtgtcact 720

gttggtgaca ttggtgccaa gatgggtaga gatggtatcg ataacggttg gatccaattc 780

tccaacgtca gaatcccaag attctttatg ttgcaaaagt tctgtaaggt ttctgctgaa 840

ggtgaagtca ccttgccacc tttggaacaa ttgtcttact ccgccttgtt gggtggtaga 900

gtcatgatgg ttttggactc ctacagaatg ttggctagaa tgtccaccat tgccttgaga 960

tacgccattg gtagaagaca attcaagggt gacaatgtcg atccaaaaga tccaaacgct 1020

ttggaaaccc aattgataga ttacccattg caccaaaaga gattgttccc atacttggct 1080

gctgcctacg tcatctccgc tggtgccctc aaggttgaag acaccatcca taacaccttg 1140

gctgaattgg acgctgccgt tgaaaagaac gacaccaagg ctatctttaa gtctattgac 1200

gacatgaagt cattgtttgt tgactctggt tccttgaagt ccactgccac ttggttgggt 1260

gctgaagcca ttgaccaatg tagacaagcc tgtggtggtc acggttactc gtcctacaac 1320

ggcttcggta aagcctacaa cgattgggtt gtccaatgta cttgggaagg tgacaacaat 1380

gtcttggcca tgagtgttgg taagccaatt gtcaagcaag ttatcagcat tgaagatgcc 1440

ggcaagaccg tcagaggttc caccgctttc ttgaaccaat tgaaggacta cactggttcc 1500

aacagctcca aggttgtttt gaacactgtt gctgacttgg acgacatcaa gactgtcatc 1560

aaggctattg aagttgccat catcagattg tcccaagaag ctgcttctat tgtcaagaag 1620

gaatctttcg actatgtcgg cgctgaattg gttcaactct ccaagttgaa ggctcaccac 1680

tacttgttga ctgaatacat cagaagaatt gacacctttg accaaaagga cttggttcca 1740

tacttgatca ccctcggtaa gttgtacgct gccactattg tcttggacag atttgccggt 1800

gtcttcttga ctttcaacgt tgcctccacc gaagccatca ctgctttggc ctctgtgcaa 1860

attccaaagt tgtgtgctga agtcagacca aacgttgttg cttacaccga ctccttccaa 1920

caatccgaca tgattgtcaa ttctgctatt ggtagatacg atggtgacat ctatgagaac 1980

tactttgact tggtcaagtt gcagaaccca ccatccaaga ccaaggctcc ttactctgat 2040

gctttggaag ccatgttgaa cagaccaacc ttggacgaaa gagaaagatt tgaaaagtct 2100

gatgaaaccg ctgctatctt gtccaagtaa 2130

<210> 94
<211> 2208
<212> DNA
<213> Candida viswanathii

<400> 94
atggtcaaca tatcgaaatt gacgggttat aacaagcagg acatcaggaa tgtggtgcta 60

ttgctacagg agtttgtcaa gacctacaaa gacaacaaga tcaaactcaa ctacctgagt 120

agacctgtca tcttgttctt gagtaccttg gttgcaactg ccggtattgg ggtgtttttc 180

accttgagaa gcatcgtcac taagtacaac gagtacctac tcaacaagag attgagacgc 240

ccaagtctta tcagacaatc ctccaatatc ttgaagaacg gatcccgtga gatctttatc 300

cagaagggca acggcaaagt aacaagaatc atcatcccaa aagcaaacaa cgaccagtat 360

gccgccgaca agtatttgta taaagatttt gcccgcaacg agcaaatatt gcaacagcaa 420

aagggaaggc tcttcaattc cagattcttg aaccagttga ccattatctg gaagatcttg 480

attccaaagt tctactgcca aaacacttcc ttgttgttat cgcagtgctt ctttttgatt 540

ttcagaacat ggttgtcctt gttgattgcc aagctagatg gtcagattgt caagaacttg 600

attgctgcag acggtaggaa gtttgcccgt gacttgattt acttttttgtt gattgccttc 660

cctgcttcgt acaccaacgc cgctatcaaa tacttggaat tgagattggc gttaggattc 720

agaactaatc ttaccagata catccatgac atgtacttgg acaaaaccat gtcgtactac 780

aaagtgggat tgaacggcgc cgatatccaa aacatagacc agtacatcac cgaagatgtc 840

accaaattct gtatgtcgtt gtgttcgttg ttttcctcca tgggtaagcc attcattgac 900

ttgatctttt tcagtgttta tttgagagac aatttgggta ctggtgccat tattggcatt 960

tttgccaact attttgctac cgccatcatg ttgaaaaagg caacaccaag attcggtaag 1020

ttggctgcca aaagaaccca cttggaaggt gtttatttca accaacagtt gaacataatg 1080

accaacagtg aagagattgg gttctacaaa ggatcgaaga ttgagaagtc caagcttgcg 1140

gagaactttg acaagttgat gggtcacgta tcgagagaaa tcaatttatc gtccagctat 1200

gccgctctag aagactacgt gcttaaatac acgtggctgg cctggggtta catcttttct 1260

ggtctacctg tgtttttgga tgtgcttttc cctaaagaag acccaagtag tggccatatt 1320

gctgatatag atgacgatga ccatgcccat ggacacgggc acaccgggga agagaccagc 1380

tcaacaactg aaaacatgaa gaccttcgtc accaacaagc gattattgtt gagtcttgcc 1440

gatgctggtt ccagattgat ggttagtttg aaagaggtca ccacgttgac aggtataacg 1500

aatagagtct tcaacatgtt gactcagctc caccgtgtcc atgatcctaa atttgactac 1560

ggtgacaagt atggtttgcc tgatattcac gggacttatc aattgaacta cgatggtttg 1620

agattggaac atgttccaat tactgtgcca actgccgagg gttcttactc cacaccattg 1680

atcccagacc tcacttttga catcaagggc aagaatttgt tatttgttgg tccaaacggt 1740

tcgggcaaaa cttctgttgc cagggttctt gcaggtcttt ggcccttgta tgccgggtta 1800

gtgctgaaac cactggattt gttctttaac ccacaaaaga gttatttcac caccggaagt 1860

ttgcgtgacc aagttgttta ccctaataga tccgaaaaca ccaccaacga tcaaattttc 1920

cacatcttac actgtgtaca cttagaccat attgttaaac ggtacggatt gaaccagaac 1980

ttggatttcg ctaaaacatt gagtggaggt gagaagcaaa gattgagttt cgccagagtg 2040

ttgtttaaca gaccaagtat tgtcattctt gatgattcga cgtcggcgtt gtccccagat 2100

atggaagagt tgatgtacca ggtgttgcaa gatcacaaga tcaattacgt cacactttca 2160

aatcgtccct ctctcagtaa gttccatgat aaagtatttg aaatataa 2208

<210> 95
<211> 1998
<212> DNA
<213> Candida viswanathii

<400> 95
atgacagtgg agaatgcaaa actacagaag aactcgttgg cggttctgct cttgaaggtg 60

tacaaatcca acagatcatt attgttaaac acctcataca tcatattaat cattgctgcc 120

ttcactggcg caacgaatac cgggcgaggc acctcctcca gatcatcggc aaaagtagag 180

accgatgaag aacaatcggt taaaaagaaa caccccaagc tctctagaga gtccttccat 240

agactaagaa aagcaatctt gccaactttc tttgatagaa ctatagttta ctttttcgcc 300

aacttgactt tgttggtggt cagagcatta ttgacactta gagttgctac ccttgacggt 360

cagcttgtgg gggcattggt ttcaagaaga ataagggtgt ttgccaagta cttgttgtac 420

tggatgcttc ttggtatccc cgctgctttg acaaatgcct tgttgaactg gaccaaactg 480

aacttgagca agagcattag aatgaacttg aataataaca tcatggagga atacttgcca 540

gataacttgg acccaaacta ttattcattg atccatttga ctgataacaa gattagagac 600

ccaaatcaga gaataaccac tgatactagt cgtttgagcg atgccttggc aagcttgccc 660

ggtcacatat tgaagccaac gttggatatc atattgtgtg cgcaacagtt aagcaagagc 720

ggtgttggta atggggaagg tacgttggca ttaggtatat tggcacactt ctcaaccatg 780

atcatccgtt tcttctcccc gccatttgcc aagttggcgg ctgagagagc taaccttgaa 840

ggtcagttgc gttccgcgca ttccaagatt gttgccaaca gtgaagaaat tgctttcttg 900

ggtggtcatg accgtgagtt ggatcacatc gaccactgct actatacttt ggagagattc 960

tcgaaaggcg aatattggaa gcgagccata cacgaaatca cacaaacgtt tattgtgaag 1020

tacttttggg gtgttgcagg tttagtgttg tgttctgcac cggttttcat tgccaaatac 1080

ttgggtgagc cggaagataa gaatgttgct ggtaatttca tcaccaacag aagattgttg 1140

atgagtgcct cggattcctt ggatcgttta atctattcta gaagatactt gttgcaagtt 1200

gtcggtcatg ctaccagagt gtctgacttc ttggacactt tacatgaagt ggaggagaag 1260

aagaagagaa tcacatcgaa tgtgcagttt aacaacgacg agattacttt cgatcatgtt 1320

agattgatga ctccaacgga agtgaccttg atcccagact tgaacttttc cattaaacca 1380

ggtgaccatt tgttgattgt ggggccaaac ggttcaggta agtcgtcgtt gttcagaatg 1440

ttgggtgggt tgtggcccgt taggtttggt actattagaa ttccaaacac agagaacatg 1500

ttctacttgc cgcaaaaggc ttaccttgtt gaaggatcat tcagagagca aatcatttat 1560

ccacacaacg tgactcaaca gaagaagact gatcaacaat tgaaagagat cttgaaggtt 1620

ttgaaattgg aagattactc agggcaattg gatgaggtta agaaatggag cgaagaattg 1680

tccattggtg ctcaacaaag attggctatg gctagattgt actaccacga acctaagttt 1740

gctgtcttgg acgaatgtac ttcagctgtg tcaccagaca tggaacaact catgtaccaa 1800

cacgcacaag gtttgggtat cacgcttttg tccgttgccc atagacctgc attgtggcac 1860

ttccacaaat acttgttgga attcgacggg aagggtagtt actactttgg tacgttggat 1920

gaaaagcaca aaatgaagtt agaagaagaa gaacgactca agaaggagaa tgaaaagaag 1980

agtgtcgcca agaagtag 1998

<210> 96
<211> 1953
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 96
atgtccgcca acgaaaacat ctccagattc gacgcccctg tcggtaagga acacccagcc 60

tacgaattat tccataacca cacaagatcc ttcgtttatg gattacaacc tagagcctgt 120

caaggaatgt tggacttcga cttcatctgc aagagagaaa acccatctgt cgccggagtt 180

atctatccat tcggtggtca attcgttacc aagatgtact ggggtaccaa ggaaactctt 240

ttacctgttt accaacaagt tgaaaaggcc gctgccaagc acccagaagt tgatgttgtc 300

gttaactttg cctcttccag atctgtttac tcttccacca tggaattgtt agaataccca 360

caattcagaa ccatcgccat tattgccgaa ggagttccag aaagaagggc cagagaaatc 420

ttacacaagg cccaaaagaa gggagtcacc atcattggac cagctaccgt tggcggaatc 480

aagccaggat gtttcaaggt gggcaacacc ggcggaatga tggacaacat tgttgcctct 540

aagttataca gaccaggttc tgtggcctac gtttctaagt ctggcggcat gtctaacgaa 600

ttgaacaaca ttatctccca caccaccgac ggagtttacg aaggaattgc tattggagga 660

gacagatacc ctggaactac cttcattgac catatcttga gatacgaagc cgacccaaag 720

tgcaagatca tcgttttact tggagaagtg ggaggagtgg aagaatacag agttatcgaa 780

gctgtgaaga acggtcaaat caagaagcca atcgttgctt gggccattgg aacttgcgcc 840

tctatgttca agactgaagt gcaattcggt cacgccggtt ctatggccaa ctctgacttg 900

gaaactgcca aggctaagaa cgccgccatg aagtccgctg gtttctacgt tccagatacc 960

ttcgaagaca tgccagaagt tcttgccgaa ttatacgaaa agatggttgc caagggtgaa 1020

ttgtccagaa tctccgaacc tgaagttcca aagatcccaa ttgactactc ctgggcccaa 1080

gaacttggac ttatcagaaa gccagctgct ttcatctcta ctatttctga tgacagaggt 1140

caagaacttt tgtacgctgg tatgccaatt tctgaagtgt tcaaggaaga cattggaatc 1200

ggtggagtta tgtccttgtt gtggttcaga agaagattac cagactacgc ctctaagttt 1260

cttgaaatgg tgttaatgct tactgctgac cacggaccag ccgtatctgg agccatgaac 1320

accattatca ccaccagagc tggaaaggat ttaatttcct ctttggtggc tggattattg 1380

accattggaa ccagattcgg cggagctctt gacggagctg ccaccgaatt caccactgcc 1440

tacgacaagg gattgtctcc aagacaattc gtggatacca tgagaaagca aaacaagttg 1500

attcctggaa ttggtcatag agttaagtcc agaaacaacc cagatttcag agttgaactt 1560

gttaaggact ttgtgaagaa gaacttccca tctacccaat tgttagacta cgcccttgct 1620

gttgaagaag ttaccacctc taagaaggac aacttgattt tgaacgtgga cggagctatt 1680

gctgtgtcct ttgttgattt aatgagatcc tgtggagcct ttactgtcga agaaactgaa 1740

gactacttaa agaacggagt gttaaacgga ttgttcgtgt taggaagatc tattggatta 1800

attgcccacc atttagatca aaagagatta aagaccggat tgtacagaca tccttgggac 1860

gatatcacct acttggtggg tcaagaagct atccaaaaga agagagttga aatcagtgcc 1920

ggtgacgtgt ctaaggccaa gactagatca tag 1953

<210> 97
<211> 1494
<212> DNA

<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 97
atgtcagcga aatctattca cgaagccgac ggtaaggcct tgttagcaca ctttttgtct 60

aaggcgccag tctgggccga acaacaacca atcaacacgt ttgagatggg tacaccaaag 120

ttggcgtcct tgacgttcga agacggtgtc gccccagaac aaatcttcgc cgccgctgag 180

aagacctacc catggttgtt ggaatctggt gccaagtttg tcgccaagcc agaccaatta 240

atcaagagaa gaggtaaggc cggtttgttg gtattaaaca agtcgtggga agaatgtaag 300

ccatggatcg ccgaacgtgc cgccaagcca atcaacgtcg aaggtattga cggcgtcttg 360

agaacgttct tggttgaacc atttgtccca cacgaccaaa agcacgaata ctacatcaac 420

atccactctg tcagagaagg tgactggatc ttattctacc acgaaggcgg cgttgacgtt 480

ggtgacgtcg acgccaaggc cgccaagatc ttaatcccag tggacattga aaacgaatac 540

ccatctaacg ccacgttaac caaggaattg ttggcacacg tcccagaaga ccaacaccaa 600

accttgttag acttcatcaa ccgtttatac gccgtttacg ttgatttgca atttacgtat 660

ttggaaatca acccattggt tgtcatccca accgcccaag gtgttgaagt tcactacttg 720

gatcttgccg gtaagttaga ccaaaccgca gaatttgaat gtggtccaaa gtgggctgct 780

gcgcgttctc cagccgcttt gggtcaagtt gttaccattg acgccggttc taccaaggtc 840

tctatcgacg ccggtccagc catggttttc ccagctcctt tcggaagaga attgtctaag 900

gaagaagcgt acattgcgga attagattct aagaccggcg cttccttgaa gttgactgtg 960

ttaaatgcca agggtagaat ctggacccct gtcgctggag gcggcgcctc tgttgtttac 1020

gccgacgcca ttgcgtccgc cggttttgct gacgaattag ccaactacgg tgaatactcc 1080

ggtgctccaa acgaaaccca aacctacgaa tacgccaaaa ccgtattgga tttaatgacc 1140

cgtggtgacg ctcacccaga aggtaaggta ttgttcattg gtggcggcat cgccaacttc 1200

acccaagtgg gctctacctt caagggtatc atccgtgcct tccgtgacta ccaatcctcc 1260

ttgcacaacc acaaggtcaa gatttacgtc agaagaggtg gaccaaactg gcaagaagga 1320

ttgcgtctca tcaagtcggc tggtgacgaa ttgaatttgc caatggaaat ttacggtcca 1380

gacatgcacg tctcgggaat tgtgcctctc gctttgcttg gcaagcgtcc aaagaatgtt 1440

aagccttttg gtaccggccc ttccactgaa gcttctactc ctttaggcgt gtaa 1494

<210> 98
<211> 1953
<212> DNA
<213> Candida viswanathii

<400> 98

atgtcaggat tagaaatagc cgctgctgcc atccttggta gtcagttatt ggaagccaaa 60

tatttaattg ccgacgacgt gctgttagcc aagacagtcg ctgtcaatgc cctcccatac 120

ttgtggaaag ccagcagagg taaggcatca tactggtact ttttcgagca gtccgtgttc 180

aagaacccaa acaacaaagc gttggcgttc ccaagaccaa gaaagaatgc ccccaccccc 240

aagaccgacg ccgagggatt ccagatctac gacgatcagt ttgacctaga agaatacacc 300

tacaaggaat tgtacgacat ggttttgaag tactcataca tcttgaagaa cgagtacggc 360

gtcactgcca acgacaccat cggtgtttct tgtatgaaca agccgctttt cattgtcttg 420

tggttggcat tgtggaacat tggtgccttg cctgcgttct tgaacttcaa caccaaggac 480

aagccattga tccactgtct taagattgtc aacgcttcgc aagttttcgt tgacccggac 540

tgtgattccc caatcagaga taccgaggct cagatcagag aggaattgcc acatgtgcaa 600

ataaactaca ttgacgagtt tgccttgttt gacagattga gactcaagtc gactccaaaa 660

cacagagccg aggacaagac cagaagacca accgatactg actcctccgc ttgtgcattg 720

atttacacct cgggtaccac cggtttgcca aaagccggta tcatgtcctg gagaaaagcc 780

ttcatggcct cggttttctt tggccacatc atgaagattg actcgaaatc gaacgtcttg 840

accgccatgc ccttgtacca ctccaccgcg gccatgttgg ggttgtgtcc tactttgatt 900

gtcggtggct gtgtctccgt gtcccagaaa ttctccgcta cttcgttctg gacccaggcc 960

agattatgtg gtgccaccca cgtgcaatac gtcggtgagg tctgtcgtta cttgttgaac 1020

tccaagcctc atccagacca agacagacac aatgtcagaa ttgcctacgg taacgggttg 1080

cgtccagata tatggtctga gttcaagcgc agattccaca ttgaaggtat cggtgagttc 1140

tacgccgcca ccgagtcccc tatcgccacc accaacttgc agtacggtga gtacggtgtc 1200

ggcgcctgtc gtaagtacgg gtccctcatc agcttgttat tgtctaccca gcagaaattg 1260

gccaagatgg acccagaaga cgagagtgaa atctacaagg accccaagac cgggttctgt 1320

accgaggccg cttacaacga gccaggtgag ttgttgatga gaatcttgaa ccctaacgac 1380

gtgcagaaat ccttccaggg ttattatggt aacaagtccg ccaccaacag caaaatcctc 1440

accaatgttt tcaaaaaagg tgacgcgtgg tacagatccg gtgacttgtt gaagatggac 1500

gaggacaaat tgttgtactt tgtcgacaga ttaggtgaca ctttccgttg gaagtccgaa 1560

aacgtctccg ccaccgaggt cgagaacgaa ttgatgggct ccaaggcctt gaagcagtcc 1620

gtcgttgtcg gtgtcaaggt gccaaaccac gaaggtagag cctgttttgc cgtctgtgaa 1680

gccaaggacg agttgagcca tgaagaaatc ttgaaattga ttcactctca cgtgaccaag 1740

tctttgcctg tgtatgctca acctgcgttc atcaagattg gcaccattga ggcttcgcac 1800

aaccacaagg ttcctaagaa ccaattcaag aaccaaaagt tgccaaaggg tgaagacggc 1860

aaggatttga tctactggtt gaatggcgac aagtaccagg agttgactga agacgattgg 1920

tctttgattt gtaccggtaa agccaaattg tag 1953

<210> 99
<211> 1554
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 99
atgattgaac aactcctaga atattggtat gtcgttgtgc cagtgttgta catcatcaaa 60

caactccttg catacacaaa gactcgcgtc ttgatgaaaa agttgggtgc tgctccagtc 120

acaaacaagt tgtacgacaa cgctttcggt atcgtcaatg gatggaaggc tctccagttc 180

aagaaagagg gcagggctca agagtacaac gattacaagt ttgaccactc caagaaccca 240

agcgtgggca cctacgtcag tattcttttc ggcaccagga tcgtcgtgac caaagatcca 300

gagaatatca aagctatttt ggcaacccag tttggtgatt tttctttggg caagaggcac 360

actcttttta agcctttgtt aggtgatggg atcttcacat tggacggcga aggctggaag 420

cacagcagag ccatgttgag accacagttt gccagagaac aagttgctca tgtgacgtcg 480

ttggaaccac acttccagtt gttgaagaag catattctta agcacaaggg tgaatacttt 540

gatatccagg aattgttctt tagatttacc gttgattcgg ccacggagtt cttatttggt 600

gagtccgtgc actccttaaa ggacgaatct attggtatca accaagacga tatagatttt 660

gctggtagaa aggactttgc tgagtcgttc aacaaagccc aggaatactt ggctattaga 720

accttggtgc agacgttcta ctggttggtc aacaacaagg agtttagaga ctgtaccaag 780

ctggtgcaca agttcaccaa ctactatgtt cagaaagctt tggatgctag cccagaagag 840

cttgaaaagc aaagtgggta tgtgttcttg tacgagcttg tcaagcagac aagagacccc 900

aatgtgttgc gtgaccagtc tttgaacatc ttgttggccg gaagagacac cactgctggg 960

ttgttgtcgt ttgctgtctt tgagttggcc agacacccag agatctgggc caagttgaga 1020

gaggaaattg aacaacagtt tggtcttgga gaagactctc gtgttgaaga gattaccttt 1080

gagagcttga agagatgtga gtacttgaaa gcgttcctta atgaaacctt gcgtatttac 1140

ccaagtgtcc caagaaactt cagaatcgcc accaagaaca cgacattgcc aaggggcggt 1200

ggttcagacg gtacctcgcc aatcttgatc caaaagggag aagctgtgtc gtatggtatc 1260

aactctactc atttggaccc tgtctattac ggccctgatg ctgctgagtt cagaccagag 1320

agatggtttg agccatcaac caaaaagctc ggctgggctt acttgccatt caacggtggt 1380

ccaagaatct gtttgggtca gcagtttgcc ttgacggaag ctggctatgt gttggttaga 1440

ttggtgcaag agttctccca cgttaggctg gacccagacg aggtgtaccc gccaaagagg 1500

ttgaccaact tgaccatgtg tttgcaggat ggtgctattg tcaagtttga ctag 1554

<210> 100
<211> 1050
<212> DNA
<213> Candida viswanathii

<400> 100
atgtctgcta atatcccaaa aactcaaaaa gctgtcgtct tcgagaagaa cggtggtgaa 60

ttaaaataca aagacatccc agtgccaacc ccaaaggcca acgaattgct catcaacgtc 120

aagtactcgg gtgtctgtca cactgatttg cacgcctgga agggtgactg gccattggac 180

accaaattgc cattggttgg tggtcacgaa ggtgctggtg ttgttgtcgg catgggtgaa 240

aacgtcaagg gctggaaaat cggtgatttc gccggtatca aatggttgaa cggttcttgt 300

atgtcctgtg agttctgtca gcaaggtgct gaaccaaact gtggtgaagc tgacttgtct 360

ggttacaccc acgatggttc tttcgaacaa tacgccactg ctgatgctgt gcaagccgcc 420

agaatcccag ctggcactga tttggccgaa gttgccccaa tcttgtgtgc tggtgtcacc 480

gtctacaaag ccttgaagac tgccgacttg gctgctggtc aatgggtcgc tatctccggt 540

gctggtggtg gtttgggctc cttggctgtc caatacgccg tcgccatggg tttgagagtc 600

gttgccattg acggtggtga cgaaaagggt gactttgtca agtccttggg tgctgaagcc 660

tacattgatt tcctcaagga aaagggcatt gttgctgctg tcaagaaggc cactgatggc 720

ggtccacacg gtgctatcaa tgtttccgtt tccgaaaaag ccattgacca atctgtcgag 780

tacgttagac cattgggtaa ggttgttttg gttggtttgc cagctggctc caaggtcact 840

gctggtgttt tcgaagccgt tgtcaagtcc attgaaatca agggttctta cgtcggtaac 900

agaaaggata ctgccgaagc cgttgacttt ttctccagag gcttgatcaa gtgtccaatc 960

aagattgtgg gcttgagtga attgccacag gtcttcaagt tgatggaaga aggtaagatc 1020

ttgggtagat acgtcttgga tacctccaaa 1050

<210> 101
<211> 2115
<212> DNA
<213> Candida viswanathii

<400> 101
atggctccat ttttgcccga ccaggtcgac tacaaacacg tcgacaccct tatgttatta 60

tgtgacggga tcatccacga aaccaccgtc gaccaaatca aagacgttat tgctcctgac 120

ttccctgctg acaagtacga agagtacgtc aggacattca ccaaaccctc cgaaacccca 180

gggttcaggg aaaccgtcta caacacagtc aacgcaaaca ccacggacgc aatccaccag 240

ttcattatct tgaccaatgt tttggcatcc agggtcttgg ctccagcttt gaccaactcg 300

ttgacgccta tcaaggacat gagcttggaa gaccgtgaaa aattgttggc ctcgtggcgc 360

gactccccaa tcgctgccaa aaggaagttg ttcaggttgg tttctacgct taccttggtc 420

acgttcacga gattggccaa tgagttgcat ttgaaagcca ttcattatcc aggaagagaa 480

gaccgtgaaa aggcttatga aacccaggag attgaccctt ttaagtacca gtttttggaa 540

aaaccgaagt tttacggcgc tgagttgtac ttgccagata ttgatgtgat cattattgga 600

tctggtgccg gtgctggtgt tgtggcccac actttggcca acgatggctt caagagtttg 660

gttttggaaa agggcaaata ctttagcaac tccgagttga actttgatga caaggacggc 720

gttcaagaat tataccaaag tggaggtact ttgactacag tcaaccaaca gttgtttgtt 780

cttgctggtt ccacttttgg tggcggtacc actgtcaatt ggtcagcctg tcttaagacg 840

ccattcaagg tgcgtaagga atggtatgat gagtttggtg ttgactttgc tgctgatgaa 900

gcatacgata aagcgcagga ttatgtttgg cagcaaatgg gagcttctac cgaaggcatc 960

acccactctt tggctaacga gattattatt gaaggtggta agaaattagg ttacaaggcc 1020

aaggtattag accaaaacag cggtggtcat cctcagcaca gatgcggttt ctgttatttg 1080

ggctgtaagc acggtatcaa gcagggttct gttaataact ggtttagaga cgcagctgcc 1140

cacggttccc agttcatgca acaggttaga gttttgcaaa tacttaacaa gaaggggatc 1200

gcttacggta tcttgtgtga ggatgttgta accggcgcca agttcaccat tactggcccc 1260

aaaaagtttg ttgttgctgc cggtgctttg aacactccat ctgtgttggt caactccggc 1320

ttcaagaaca agaacatcgg taagaactta actttgcacc cagtttctgt cgtgtttggt 1380

gattttggca aagacgttca agcagaccac ttccacaact ccatcatgac tgccctttgt 1440

tcagaagccg ctgatttaga cggcaagggc catggatgca gaattgaaac catcttgaac 1500

gctccattca tccaggcttc attcttacca tggagaggta gtaacgaggc tagacgagac 1560

ttgttgcgtt acaacaacat ggtggcgatg ttgctcctta gtcgtgacac caccagtggt 1620

tccgtttctg ctcatccaac caaacctgaa gctttggttg tcgagtacga cgtgaacaag 1680

tttgacagaa actcgatctt gcaggcattg ttggtcactg ctgacttgtt gtatatccaa 1740

ggtgccaaga gaatccttag tccacaggca tgggtgccaa tttttgaatc cgacaagcca 1800

aaggataaga gatcaatcaa ggacgaggac tatgtcgaat ggagagccaa ggttgccaag 1860

attcctttcg acacctacgg ctcaccttat ggttcggcac atcaaatgtc ttcttgccgt 1920

atgtcaggta agggtcctaa atacggtgct gttgacaccg atggtagatt gtttgaatgt 1980

tcgaatgttt atgttgccga tgcaagtctt ttgccaactg caagcggtgc caaccctatg 2040

gtcaccacca tgactcttgc cagacatgtt gcgttaggtt tggcagactc cttgaagacc 2100

aaagccaagt tgtag 2115

<210> 102
<211> 6120
<212> DNA
<213> Candida viswanathii

<400> 102
atgtctactc atagaccttt ccaattgacc cacggttcca tcgaacacac cttgttggtg 60

ccaaacgagt tgttcttcaa ctattcacag ttaaaagacg aattcataaa gaccttgcct 120

gaaccaaccg aaggtttcgc tggcgacgat gaaccttcca gtcctgctga attgtacggc 180

aaattcctcg gctacatcag tgacaacacc gttcaattcc cccagatctt acaattgtcc 240

ttgcaagact tccagcagcg attcttggac aaccacgaca acatccactc ctttgccgtc 300

agattattag aagatgaagc ttatccaaca acaatcacca aagtcaagga aaatatcatc 360

aagaactact acaaagccat caagtccatc gacaaggtcg agtcaaactt gttgtaccac 420

tgcaaacatg acgccaagtt ggccgctata ttcggtggtc aaggtaacac cgacgactac 480

tttgaagaat tgcgtgaatt gtacacctta taccagggct tgattgagga cctccttatc 540

tccattgccg acaagttgga cgagttatac ccttcttttg acaagatcta cacccagggt 600

ttgaacatct tgggctggtt gaagcaccca gaaaccaccc ctgaccaaga ttacttgttg 660

tccgtaccag tgagttgtcc tgttatctgt atcatccaat tgtgtcacta caccatcacc 720

tgcaaagttc ttggtttgac ccctggtgaa tttagagact cgttgaagtg gtccaccggt 780

cactcccaag gtttggttac tgctaccgct atttccagtt ccgactcctg ggactccttc 840

aaacaaaact ccattgctgc cgtctccttg atgcttttca ttggtgccag atgtttgatg 900

gcttacccaa gaactacctt gccaccaacc atgttgcaag actccttgga acacggtgaa 960

ggtagaccat ctccaatgtt gtcagttaga gacttgacca tcacccaagt tgagaagttt 1020

attgaacaga ccaactctca cttgccaaag gaaaagcaca ttgccgtcag tttggtcaat 1080

ggtgccagaa atttggttct ttctggtccc ccggagtccc tttacggttt caacttgaac 1140

ttgagaaacc aaaaggctcc aatgggattg gaccaatcac gtgttccatt cagtgaacgt 1200

aagttgaagt gttccaacag attcttgcca atttttgcac cattccactc tcacttgttg 1260

gctgatgcca ctgagtacat tttggatgat gtcaaagaac accgtttgtc tttccagaaa 1320

ttgaagattg cagtctacga tacctacgac ggctccaact tccaagagag caaggaacca 1380

attattgaca gactcgtcaa gttgatcacc gagttgccag ttcactggga aaccgccacc 1440

aaccacaggg ccacccacat tttggatttc ggcccaggtg gtgtctccgg tttgggtgtt 1500

ttgacccaca gaaacaagga aggtactggt gctagaataa ttgttgctgg tgctcttgac 1560

tccaacccaa ttgacgatga gtatggtttc aagcacgaaa tcttccagac ttctgccgac 1620

aagtccatca agtgggctag cgattggttg gaagaattca aaccaacttt ggtcaagact 1680

tcccagggaa agatctacgt caacaccaag ttctcgcaat tgttgggcag agctcccttg 1740

atggtcccag gtatgacacc aaccactgtt aacccagaca tcattgccgc ctctttgaat 1800

gctggctatc acattgaatt agccggtggt ggttatttcg ccggcaggat catgaccaaa 1860

gccattgacc aaattgttgc cgacatcaag ccaggttacg gtttgggtat caacttgatt 1920

tacgtcaacc cattcatgtt gcaatggggt attccattga ttaaggagtt gagagaaaag 1980

ggttatccaa tccaatcttt gaccattggt gctggtgttc catctttgga agttgccact 2040

gaatacattg aagagttggg tctcacgcac ttgggcttga aaccaggttc gattgacgcc 2100

atcagccaag tcatcaccat tgccaaggct catccaaagt tcccaattgt cttgcaatgg 2160

actggtggta gaggtggtgg ccaccactct tttgaagatt tccaccaacc aatcctccag 2220

atgtactcca agatcagaag atgcccaaac attgtcttgg ttgctggttc cgggtttggt 2280

tctgacgaag acacctaccc atacttgact ggttcttggt ccaagagatt caactaccca 2340

ccaatgccat acgatggtgt cttgtttggt tccagagtca tgaccgccaa ggaggcccac 2400

acttcgttgg aagctaagaa attgattgcc tcgtgcccag gtgtcccaga tgagaagtgg 2460

gagcaaacct acaagaagcc aaccggtggt atcatcactg ttagatctga aatgggtgag 2520

ccaatccaca agattgccac cagaggtgtc atgttctgga aggaattgga cgacactatc 2580

ttcaaccttc caaagccaaa agccttggaa gccatcaaga agaagagaga ctacatcatc 2640

aagaagttga acagcgactt ccaaaagcca tggtttggta agaatgcttc tggtgtttgt 2700

gacttgcaag aaatgaccta cgaggaaatc gccaacagat tggttgagtt gatgtacgtc 2760

aagaagtctc aaagatggat tgacgtttcc ttaagaaact tgtacggtga ctttttgaga 2820

agagttgaag aaagattcac ttctgctgcc ggcgtggttt ccttgttgca aaacttcatc 2880

cagttgaacg acccagaaac attcagtgcc gagttcttca acaagttccc acaagccaag 2940

gaacaattga tttccgaaga agacgctgac cacttttgt tgttggctgc cagaccaggg 3000

caaaagccgg ttccattcgt gccagtcttg gacgaaagat ttgaattctt cttcaagaaa 3060

gattctcttt ggcaatctga agacttggaa agtgttgtcg acgaagacgt tcaaagaact 3120

tgtatcttgc acggtccagt tgcctcccaa ttcaccaaga aggttgatga accaattggc 3180

gaaatcttgg actctatcca cgagggccat attgccaagt tgatcaagga tgaatacgct 3240

ggtgatgcat ccaagatccc agttgttgag tacttcggtg gtttcaagac cgacaaggtt 3300

aatgctaaca atgttcaagt caatgctacc agaaaggaaa ccgtctacga aattggttcc 3360

aagttgccag ccaggcaaga ctggttggac ttgttggccg gtactgaatt gaactggttg 3420

cacgctttca tctccaccaa cagaattgtc caaggctcca agcacgtcgc caacccattg 3480

cacgacattt tggctcccgt tgccagatcc agtgtttcca ttgacaaggc taccaagaaa 3540

ttgactgctt atgaaaaggt caagggtgag ttggttccag ttgttgaaat tgaattggtc 3600

aagccaaaca ccattcaatt gtctttgatt gaacacagaa ctgctgatgg caaaccagtt 3660

gctttgccat tcttgtacaa gtacgaccca actgatgggt ttgcaccagt cttggagatc 3720

atggaaaaca gaaacgacag aatcaaggaa ttctactgga agttgtggtt cggtgcttcc 3780

gtcccttacg acaatgacat cgatgtcgaa gagcaaatct tgggtgacga aatcaccatt 3840

tcttctcaag acattggtga attcacacac gctattggta acaagtgtga agcctttgtc 3900

aacagaccag gtaaggtcac tttggctcca atggatttcg ccattgttgt tggttggaaa 3960

gctatcatca agtccatctt cccaaagacc gttgacggtg acttgttgaa gttggtccac 4020

ttgtccaacg gttacaagat gatccctggt gcagctccat tgcaaaaggg cgatgttgtt 4080

tccactagat ctgacatcaa ggctgttttg aaccaaccaa gtggtaagtt ggttgaagtt 4140

gttggtacca tcttccgtga aggcaagcca gttatggaag tcacttcaca attcttgtac 4200

cgtggtgaat acgacgacta ctgcaacacc ttccaaaagg tcactgaaac tccagttcaa 4260

gtctcattca agtctcctaa ggatttggct gttttgagat ccaaggaatg gttccatttg 4320

gaaaaggatg tcgagtttga tgctttgact ttcagatgtg aatccactta caagttcaag 4380

tctgccaacg tctactcgtc catcagaacg accggtcaag ttttcttgga gttgtccacc 4440

aaagaagtta tccaagttgg ttctgttgac tatgaagctg gtacctctta tggtaaccca 4500

gtcactgact acttgaacag aaacggtaag accattgaag aggctgttac ttttgagaat 4560

gccatcccct tgtcgtctgg tgaagagttg accaccaagg ctccgggtac caacgagcca 4620

tatgctattg tttctggtga ctacaaccca atccacgttt ccagagtctt ttctgcttac 4680

gccaagttgc caggtactat cacccacggt atgtactctt ctgccgccat cagagccttg 4740

gttgaagagt gggctgccaa caacgttgcc ccaagagtca gagccttcaa gtgtgaattt 4800

gttggtatgg ttttgccaaa cgacactttg caaaccacta tggaacacgt tggtatgatc 4860

aatggccgta agatcatcaa ggtcaagact gtcaatgccg agaccgagac tccagtcttg 4920

cttgctgaag ccgaaattga acaaccaacc accacctatg ttttcactgg tcaaggttcc 4980

caagaacaag gcatgggtat ggatttgtac aactcttctg aagttgcccg taacgtttgg 5040

gataccgccg acaagcattt catcaaccac tatggcttct ccatcttgga cattgtgcaa 5100

aacaacccta aggaattgac tatccacttt ggaggtgcta aaggtagagc tatcagagac 5160

aactacattg gtatgatgtt tgaaacaatt ggtgaagacg gttctttgaa gtccgaaaag 5220

atcttcaagg acattgacga aaacaccact tcctacacct ttgtttctga cactgggttg 5280

ttgtctgcta ctcaattcac ccaacccgct ttgactttga tggagaaggc tgcctacgac 5340

gatatcaagt ctaaaggatt gattccaagt gacatcatgt ttgctggtca ctctcttggt 5400

gaatactctg ctttgacttc cttggccaac gttatgccta ttgaatcctt ggttgatgtt 5460

gtcttctaca gaggtatgac catgcaagtt gctgttccaa gagacgagtt tggtagatcc 5520

aactacggta tggttgctgt caacccaacc agagtcagcc caacatttga cgatgccgcc 5580

atgagatttg ttgttgacga gactgccaag agaaccacct ggttgttgga aattgtcaac 5640

tacaatgttg aaaaccaaca gtacgttgct gctggtgact tgagagcctt ggataccttg 5700

accaacgtgt tgaatgtttt gaagatcaac aagattgata ttgtcagatt gcaagaacaa 5760

ttatccctcg acaaggtcaa ggagcacttg tacgagattg ttgatgaagt tgctgccaag 5820

tccattgcta agccacaacc aattgaatta gaaagaggtt ttgctgttat cccattgaag 5880

ggtatttctg tcccattcca ctcttcctac ttgatgtctg gtgtcaagcc attccagaga 5940

ttcttgtgca agaagattcc aaaggcttcc atcaaaccac aagatttgat tggcaagtac 6000

attcctaact tgactgctaa gccatttgaa cttactaagg aatacttcca ggatgtctac 6060

gacttgacta aatctgaaaa gatcaaggct atcttggaca actgggaaaa atacgaatag 6120

<210> 103
<211> 5661
<212> DNA
<213> Candida viswanathii

<400> 103
atgaagccag agattgaaca agaattatcc cacaccttgt taacagaatt gttagcttat 60

cagttcgctt ctccagtcag atggatcgaa acccaagatg tcttcttgaa gcaacacaac 120

accgaaagaa tcatcgaaat cggcccttcc ccaaccttgg ccggtatggc caacagaacc 180

atcaaggcca aataccaatc ctacgacgcc gcgttgtcct tgcaaagaga agtcttatgc 240

tactctaagg acgccaagga gatctactac aagccagatc cagcagatct tgctccaaag 300

gaagaaccaa agaaggaaga agctgccgcc gctccagccg ctacaccagc tgctgctgct 360

gctgctgcta ctcctgctgc tgccccagtc gccgctgctc cagccccatc tgctggccct 420

gctgaatcca tcccagatga accagtcaag gcttccttgt tgatccacgt cttggttgct 480

cagaaattaa agaaaccatt ggatgctgtt ccaatgtcca aggctatcaa agatttagtt 540

aacggtaagt ccactgtcca gaacgaaatt cttggtgact tgggtaaaga attcggttcc 600

actcctgaaa aaccagaaga taccccattg gaagaattgg ccgaacagtt ccaagactcc 660

ttcagtggtc aattgggtaa gacttctact tcattgattg gtagattgat gtcttctaag 720

atgcctggtg gtttctcaat caccgctgcc agaaaatact tggaatccag attcggtttg 780

ggtgccggta gacaagactc tgtcttgttg gttgctttga ccaacgaacc tgcaagcaga 840

ttgggttctg aggccgaagc taagaccttc ttggacacca tggctcagaa atatgcctca 900

tctgctggta tttccttgtc gtcagcttct gccggtgccg gtgctggagg tgccgccggt 960

ggcgccgttg ttgacagtgc tgctttggac gccttgactg ctgaaaacaa gaaattggct 1020

agacaacaat tagaggtctt ggctagatac ttgcaagtcg acttgaactc aggagctaag 1080

tcttttatca aagaaaaaga agcttccgct gttttgcaga aagaattgga cttgtgggaa 1140

gccgaacatg gtgaattcta cgccagaggt atcaaaccaa ctttctcagc tttgaaagca 1200

agaacctatg attcctactg gaactgggcc agacaagatg ttttgtccat gtactttgat 1260

attttgtttg gtaagttgac ctccgttgac agagaaacca tcgaccaatg tatccaaatt 1320

atgaacagat ccaacccaac tttgatcaag ttcatgcaat accacattga ccactgccca 1380

gaatacaagg gtgagactta caagttggcc aagagattgg gtcaacagtt gattgacaac 1440

tgtaagcaaa ccttgaatga agacccagtg tacaaggacg tttctagaat cactggtcca 1500

aagaccaccg tctgcgccaa gggtaacatt gaatacgaag aagccgaaaa ggattctgtt 1560

agaaagtttg aacagtacgt ctacgaaatg gcccaaggtg gtgaaatgac caagattgcc 1620

caaccaacta ttcaagaaga cttggccaga gtttacaaag ccatctccaa gcaagcttcc 1680

agagacagca agttggaatt gcagaaagtc tacgagcaat tgttgaaggt tgttgctggt 1740

tcagacgaaa ttgaaactca gcaattaacc aaggacatct tgcaagctcc aactggctcc 1800

aacaccccaa ctgatgaaga tgaaatttcc accgccgact ctgacgatga aattgcttca 1860

ttgccagaca agacttcaat tgcccaacca gtttcttcaa ctgttccacc ccagaccatc 1920

ccattcttgc acattcaaaa gaagaccaac gaaggctggg aatacgaccg caagttgtct 1980

gccctttact tggacggttt ggaatccgct gctgtcaacg gtctcacctt caaggacaag 2040

tacgttttgg ttaccggtgc tggtgctgga tccattggtg ccgaaatctt gcaaggtttg 2100

atcagtggtg gtgccaaggt tgttgttacc acctctagat tctccaagaa ggttactgag 2160

tactaccaaa acatgtacgc cagatacggt gctgccggtt ctactttgat tgttgttcca 2220

ttcaaccaag gttctaaaca agatgttgac gctttggttc aatacatcta cgacgatcca 2280

aagaagggtg gtttaggctg ggacttggat gccattatcc cattcgctgc tatcccagaa 2340

aatggtaacg gtatcgacaa cattgattct aaatccgaat ttgcccacag aattatgttg 2400

accaaccttt tgagattgtt gggtgctgtc aaatccaaga agactaccga caccagacca 2460

gctcaatgta tcttgccaat gtctcctaac cacggtactt tcggtttcga tgggttgtac 2520

tctgaatcca agatttcctt ggaaaccttg ttcaacagat ggtactccga agattggggc 2580

tccaagttga ccgtctgtgg tgccgttatt ggttggacca gaggtactgg tttgatgagc 2640

gccaacaaca tcattgccga aggtatcgaa aagattggtg tcagaacctt ctcccaaaag 2700

gaaatggctt tcaacatctt gggtttgttg actccagaga ttgtcaagtt gtgccaagaa 2760

gaaccagtta tggccgactt gaacggtggt ttgcaattca ttgaaaactt gaaggatttc 2820

acttccaagt tgagatctga cttgatggaa tccgctgaag ttagaagagc tgtctccatt 2880

gaatccgcca tcgaacaaaa ggttgtcaat ggtgacaatg ttgatgccaa ctacaccaag 2940

gttaccgttc aaccaagagc caacatgaaa ttcgacttcc caaccttgaa atcgtacgat 3000

gagatcaaga aggttgctcc agaattggaa ggcatgttgg acttggaatc cgtcgttgtt 3060

gtcaccggtt ttgccgaagt tggtccatgg ggtaacgcca gaaccagatg ggaaatggaa 3120

tccaagggtg aattctcctt ggaaggtgcc attgaaatgg cctggatcat gggtttcatc 3180

aagtaccaca acggtaactt gaagggtaag ccttactctg gttgggttga tgccaagacc 3240

caaactccaa tcgatgacaa ggacatcaag gccaagtacg aagaagagat cttggaccac 3300

tctggtatta gattgattga gccagaattg ttcaatggct acgatccaaa gaagaagcag 3360

atgatccaag aagttgtcat ccaacatgac ttggaaccat ttgaagcctc caaggaaact 3420

gctgaacaat acaaacacga acacggtgac aagtgtgaga tctttgaaat tgaagaatcc 3480

ggtgaataca ctgttagaat cttgaaaggt gctaccttgt ttgttccaaa ggctttgaga 3540

tttgacagat tggttgctgg tcaaattcca actggttggg atgctcgtac ctacggtatt 3600

ccagaagata ccattaacca agttgatcct atcactttgt acgtcttggt tgctaccgtt 3660

gaagctttgt tgtctgctgg tatcaccgac ccatatgaat tctacaagta cgtccacgtt 3720

tccgaagttg gtaactgttc tggttccggt atgggtggtg tctctgcctt gagaggaatg 3780

ttcaaggaca gatacgccga cagaccagtg caaaacgata tcttgcaaga atctttcatc 3840

aacaccatgt ccgcctgggt taacatgttg ttgttgtctt cttcgggtcc aatcaagacc 3900

ccagttggtg cctgtgctac cgctgttgaa tccgttgaca ttggtattga aactattttg 3960

tctggtaagg ctaaggttgt tatggttggt ggttacgatg acttccagga agaaggttct 4020

tatgaattcg ccaacatgaa tgccacttcc aactcccttg acgagtttgc tcacggcaga 4080

actccaaagg agatgtccag accaactacc accaccagac acggtttcat ggaggcccaa 4140

ggttctggta tccaagttat tatgactgct gacttggcca tcaagatggg tgttccaatt 4200

cacgctgtgt tggccatgtc tgctactgct accgacaaga ttggtagatc tgttccagct 4260

ccaggtaagg gtattttgac cactgccagg gaacaccacg gtaacttgaa gtacccatct 4320

ccagctttga acatcaagta cagaaagaga caattgaagg ctagattaga ccaaatcaag 4380

gcttgggaag aagctgaaat tgcttacttg caagacgaag ctgagttggc caaggaagaa 4440

atgggcgatg agttctccat gcacgaattc ttgaaggaaa gaactgaaga agtgtaccgt 4500

gaatccaaga gacaagtttc tgacgctaag aagcaatggg gtaaccaatt ctacaagtct 4560

gacccaagaa ttgctccatt gagaggtgcc ttggctgctt tcaacttgac cattgacgat 4620

cttggtgttg cttccttcca cggtacttct accgtcgcca acgataagaa cgaatccgcc 4680

actattaaca gcatgatgca acacttgggc agatctgaag gtaacccagt gtttggtgtt 4740

ttccagaagt acttgactgg tcatccaaag ggtgctgctg gtgcttggat gttgaacggt 4800

gccatccaga tcttggagtc tggtattgtt ccaggtaaca gaaatgccga taacgttgac 4860

aaggtcttgg aagaatacga gtacgtcttg tacccatcca gatccatcca aactgacggt 4920

atcaaggccg tttccgtgac ctctttcggt ttcggtcaaa aaggtgctca agctgttgtc 4980

gtccacccag actacttgtt tgctgttttg gatagatcta cttatgagga ctacgccacc 5040

agagtttctg ccagaaacaa gaagacttac cgttacatgc acaatgctat tactagaaac 5100

actatgtttg ttgctaagga taaggctcca tatgccgatg aattggaaca accagtttac 5160

ttggacccat tagcccgtgt tgaaaacgct aaggaaaagc ttgccttcag caacaagagt 5220

atccaatcca accaagctta tgctggtgaa aatgccagaa ccactgccaa ggctttggct 5280

gccttgaaca agtcatccaa gggtgttggt gtcgacgttg aattattgtc tgagctcaac 5340

ttggagaatg aaacttttgt tgcaagaaac ttcactcctg gtgaaatcca atactgctcc 5400

aagagtgcca acccacaagc ttcatacacc ggcacttggt ctgctaaaga agctgtcttc 5460

aaggcattag gtgttgaatc taaaggtgct ggtgccagct tggttgacat tgagatcact 5520

cgtgacgtca acggcgctcc acaagttgtc ttgcacggtg atgcggcaaa atcagccgcc 5580

aaagctggtg tcaagaacgt caagatttcc atctcccatg acgacttcca agccactgct 5640

gttgccttga gtgaattcta g 5661

<210> 104
<211> 1029
<212> DNA
<213> Candida viswanathii

<400> 104
atgttgagat cttcagtccg tactttctcc acccagtcca gagtattagc caactacggt 60

ttcgtaggct tgggtctcat gggccagcac atggccagac acgtctacaa ccagttgcag 120

ccagcagaca agttgtatgt ccacgacgtc aacccccagc acaccaccca gttcgtcacc 180

gacgtgacca cccagaagcc acagaacgcc acacaattga cgcccttgtc ctccttgaaa 240

gagttcacca ccgagccaga gtcccagttg gacttcatcg tcaccatggt ccccgagggc 300

aagcacgtca aggccgttgt ctccgagcta gtcgaccact acaatgcgtc gggaaaatac 360

gacccatcca agaagttgac ctttgtggac tcctccacca tcgacatccc cacctccagg 420

gaggtccacc agctcgttgc cgacaagtta caaggcgcca cgttcatcga cgccccggtt 480

tcgggtggtg tcgctggtgc caggaacgga accttgtcgt tcatggtgtc gcgggacacc 540

aaggaagacg tcgaccctaa cctcgtcacg cttttgaact acatgggcag caacatcttc 600

ccatgtggtg gaacccacgg gaccggcttg gctgccaagt tggcaaacaa ctacttgttg 660

gcgatcacga acatcgccgt cgcagatagc ttccagttgg caaactcgtt cgggttgaac 720

ttgcagaact acgccaagtt ggtgtcgacc tccacaggta agtcctgggc tagtgtcgat 780

aactgcccaa tccccggtgt ctaccctgaa aagaacttga cttgtgataa cggatacaag 840

ggtgggtttg tcacgaagtt gacgagaaag gatgtcgtct tggctacgga gtctgctaag 900

gctaacaacc agttccttat gcttggcgaa gtcggtagat actggtacga caaggcttgt 960

gaagatgaaa agtacgccaa cagagacttg tctgttcttt tcgaattctt gggtgatctt 1020

aaaaaataa 1029

<210> 105
<211> 1650
<212> DNA
<213> Candida viswanathii

<400> 105
atgttatcca gagttctttt caagactaaa ccaagagttc ctactaaatc aatcaccgcc 60

atggccatca gaaacaaatc catcgtgact ttatcctcca ccacctccac atacccaacc 120

gaccacacga ccccgtccac ggagccatac atcacgccat ccttcgtgaa caacgagttc 180

atcaagtcgg actccaacac ctggttcgac gtgcacgacc cggccacgaa ctacgtcgtg 240

tccaaggtgc cacagtcgac gccggaggag ttggaagagg cgatcgcgtc ggcccatgcc 300

gcgttcccca agtggcgcga caccagcatc atcaagcgtc aggggatcgc gttcaagttt 360

gtgcagttgt tgcgcgagaa catggacaga atcgcaagcg tcattgtctt ggaacagggt 420

aagacgtttg tcgatgccca gggtgacgtg actagaggat tgcaggttgc tgaggctgcg 480

tgcaacatca ctaatgactt gaagggtgag tcgttggaag tgtctactga tatggagacc 540

aagatgatta gagaaccttt gggtgttgtg ggatccatct gtccttttaa cttcccagct 600

atggtcccat tgtggtcttt gcctttggtt ttggtcacgg gtaacactgc tgtgattaag 660

ccttccgaga gagtcccggg cgcaagtatg attatttgtg aattggccgc caaggctggt 720

gttccacctg gtgtgttgaa cattgtccac ggtaagcacg acaccgtcaa caagttgatt 780

gaggacccaa gaatcaaggc attgactttt gttggtggtg acaaggccgg taagtacatt 840

tacgaaaagg gttccagttt gggcaagaga gtgcaggcca acttgggtgc taagaaccac 900

ttggttgtgt tgccagacgc acacaagcag agttttgtca atgccgtcaa cggtgccgct 960

ttcggtgctg ctggacagag atgtatggct atttctgtct tggtcaccgt gggtaagacc 1020

aaggaatggg tgcaggatgt catcaaggac gccaagttgt tgaacaccgg aagtggattt 1080

gacccaaaga gtgacttggg tccagtcatc aacccagagt ccttgactcg tgctgaagaa 1140

atcattgctg attccgtggc caacggtgcc gtgttggaat tggacggaag aggatacaga 1200

ccagaagacg ctagattcgc caagggtaac ttcttgggtc caaccatctt gaccaacgtc 1260

aagccaggct tgagagcata cgacgaagag attttcgctc ctgttttgtc tgtggttaac 1320

gtcgacacca ttgacgaagc cattgagttg atcaacaaca acaagtacgg taacggtgtt 1380

tcattattta cttcctccgg tggctcagcc cagtatttca ccaagagaat cgacgtcggt 1440

caagtcggta tcaatgtccc aatccctgtt ccattgccta tgttctcctt cactggttcc 1500

agaggctcct tcttgggtga cttgaacttc tacggtaagg ccggtatcac cttcttgacc 1560

aagccaaaga ccatcactag tgcctggaag accaacttga ttgatgacga gatcttgaaa 1620

ccatctacct cgatgcctgt ccaacagtaa 1650

<210> 106
<211> 726
<212> DNA
<213> Candida viswanathii

<400> 106
atgtccgacg aggaatcaga tatcttatac gaggtcagag acagaaccgc catcatcacc 60

ttgaacatcc ccaagagatt gaacgcattg aacggtgctc aatacttgaa gttgggtaag 120

ttcttggaga gagccaacaa cgaagaggac accgtcttga cattgatcca ggccctgggc 180

agattcttct ccgccggtgc caatttcgcc gacaacgata tggccaaggt cgaaatgtcc 240

aagttgttca gtcacgagta ctggttggaa agattcgtcg ccagaaacat ctggttgacc 300

aacttgttca acgaccacaa gaagatcttg gctgctgctg tcaatggtcc agttatcggt 360

ttgagcactg gtttgttgtt gttggtcgat ttggtctacg tccacgactt gaacaagttc 420

tacctcttgg ccccatttgc caacttgggt ttggttgccg aaggtgcttc ctctgccact 480

ttgttcaaca gattgggctg gtcaaaggct tctgaggcct tgttgttggc caagccaatc 540

ggcggccaag actgttacaa cgccggtttc atcaacaagc actacgacgg taagttttcc 600

tccactgaag agttcaacga acacgtctac aaggagttga cggaagcttt tgaaaacttg 660

cacgatgact ccattttgca gaacaagcaa ttgttgaagt tgtccagaga ccaggccatc 720

aactag 726

<210> 107
<211> 807
<212> DNA
<213> Candida viswanathii

<400> 107
atgtccgacg accttatcac ctacgaagtc aaagaccgag ctgccgtgat caccttgaac 60

aaccccaaga agctcaacgc cttatcgatc ccgcagtacg acaccatctg caagctctta 120

gaacgagcca acgccgaaga agacaccgtc atcaccttgc tccagtccac gggccgggtg 180

ttctctgccg gggccaacgc cgactccatc gtggggcagg atgccgagct cgagacctgg 240

ttgaacatgt cggtggccaa gcagacgttc ttggtgcaga cgttcctcgc acacaagaag 300

atccttgccg tcgccttgaa cggccccgtg attggcttat cggcggcgtt cgtggcgctc 360

tgcgacttgg tctacgtgca caacgccgca aagacgttct tcttgacccc gttcgccaac 420

atcgggatcc ttgccgaggg cggcacctca gccacgttgc ccatgcgcgt ggggtggtcc 480

agggccgcgg aagcgttgtt gttgtcaaag aggatttcgg gagatgactt gcagagagcg 540

gggttcttca ataaggacta caaggggcag ttcaagtccg cggaggagtt taacgaggtc 600

gtcttgaagg agttgcttga cgccacggaa aacttgcatg aggactcgat catccagaac 660

aaggagttgt tgaaggctat tttcaagcca aagatcagtg aggtcaactc gcaggaggtg 720

tcaagaggtg tgtacaagtg gacctctggg gtgccaatgg atagatttaa aaaattgctt 780

aatggtgagt tgaaacataa attatag 807

<210> 108
<211> 876
<212> DNA
<213> Candida viswanathii

<400> 108
atgccaaaca ctttagatca caactacctt aagaaaagcg tctggaagtc agatatcttt 60

gccggcaaag tcatcttcat cactggtggt gccggtacca tctgtcgtgt gcaagccgaa 120

gccatggtct tgttgggtgc taatgctgcc atcattggta gaaacgtcga aaagactgaa 180

gaagccgcta aggagattgc ttccttgaga ccaggagcca aggtcattgg tattggtgcc 240

gttgacgtgc gtaagatcca aactatcaag gatgccgttg acagaactgt tgccgagttg 300

ggcagaatcg actatgtcat tgctggtgcc gccggtaact tcctttgtga cttcaaccac 360

ttgtctgcca acgccttcaa gtccgtgatt gacattgatt tgttgggttc tttcaacacc 420

gtcaaggtta cctttgacca gttgagaaag aacaaaggtg ccgtgttgtt tgtcagtgcc 480

actttgcact acaccggtgt tcctatgcaa tcgcacgttg gtgctgccaa gtctggtgtt 540

gatgccttga gcaacgccct tgctgttgaa ttgggtccat tgggtatcag attcaactgt 600

attgccccag gtgccattgc tggtactgaa ggtatgctga gattggcccc acctaccgac 660

acgccattgg agacgaagat tccattgcag agacaaggta ctactgagga cattgctgat 720

gccactgtgt ttttgttctc gccagcttcc tcctacatta ctggtgatgt tttggttgtt 780

gatggtgcta tgtggcaaac tggcggaggt attcttaacg acttctaccc agacatcatt 840

attcaccaga acgccgaccc agaaggtaag ttgtaa 876

<210> 109
<211> 879
<212> DNA
<213> Candida viswanathii

<400> 109
atgccaaaca ctttaactga agcatacaaa caagaaagtt cttggaagcc agatttgttc 60

aagggcaaag tcgtcttcat caccggtggt gcaggttcga tctgccgtgt ccaggccgaa 120

gccatggtct tgttgggtgc caacgctgcc atcattggta gaaacgtcga aaagaccgaa 180

tccgccgcca aagaaatcgc cagcttgaga tcaggagcca aggttcttgg tattggtggc 240

attgatgtgc gtaaagttga cagcttgaag agcgccgtcg ataagaccgt tgccgaattg 300

ggcagaatcg actttgtcat tgctggtgcc gccggtaact tcctttgtga cttcaaccac 360

ttgtcttcca acgccttcaa atccatcatt gacattgact tgttgggttc tttcaacacc 420

gtcaaggtca cttttgacca attgagaaag aacaagggtg ccatcttgtt tgtcagtgcc 480

accttgcact actacggtgt tcctttccaa attggtgttg gtgctgccaa ggctggtgtc 540

gacgccttga gcaatgccct tgccgttgag ttgggtccat tgggcatcag atccaactgt 600

attgccccag gtccaatcga cggtaccgaa ggtgttgaaa gattggtcag agcttccaag 660

gccgaagccg ccaagaaggt tccattgcaa agattgggta ctacacaaga cattgctgac 720

ggtaccgtct acttgttctc cccagccgcc tcctttgtca ctggtgacgt tttggttgtt 780

gatggtgcta gctggcagat ctcctctggt gttggtgcca ccaactaccc agtctccatc 840

ctcaatgcga tcgacgcacc gaagggtggt aagttgtaa 879

<210> 110
<211> 4262
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 110
aaacgccagc aacgcggcct ttttacggtt cctggccttt tgctggcctt ttgctcacat 60

gttctttcct gcgttatccc ctgattctgt ggataaccgt attaccgcct ttgagtgagc 120

tgataccgct cgccgcagcc gaacgaccga gcgcagcgag tcagtgagcg aggaagcgag 180

ttcggcatgg cagatcatca tgcctgcagg agctccaatt gtaatatttc gggagaaata 240

tcgttggggt aaaacaacag agagagagag ggagagatgg ttctggtaga attataatct 300

ggttgttgca aatgctactg atcgactctg gcaatgtctg tagctcgcta gttgtatgca 360

acttaggtgt tatgcataca cacggttatt cggttgaatt gtggagtaaa aattgtctga 420

gttgtgtctt agctactggc tggcccccg cgaaagataa tcaaaattac acttgtgaat 480

ttttgcacac acaccgatta acatttccct tttttgtcca ccgatacacg cttgcctctt 540

ctttttttc tctgtgcttc cccctcctgt gacttttttcc accattgata taaaatcaac 600

tccatttccc taaaatctcc ccagattcta aaaacaactt cttctcttct gcttttcctt 660

tttttttgtt atatttattt accatcccct ttttttgaat agttattccc cactaacatt 720

gttcaaatct tcacgacata agaagagccc gggtctagat gtgtgctctt ccgagtgact 780

```
cttttgataa gagtcgcaaa tttgatttca taagtatata ttcattatgt aaagtagtaa      840

atggaaaatt cattaaaaaa aaagcaaatt tccgttgtat gcatactccg aacacaaaac      900

tagccccgga aaaaccctta gttgatagtt gcgaatttag gtcgaccata tgcgacgggt      960

acaacgagaa ttgtattgaa ttgatcaaga acatgatctt ggtgttacag aacatcaagt     1020

tcttggacca gactgagaat gcacagatat acaaggcgtc atgtgataaa atggatgaga     1080

tttatccaca attgaagaaa gagtttatgg aaagtggtca accagaagct aaacaggaag     1140

aagcaaacga agaggtgaaa caagaagaag aaggtaaata agtattttgt attatataac     1200

aaacaaagta aggaatacag atttatacaa taaattgcca tactagtcac gtgagatatc     1260

tcatccattc cccaactccc aagaaaaaaa aaaagtgaaa aaaaaaatca aacccaaaga     1320

tcaacctccc catcatcatc gtcatcaaac ccccagctca attcgcaatg gttagcacaa     1380

aaacatacac agaaagggca tcagcacacc cctccaaggt tgcccaacgt ttattccgct     1440

taatggagtc caaaaagacc aacctctgcg cctcgatcga cgtgaccaca accgccgagt     1500

tcctttcgct catcgacaag ctcggtcccc acatctgtct cgtgaagacg cacatcgata     1560

tcatctcaga cttcagctac gagggcacga ttgagccgtt gcttgtgctt gcagagcgcc     1620

acgggttctt gatattcgag gacaggaagt ttgctgatat cggaaacacc gtgatgttgc     1680

agtacacctc gggggtatac cggatcgcgg cgtggagtga catcacgaac gcgcacggag     1740

tgactgggaa gggcgtcgtt gaagggttga aacgcggtgc ggagggggta gaaaaggaaa     1800

ggggcgtgtt gatgttggcg gagttgtcga gtaaaggctc gttggcgcat ggtgaatata     1860

cccgtgagac gatcgagatt gcgaagagtg atcgggagtt cgtgattggg ttcatcgcgc     1920

agcgggacat ggggggtaga gaagaagggt ttgattggat catcatgacg cctggtgtgg     1980

ggttggatga taaaggcgat gcgttgggcc agcagtatag gactgttgat gaggtggttc     2040

tgactggtac cgatgtgatt attgtcggga gagggttgtt tggaaaagga agagaccctg     2100

aggtggaggg aaagagatac agggatgctg gatggaaggc atacttgaag agaactggtc     2160

agttagaata aatattgtaa taaataggtc tatatacata cactaagctt ctaggacgtc     2220

attgtagtct tcgaagttgt ctgctagttt agttctcatg atttcgaaaa ccaataacgc     2280

aatggatgta gcagggatgg tggttagtgc gttcctgaca aacccagagt acgccgcctc     2340

aaaccacgtc acattcgccc tttgcttcat ccgcatcact tgcttgaagg tatccacgta     2400

cgagttgtaa tacaccttga agaacggctt cgtctacgcg cgagacgaaa gggcctcgtg     2460

atacgcctat ttttataggt taatgtcatg ataataatgg tttcttagac gtcaggtggc     2520

acttttcggg gaaatgtgcg cggaacccct atttgtttat ttttctaaat acattcaaat     2580

atgtatccgc tcatgagaca ataaccctga taaatgcttc aataatattg aaaaaggaag     2640

agtatgagta ttcaacattt ccgtgtcgcc cttattccct tttttgcggc attttgcctt     2700
```

cctgtttttg ctcacccaga aacgctggtg aaagtaaaag atgctgaaga tcagttgggt 2760

gcacgagtgg gttacatcga actggatctc aacagcggta agatccttga gagttttcgc 2820

cccgaagaac gttttccaat gatgagcact tttaaagttc tgctatgtgg cgcggtatta 2880

tcccgtattg acgccgggca agagcaactc ggtcgccgca tacactattc tcagaatgac 2940

ttggttgagt actcaccagt cacagaaaag catcttacgg atggcatgac agtaagagaa 3000

ttatgcagtg ctgccataac catgagtgat aacactgcgg ccaacttact tctgacaacg 3060

atcggaggac cgaaggagct aaccgctttt ttgcacaaca tgggggatca tgtaactcgc 3120

cttgatcgtt gggaaccgga gctgaatgaa gccataccaa acgacgagcg tgacaccacg 3180

atgcctgtag caatggcaac aacgttgcgc aaactattaa ctggcgaact acttactcta 3240

gcttcccggc aacaattaat agactggatg gaggcggata aagttgcagg accacttctg 3300

cgctcggccc ttccggctgg ctggtttatt gctgataaat ctggagccgg tgagcgtggg 3360

tctcgcggta tcattgcagc actggggcca gatggtaagc cctcccgtat cgtagttatc 3420

tacacgacgg ggagtcaggc aactatggat gaacgaaata gacagatcgc tgagataggt 3480

gcctcactga ttaagcattg gtaactgtca gaccaagttt actcatatat actttagatt 3540

gatttaaaac ttcattttta atttaaaagg atctaggtga agatcctttt tgataatctc 3600

atgaccaaaa tcccttaacg tgagttttcg ttccactgag cgtcagaccc cgtagaaaag 3660

atcaaaggat cttcttgaga tccttttttt ctgcgcgtaa tctgctgctt gcaaacaaaa 3720

aaaccaccgc taccagcggt ggtttgtttg ccggatcaag agctaccaac tctttttccg 3780

aaggtaactg gcttcagcag agcgcagata ccaaatactg ttcttctagt gtagccgtag 3840

ttaggccacc acttcaagaa ctctgtagca ccgcctacat acctcgctct gctaatcctg 3900

ttaccagtgg ctgctgccag tggcgataag tcgtgtctta ccgggttgga ctcaagacga 3960

tagttaccgg ataaggcgca gcggtcgggc tgaacggggg gttcgtgcac acagcccagc 4020

ttggagcgaa cgacctacac cgaactgaga tacctacagc gtgagctatg agaaagcgcc 4080

acgcttcccg aagggagaaa ggcggacagg tatccggtaa gcggcagggt cggaacagga 4140

gagcgcacga gggagcttcc agggggaaac gcctggtatc tttatagtcc tgtcgggttt 4200

cgccacctct gacttgagcg tcgatttttg tgatgctcgt caggggggcg gagcctatgg 4260

aa 4262

<210> 111
<211> 7638
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic

polynucleotide

<400> 111

cccgtagaaa agatcaaagg atcttcttga gatccttttt ttctgcgcgt aatctgctgc 60

ttgcaaacaa aaaaaccacc gctaccagcg gtggtttgtt tgccggatca agagctacca 120

actctttttc cgaaggtaac tggcttcagc agagcgcaga taccaaatac tgtccttcta 180

gtgtagccgt agttaggcca ccacttcaag aactctgtag caccgcctac atacctcgct 240

ctgctaatcc tgttaccagt ggctgctgcc agtggcgata agtcgtgtct taccgggttg 300

gactcaagac gatagttacc ggataaggcg cagcggtcgg gctgaacggg gggttcgtgc 360

acacagccca gcttggagcg aacgacctac accgaactga gatacctaca gcgtgagcta 420

tgagaaagcg ccacgcttcc cgaagggaga aaggcggaca ggtatccggt aagcggcagg 480

gtcggaacag gagagcgcac gagggagctt ccagggggaa acgcctggta tctttatagt 540

cctgtcgggt ttcgccacct ctgacttgag cgtcgatttt tgtgatgctc gtcaggggggg 600

cggagcctat ggaaaaacgc cagcaacgcg gccttttttac ggttcctggg cttttgctgg 660

ccttttgctc acatgttctt tcctgcgtta tcccctgatt ctgtggataa ccgtattacc 720

gcctttgagt gagctgatac cgctcgccgc agccgaacga ccgagcgcag cgagtcagtg 780

agcgaggaag cggaagagcg cccaatacgc aaaccgcctc tccccgcgcg ttggccgatt 840

cattaatgca gctggcacga caggtttccc gactggaaag cgggcagtga gcgcaacgca 900

attaatgtga gttagctcac tcattaggca ccccaggctt tacactttat gcttccggct 960

cgtatgttgt gtggaattgt gagcggataa caatttcaca caggaaacag ctatgaccat 1020

gattacgcca agctatttag gtgacactat agaatactca agctatgcat caagcttggt 1080

accgagctcg gatccactag taacggccgc cagtgtgctg gaattcgccc ttccgttaaa 1140

caaaaatcag tctgtaaaaa aggttctaaa taaatattct gtctagtgta cacattctcc 1200

caaaatagtg aaatccagct ctacaatttg gctttaccgg tacaaatcaa agaccaatcg 1260

tcttcagtca actcctggta cttgtcgcca ttcaaccagt agatcaaatc cttgccgtct 1320

tcacccttg gcaacttttg gttcttgaat tggttcttag gaaccttgtg gttgtgcgaa 1380

gcctcaatgg tgccaatctt gatgaacgca ggttgagcat acacaggcaa agacttggtc 1440

acgtgagagt gaatcaattt caagatttct tcatggctca actcgtcctt ggcttcacag 1500

acggcaaaac aggctctacc ttcgtggttt ggcaccttga caccgacaac gacggactgc 1560

ttcaaggcct tggagcccat caattcgttc tcgacctcgg tggcggagac gttttcggac 1620

ttccaacgga aagtgtcacc taatctgtcg acaaagtaca acaatttgtc ctcgtccatc 1680

ttcaacaagt caccggatct gtaccacgcg tcacctttt tgaaaacatt ggtgaggatt 1740

ttgctgttgg tggcggactt gttaccataa taaccctgga aggatttctg cacgtcgtta 1800

gggttcaaga ttctcatcaa caactcacct ggctcgttgt aagcggcctc ggtacagaac 1860

ccggtcttgg ggtccttgta gatttcactc tcgtcttctg ggtccatctt ggccaatttc 1920

tgctgggtag acaataacaa gctgatgagg gacccgtact tacgacaggc gccgacaccg 1980

tactcaccgt actgcaagtt ggtggtggcg ataggggact cggtggcggc gtagaactca 2040

ccgatacctt caatgtggaa tctgcgcttg aactcagacc atatatctgg acgcaacccg 2100

ttaccgtagg caattctgac attgtgtctg tcttggtctg gatgaggctt ggagttcaac 2160

aagtaacgac agacctcacc gacgtattgc acgtgggtgg caccacataa tctggcctgg 2220

gtccagaacg aagtagcgga gaatttctgg gacacggaga cacagccacc gacaatcaaa 2280

gtaggacaca accccaacat ggccgcggtg gagtggtaca agggcatggc ggtcaagacg 2340

ttcgatttcg agtcaatctt catgatgtgg ccaaagaaaa ccgaggccat gaaggctttt 2400

ctccaggaca tgataccggc ttttggcaaa ccggtggtac ccgaggtgta aatcaatgca 2460

caagcggagg agtcagtatc ggttggtctt ctggtcttgt cctcggctct gtgttttgga 2520

gtcgacttga gtctcaatct gtcaaacaag gcaaactcgt caatgtagtt tatttgcaca 2580

tgtggcaatt cctctctgat ctgagcctcg gtatctctga ttggggaatc acagtccggg 2640

tcaacgaaaa cttgcgaagc gttgacaatc ttaagacagt ggatcaatgg cttgtccttg 2700

gtgttgaagt tcaagaacgc aggcaaggca ccaatgttcc acaatgccaa ccacaagaca 2760

atgaaaagcg gcttgttcat acaagaaaca ccgatggtgt cgttggcagt gacgccgtac 2820

tcgttcttca agatgtatga gtacttcaaa accatgtcgt acaattcctt gtaggtgtat 2880

tcttctaggt caaactgatc gtcgtagatc cactagtaac ggccgccagt gtgctggaat 2940

tcgcccttgg gctaacgaaa aggaaaccgc tgacgttaaa ggtatctacg gttgtttcgg 3000

tatgaccccc ggggatctga cgggtacaac gagaattgta ttgaattgat caagaacatg 3060

atcttggtgt tacagaacat caagttcttg gaccagactg agaatgcaca gatatacaag 3120

gcgtcatgtg ataaaatgga tgagatttat ccacaattga agaaagagtt tatggaaagt 3180

ggtcaaccag aagctaaaca ggaagaagca aacgaagagg tgaaacaaga agaagaaggt 3240

aaataagtat tttgtattat ataacaaaca aagtaaggaa tacagattta tacaataaat 3300

tgccatacta gtcacgtgag atatctcatc cattccccaa ctcccaagaa aaaaaaaaag 3360

tgaaaaaaaa aatcaaaccc aaagatcaac ctccccatca tcatcgtcat caaaccccca 3420

gctcaattcg caatggttag cacaaaaaca tacacagaaa gggcatcagc acacccctcc 3480

aaggttgccc aacgtttatt ccgcttaatg gagtccaaaa agaccaacct ctgcgcctcg 3540

atcgacgtga ccacaaccgc cgagttcctt tcgctcatcg acaagctcgg tccccacatc 3600

tgtctcgtga agacgcacat cgatatcatc tcagacttca gctacgaggg cacgattgag 3660

ccgttgcttg tgcttgcaga gcgccacggg ttcttgatat tcgaggacag gaagtttgct 3720

gatatcggaa acaccgtgat gttgcagtac acctcggggg tataccggat cgcggcgtgg 3780

agtgacatca cgaacgcgca cggagtgact gggaagggcg tcgttgaagg gttgaaacgc 3840

ggtgcggagg gggtagaaaa ggaaaggggc gtgttgatgt tggcggagtt gtcgagtaaa 3900

ggctcgttgg cgcatggtga atatacccgt gagacgatcg agattgcgaa gagtgatcgg 3960

gagttcgtga ttgggttcat cgcgcagcgg gacatggggg gtagagaaga agggtttgat 4020

tggatcatca tgacgcctgg tgtggggttg gatgataaag gcgatgcgtt gggccagcag 4080

tataggactg ttgatgaggt ggttctgact ggtaccgatg tgattattgt cgggagaggg 4140

ttgtttggaa aaggaagaga ccctgaggtg gagggaaaga gatacaggga tgctggatgg 4200

aaggcatact tgaagagaac tggtcagtta gaataaatat tgtaataaat aggtctatat 4260

acatacacta agcttctagg acgtcattgt agtcttcgaa gttgtctgct agtttagttc 4320

tcatgatttc gaaaaccaat aacgcaatgg atgtagcagg gatggtggtt agtgcgttcc 4380

tgacaaaccc agagtacgcc gcctcaaacc acgtcacatt cgccctttgc ttcatccgca 4440

tcacttgctt gaaggtatcc acgtacgagt tgtaatacac cttgaagaac ggcttcgtct 4500

acggtcgacg acgggtacaa cgagaattgt attgaattga tcaagaacat gatcttggtg 4560

ttacagaaca tcaagttctt ggaccagact gagaatgcac agatatacaa ggcgtcatgt 4620

gataaaatgg atgagattta tccacaattg aagaaagagt ttatggaaag tggtcaacca 4680

gaagctaaac aggaagaagc aaacgaagag gtgaaacaag aagaagaagg taaataagta 4740

ttttgtatta tataacaaac aaagtaagga atacagattt atacaataaa ttgccatact 4800

agtcacgtga gatatctcat ccattcccca actcccaaga aaaaaaaaaa gtgaaaaaaa 4860

aaatcaaacc caaagatcaa cctccccatc atcatcgtca tcaaaccccc agctcaattc 4920

gcagagctcg gtacccgggg atctggaatc cctcggcgtc ggtcttgggg gtgggggcat 4980

tctttcttgg tcttgggaac gccaacgctt tgttgtttgg gttcttgaac acggactgct 5040

cgaaaaagta ccagtatgat gccttacctc tgctggcttt ccacaagtat gggagggcat 5100

tgacagcgac tgtcttggct aacagcacgt cgtcggcaat taaatatttg gcttccaata 5160

actgactacc aaggatggca gcagcggcta tttctaatcc tgacatgttt ctcgtacgta 5220

gtagtgaatg aagggaaggt ggaataatat caagggcgaa ttctgcagat atccatcaca 5280

ctggcggccg ctcgagcatg catctagagg gcccaattcg ccctatagtg agtcgtatta 5340

caattcactg gccgtcgttt tacaacgtcg tgactgggaa aaccctggcg ttacccaact 5400

taatcgcctt gcagcacatc ccccttttcgc cagctggcgt aatagcgaag aggcccgcac 5460

cgatcgccct tcccaacagt tgcgcagcct atacgtacgg cagtttaagg tttacaccta 5520

taaaagagag agccgttatc gtctgtttgt ggatgtacag agtgatatta ttgacacgcc 5580

ggggcgacgg atggtgatcc ccctggccag tgcacgtctg ctgtcagata aagtctcccg 5640

tgaactttac ccggtggtgc atatcgggga tgaaagctgg cgcatgatga ccaccgatat 5700

ggccagtgtg ccggtctccg ttatcgggga agaagtggct gatctcagcc accgcgaaaa 5760

tgacatcaaa aacgccatta acctgatgtt ctggggaata taaatgtcag gcatgagatt 5820

atcaaaaagg atcttcacct agatcctttt cacgtagaaa gccagtccgc agaaacggtg 5880

ctgaccccgg atgaatgtca gctactgggc tatctggaca agggaaaacg caagcgcaaa 5940

gagaaagcag gtagcttgca gtgggcttac atggcgatag ctagactggg cggttttatg 6000

gacagcaagc gaaccggaat tgccagctgg ggcgccctct ggtaaggttg ggaagccctg 6060

caaagtaaac tggatggctt tctcgccgcc aaggatctga tggcgcaggg gatcaagctc 6120

tgatcaagag acaggatgag gatcgtttcg catgattgaa caagatggat tgcacgcagg 6180

ttctccggcc gcttgggtgg agaggctatt cggctatgac tgggcacaac agacaatcgg 6240

ctgctctgat gccgccgtgt tccggctgtc agcgcagggg cgcccggttc tttttgtcaa 6300

gaccgacctg tccggtgccc tgaatgaact gcaagacgag gcagcgcggc tatcgtggct 6360

ggccacgacg ggcgttcctt gcgcagctgt gctcgacgtt gtcactgaag cgggaaggga 6420

ctggctgcta ttgggcgaag tgccggggca ggatctcctg tcatctcacc ttgctcctgc 6480

cgagaaagta tccatcatgg ctgatgcaat gcggcggctg catacgcttg atccggctac 6540

ctgcccattc gaccaccaag cgaaacatcg catcgagcga gcacgtactc ggatggaagc 6600

cggtcttgtc gatcaggatg atctggacga agagcatcag gggctcgcgc cagccgaact 6660

gttcgccagg ctcaaggcga gcatgcccga cggcgaggat ctcgtcgtga cccatggcga 6720

tgcctgcttg ccgaatatca tggtggaaaa tggccgcttt tctggattca tcgactgtgg 6780

ccggctgggt gtggcggacc gctatcagga catagcgttg gctacccgtg atattgctga 6840

agagcttggc ggcgaatggg ctgaccgctt cctcgtgctt tacggtatcg ccgctcccga 6900

ttcgcagcgc atcgccttct atcgccttct tgacgagttc ttctgaatta ttaacgctta 6960

caatttcctg atgcggtatt ttctccttac gcatctgtgc ggtatttcac accgcataca 7020

ggtggcactt ttcggggaaa tgtgcgcgga acccctattt gtttatttttt ctaaatacat 7080

tcaaatatgt atccgctcat gagacaataa ccctgataaa tgcttcaata atagcacgtg 7140

aggagggcca ccatggccaa gttgaccagt gccgttccgg tgctcaccgc gcgcgacgtc 7200

gccggagcgg tcgagttctg gaccgaccgg ctcgggttct cccgggactt cgtggaggac 7260

gacttcgccg gtgtggtccg ggacgacgtg accctgttca tcagcgcggt ccaggaccag 7320

gtggtgccgg acaacaccct ggcctgggtg tgggtgcgcg gcctggacga gctgtacgcc 7380

gagtggtcgg aggtcgtgtc cacgaacttc cgggacgcct ccgggccggc catgaccgag 7440

atcggcgagc agccgtgggg gcgggagttc gccctgcgcg acccggccgg caactgcgtg 7500

cacttcgtgg ccgaggagca ggactgacac gtgctaaaac ttcattttta atttaaaagg 7560

atctaggtga agatcctttt tgataatctc atgaccaaaa tcccttaacg tgagttttcg 7620

ttccactgag cgtcagac 7638

<210> 112
<211> 7963
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 112
cccgtagaaa agatcaaagg atcttcttga gatccttttt ttctgcgcgt aatctgctgc 60

ttgcaaacaa aaaaaccacc gctaccagcg gtggtttgtt tgccggatca agagctacca 120

actctttttc cgaaggtaac tggcttcagc agagcgcaga taccaaatac tgtccttcta 180

gtgtagccgt agttaggcca ccacttcaag aactctgtag caccgcctac atacctcgct 240

ctgctaatcc tgttaccagt ggctgctgcc agtggcgata agtcgtgtct taccgggttg 300

gactcaagac gatagttacc ggataaggcg cagcggtcgg gctgaacggg gggttcgtgc 360

acacagccca gcttggagcg aacgacctac accgaactga gatacctaca gcgtgagcta 420

tgagaaagcg ccacgcttcc cgaagggaga aaggcggaca ggtatccggt aagcggcagg 480

gtcggaacag gagagcgcac gagggagctt ccagggggaa acgcctggta tctttatagt 540

cctgtcgggt ttcgccacct ctgacttgag cgtcgatttt tgtgatgctc gtcaggggg 600

cggagcctat ggaaaaacgc cagcaacgcg gccttttac ggttcctggg cttttgctgg 660

ccttttgctc acatgttctt tcctgcgtta tcccctgatt ctgtggataa ccgtattacc 720

gcctttgagt gagctgatac cgctcgccgc agccgaacga ccgagcgcag cgagtcagtg 780

agcgaggaag cggaagagcg cccaatacgc aaaccgcctc tccccgcgcg ttggccgatt 840

cattaatgca gctggcacga caggtttccc gactggaaag cgggcagtga gcgcaacgca 900

attaatgtga gttagctcac tcattaggca ccccaggctt tacactttat gcttccggct 960

cgtatgttgt gtggaattgt gagcggataa caatttcaca caggaaacag ctatgaccat 1020

gattacgcca agctatttag gtgacactat agaatactca agctatgcat caagcttggt 1080

accgagctcg gatccactag taacggccgc cagtgtgctg gaattcgccc tttaaacgtt 1140

gggcaacctt ggaggggtgt gctgatgccc tttctgtgta tgtttttgtg ctaaccattg 1200

cgaattgagc tgggggtttg atgacgatga tgatggggag gttgatcttt gggtttgatt 1260

ttttttttca ctttttttttt ttcttgggag ttggggaatg gatgagatat ctcacgtgac 1320

tagtatggca atttattgta taaatctgta ttccttactt tgtttgttat ataatacaaa 1380

atacttattt accttcttct tcttgtttca cctcttcgtt tgcttcttcc tgtttagctt 1440
ctggttgacc actttccata aactctttct tcaattgtgg ataaatctca tccattttat 1500
cacatgacgc cttgtatatc tgtgcattct cagtctggtc caagaacttg atgttctgta 1560
acaccaagat catgttcttg atcaattcaa tacaattctc gttgtacccg tcgcatatgg 1620
tcgacctaaa ttcgcaacta tcaactaagg gtttttccgg ggctagtttt gtgttcggag 1680
tatgcataca acggaaattt gctttttttt taatgaattt tccatttact actttacata 1740
atgaatatat acttatgaaa tcaaatttgc gactcttatc aaaagagtca ctctcttcta 1800
ttcttaactg gacaagattt cagcagcttc ttcggacttt tcacctcttt ctctgacttc 1860
aaggtctgga cggttcaaca tgtcttccaa agccttggag tatggagcct tggtgttttc 1920
tggtgggttc aaagccttga cagtttcgaa gtagtgttcg tagacgttac catcatatct 1980
accaatagca gcattggtca tcatgtcagt caagttgaaa ccatcagtca aaccaacaac 2040
gtttggtcta accttggcac acaaggctgg gaagtgctcg gaggaaatct ttctgctgac 2100
atctggggta atgataccgt actgcaagaa ggtggcacca aagttggtca agatccagtc 2160
ggcatacaag aaacccaaag gtctcaagac ttcagaccat tctggggtga tacgcttgaa 2220
gaaaccgtca atcaagaatc tgtgagcgtt caacttggaa accaaaacca aacttggacc 2280
ggtaatgtcg aaacccttgt cttcagcaac agcaccaatg tctctggcca agccggaaag 2340
ggcgtggtcg aaagccttaa ccaacttggc tgggtcgtcc aagtcggcaa cgctggagag 2400
aaccaatccc ttttgttctg gctccttcaa caagtctcta accattggct tggcaacgtt 2460
catggccaag atgttgttgt caccttccca ggtacattgg acaacccagt cggagtaggc 2520
tttaccgaaa ccgttgtaag acgagtaacc gtgaccacca caagcttgac gagcttcgtc 2580
aatggcttca gcagtcaacc aggtacaggt ggacttcaaa caaccggaag cgacgaacaa 2640
tttcttggat tcgacaatgg cagcgtcaat ggcttccttt tcaccagcac tgacagcttc 2700
gtccaacttg tcgttggttg cgttcatggt ttgttctaag tacaaggcac cttgggagaa 2760
caagtaagcg gcagccaaga atgggaacaa tctcttttga tgcaatgggt agtcaatcaa 2820
cttggtttca atggtatcgg tgtccttctt cttgaattgt cttctgtgga tggcgtatct 2880
caaggcaatg gtgatgaatc tactggtcat tctgtaggag tccatcatca tggtgactct 2940
accaccaatc aaagccgagt aagacaattg ttcagatggt ggcatggtga cttcacccag 3000
acgggaaacc ttacagtact tttgcaacat aaagaatctt gggattctga cgttggagaa 3060
ctggatccaa ccgttatcga taccgtctct acccatcttg gcaccaatgt caccaacagt 3120
cacacctggc tcgaggtcgt ggttggagtc tctcaatggg acaacaaagg tcttgacacc 3180
gtagtccttt cctttgacct tcaatctggc gtagacggag cagtgggtgg cggagtgcgc 3240

agcaccaccg atccaccact tggtggcacc aatgtgtggg gtgttgatga caaactcgtc 3300

agtgtcttcg tcaaaagtgg cggtggtttc aataccttgg acgttggaac cgtgggccaa 3360

ctcagtcata ccaaaacaac catagatacc tctcaacttg tcgataccct tatttatggt 3420

ccagtagaag aactgggagt tggtaccgtt accacggaca caggaaacaa acaaacccaa 3480

gttgacaccg attctggttc cgacttgtgg gtcaaagaca ccgaggatcg acaatctctg 3540

ggcctgttgg tctgggtact cgtgctcaaa gtatctggag agtctggcaa tcttcttggc 3600

ggtgacttct ctttgttggt ctttggtcaa gttgtagtat gaagcgtcga ccttcaagat 3660

agggtctttt tgcatttgtt cgaccatgtt gctgatgatc tcggatcttt cttgggaacc 3720

ttccaagaag tagttcaact cctttgggtt gaacttggtg agttctcttt ctttttgaag 3780

ttcggtaggc atgattgatt gttatagttt ctttctttct tttgaggatg accagatgtt 3840

tgataataat cagataggta ggaaaagaat aaaaaaaaac taatatggta ggaaagaaga 3900

aaaaaaatgg tggatgtgaa gaattgtgaa aaggggggcc aatgcctgaa gtttatataa 3960

agattcaggc tccagagttt gccctgttga ttctgcgggg ggccacacac aagcacactg 4020

cagacaggaa tagtgggggt agtgtgcatg tggataaaaa tgcctgtgtt ttttttttct 4080

gtttcttctg atgccattgt cgcacgggca caactttggc tgtctgcccc cccaggcgta 4140

ataaccgagg attatctcac aactacagtt gtcacttttg ccagcaaaat aggaatcaac 4200

aagcaatggg ttatataaca ctgcttatac agactactac caatcccaat acacaccact 4260

gagcctacta ctcacagaaa ttgagcgacc accacccaat tgatcgggaa tatgtctccc 4320

tgattatttt tccacacttt aattacccca taaaacggtt acatcaagga taatgaagtg 4380

ggatatgcaa agtttcaatt ttttcctctc cgctctaacc aagggattct ccatctctat 4440

ccgtggtcca aacagggcca aacaccaccc aagctctact ctagaacctt ctggtatcta 4500

cgtgccagct taattacatc aactcaacga ctcttcttcc cttctgcagg catgatgatc 4560

tgccatgccg aactagacga agccgttctt caaggtgtat tacaactcgt acgtggatac 4620

cttcaagcaa gtgatgcgga tgaagcaaag ggcgaatgtg acgtggtttg aggcggcgta 4680

ctctgggttt gtcaggaacg cactaaccac catccctgct acatccattg cgttattggt 4740

tttcgaaatc atgagaacta aactagcaga caacttcgaa gactacaatg acgtcctaga 4800

agcttagtgt atgtatatag acctatttat tacaatattt attctaactg accagttctc 4860

ttcaagtatg ccttccatcc agcatccctg tatctctttc cctccacctc agggtctctt 4920

ccttttccaa acaaccctct cccgacaata atcacatcgg taccagtcag aaccacctca 4980

tcaacagtcc tatactgctg gcccaacgca tcgcctttat catccaaccc cacaccaggc 5040

gtcatgatga tccaatcaaa cccttcttct ctacccccca tgtcccgctg cgcgatgaac 5100

ccaatcacga actcccgatc actcttcgca atctcgatcg tctcacgggt atattcacca 5160

tgcgccaacg agcctttact cgacaactcc gccaacatca acacgcccct ttccttttct 5220

acccccctccg caccgcgttt caacccttca acgacgccct tcccagtcac tccgtgcgcg 5280

ttcgtgatgt cactccacgc cgcgatccgg tatacccccg aggtgtactg caacatcacg 5340

gtgtttccga tatcagcaaa cttcctgtcc tcgaatatca agaacccgtg gcgctctgca 5400

agcacaagca acggctcaat cgtgccctcg tagctgaagt ctgagatgat atcgatgtgc 5460

gtcttcacga gacagatgtg gggaccgagc ttgtcgatga gcgaaaggaa ctcggcggtt 5520

gtggtcacgt cgatcgaggc gcagaggttg gtctttttgg actccattaa gcggaaaagg 5580

gcgaattctg cagatatcca tcacactggc ggccgctcga gcatgcatct agagggccca 5640

attcgcccta tagtgagtcg tattacaatt cactggccgt cgttttacaa cgtcgtgact 5700

gggaaaaccc tggcgttacc caacttaatc gccttgcagc acatcccccct ttcgccagct 5760

ggcgtaatag cgaagaggcc cgcaccgatc gcccttccca acagttgcgc agcctatacg 5820

tacggcagtt taaggtttac acctataaaa gagagagccg ttatcgtctg tttgtggatg 5880

tacagagtga tattattgac acgccggggc gacggatggt gatccccctg gccagtgcac 5940

gtctgctgtc agataaagtc tcccgtgaac tttacccggt ggtgcatatc ggggatgaaa 6000

gctggcgcat gatgaccacc gatatggcca gtgtgccggt ctccgttatc ggggaagaag 6060

tggctgatct cagccaccgc gaaaatgaca tcaaaaacgc cattaacctg atgttctggg 6120

gaatataaat gtcaggcatg agattatcaa aaaggatctt cacctagatc cttttcacgt 6180

agaaagccag tccgcagaaa cggtgctgac cccggatgaa tgtcagctac tgggctatct 6240

ggacaaggga aaacgcaagc gcaaagagaa agcaggtagc ttgcagtggg cttacatggc 6300

gatagctaga ctgggcggtt ttatggacag caagcgaacc ggaattgcca gctggggcgc 6360

cctctggtaa ggttgggaag ccctgcaaag taaactggat ggctttctcg ccgccaagga 6420

tctgatggcg caggggatca agctctgatc aagagacagg atgaggatcg tttcgcatga 6480

ttgaacaaga tggattgcac gcaggttctc cggccgcttg ggtggagagg ctattcggct 6540

atgactgggc acaacagaca atcggctgct ctgatgccgc cgtgttccgg ctgtcagcgc 6600

aggggcgccc ggttcttttt gtcaagaccg acctgtccgg tgccctgaat gaactgcaag 6660

acgaggcagc gcggctatcg tggctggcca cgacgggcgt tccttgcgca gctgtgctcg 6720

acgttgtcac tgaagcggga agggactggc tgctattggg cgaagtgccg gggcaggatc 6780

tcctgtcatc tcaccttgct cctgccgaga aagtatccat catggctgat gcaatgcggc 6840

ggctgcatac gcttgatccg gctacctgcc cattcgacca ccaagcgaaa catcgcatcg 6900

agcgagcacg tactcggatg gaagccggtc ttgtcgatca ggatgatctg gacgaagagc 6960

atcaggggct cgcgccagcc gaactgttcg ccaggctcaa ggcgagcatg cccgacggcg 7020

aggatctcgt cgtgacccat ggcgatgcct gcttgccgaa tatcatggtg gaaaatggcc 7080

gcttttctgg attcatcgac tgtggccggc tgggtgtggc ggaccgctat caggacatag 7140

cgttggctac ccgtgatatt gctgaagagc ttggcggcga atgggctgac cgcttcctcg 7200

tgctttacgg tatcgccgct cccgattcgc agcgcatcgc cttctatcgc cttcttgacg 7260

agttcttctg aattattaac gcttacaatt tcctgatgcg gtattttctc cttacgcatc 7320

tgtgcggtat ttcacaccgc atacaggtgg cacttttcgg ggaaatgtgc gcggaacccc 7380

tatttgttta tttttctaaa tacattcaaa tatgtatccg ctcatgagac aataaccctg 7440

ataaatgctt caataatagc acgtgaggag ggccaccatg gccaagttga ccagtgccgt 7500

tccggtgctc accgcgcgcg acgtcgccgg agcggtcgag ttctggaccg accggctcgg 7560

gttctcccgg gacttcgtgg aggacgactt cgccggtgtg gtccgggacg acgtgaccct 7620

gttcatcagc gcggtccagg accaggtggt gccggacaac accctggcct gggtgtgggt 7680

gcgcggcctg gacgagctgt acgccgagtg gtcggaggtc gtgtccacga acttccggga 7740

cgcctccggg ccggccatga ccgagatcgg cgagcagccg tgggggcggg agttcgccct 7800

gcgcgacccg gccggcaact gcgtgcactt cgtggccgag gagcaggact gacacgtgct 7860

aaaacttcat ttttaattta aaaggatcta ggtgaagatc ctttttgata atctcatgac 7920

caaaatccct taacgtgagt tttcgttcca ctgagcgtca gac 7963

<210> 113
<211> 751
<212> DNA
<213> Candida viswanathii

<400> 113
aagggaagaa gagtcgttga gttgatgtaa ttaagctggc acgtagatac cagaaggttc 60

tagagtagag cttgggtggt gtttggccct gtttggacca cggatagaga tggagaatcc 120

cttggttaga gcggagagga aaaaattgaa actttgcata tcccacttca ttatccttga 180

tgtaaccgtt ttatggggta attaaagtgt ggaaaaataa tcagggagac atattcccga 240

tcaattgggt ggtggtcgct caatttctgt gagtagtagg ctcagtggtg tgtattggga 300

ttggtagtag tctgtataag cagtgttata taacccattg cttgttgatt cctattttgc 360

tggcaaaagt gacaactgta gttgtgagat aatcctcggt tattacgcct ggggggggcag 420

acagccaaag ttgtgcccgt gcgacaatgg catcagaaga aacagaaaaa aaaaacacag 480

gcatttttat ccacatgcac actaccccca ctattcctgt ctgcagtgtg cttgtgtgtg 540

gccccccgca gaatcaacag ggcaaactct ggagcctgaa tctttatata aacttcaggc 600

attggccccc cttttcacaa ttcttcacat ccaccatttt ttttcttctt tcctaccata 660

ttagtttttt tttattcttt tcctacctat ctgattatta tcaaacatct ggtcatcctc 720

aaaagaaaga aagaaactat aacaatcaat c 751

<210> 114
<211> 991
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 114
ttgtccaatg taataatttt tccatgacta aaaagtgtgt gttggtgtaa agaagaaagt 60

ggaagggacg ttggtgatgg tgagttcgtc tatccctttt ttatagttgc ttgtatagta 120

ggctactctt ctagggactc gatgggggaa ggttcttgat atttgcttag ttcgagaagg 180

ttccagatga gcgagacatt tttggtacga cattgggtgg atgatctgca cgacattttg 240

tgattcttgc gacacgctgc actaccaagt gtagtctggc tgaacggatc acaagataaa 300

cctctgaaaa attatctcag ggcatgcaac aacaattata catagaagag ggagtcacga 360

tatacacctg tgaaggaatc atgtggtcgg ctctccttga actttgaatt catgcaatta 420

ttaagaagaa gcacaggtga gcaacccacc atacgttcat ttgcaccacc tgatgattaa 480

aagccaaaga aagaaaaaaa aaaaagaaac aggcggtggg aattgttaca acccacgcga 540

acccgaaaat ggagcaatct tccccggggc ctccaaatac caactcaccc gagagagaga 600

aagagacacc acccaccacg agacggagta tatccaccaa ggtaagtaac tcagggttaa 660

tgatacaggt gtacacagct ccttccctag ccattgagtg ggtatcacat gacactggca 720

ggttacaacc acgtttagta gttattttgt gcaattccat ggggatcagg aagtttggtt 780

tggtgggtgc gtctactgat tccccttttgt ctctgaaaat cttttcccta gtggaacact 840

ttggctgaat gatataaatt caccttgatt cccaccctcc cttctttctc tctctctctg 900

ttacacccaa ttgaattttc ttttttttttt ttttactttc cctccttctt tatcatcaaa 960

gataagtaag tttatcaatt gcctattcag a 991

<210> 115
<211> 351
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 115
aaagtacggt gttgttgaca agttgtccca cgtttctacc ggtggtggtg cttctttgga 60

attgttagaa ggtaaggaat tgccaggtgt gactgctctt tccaacaagg cctaaatcag 120

aaatatttat tctaataaat aggatgcaaa aaaaaaaccc cccttaataa aaaaaaaaga 180

aacgattttt tatctaatga agtctatgta tctaacaaat gtatgtatca atgtttattc 240

cgttaaacaa aaatcagtct gtaaaaaagg ttctaaataa atattctgtc tagtgtacac 300

attctcccaa aatagtgaaa tccagctcta caatttagct ctaccagtac a 351

<210> 116
<211> 184
<212> DNA
<213> Candida viswanathii

<400> 116
gaatagaaga gagtgactct tttgataaga gtcgcaaatt tgatttcata agtatatatt 60

cattatgtaa agtagtaaat ggaaaattca ttaaaaaaaa agcaaatttc cgttgtatgc 120

atactccgaa cacaaaacta gccccggaaa aacccttagt tgatagttgc gaatttaggt 180

cgac 184

<210> 117
<211> 531
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 117
gagctccaat tgtaatattt cgggagaaat atcgttgggg taaaacaaca gagagagaga 60

gggagagatg gttctggtag aattataatc tggttgttgc aaatgctact gatcgactct 120

ggcaatgtct gtagctcgct agttgtatgc aacttaggtg ttatgcatac acacggttat 180

tcggttgaat tgtggagtaa aaattgtctg agttgtgtct tagctactgg ctggcccccc 240

gcgaaagata atcaaaatta cacttgtgaa tttttgcaca cacaccgatt aacatttccc 300

ttttttgtcc accgatacac gcttgcctct tctttttttt ctctgtgctt ccccctcctg 360

tgacttttttc caccattgat ataaaatcaa ctccatttcc ctaaaatctc cccagattct 420

aaaaacaact tcttctcttc tgcttttcct tttttttttgt tatatttatt taccatccct 480

ttttttttgaa tagttattcc ccactaacat tgttcaaatc ttcacgacat a 531

<210> 118
<211> 745
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 118
aaaatcagag gctactccgg agatggcaca ttatcacgtg ggcagtccca tctctctgac 60

aatggcactg acaatgccat gtcattctag aagtcgctcg ttccattgct agtcgtgtag 120

gtgtcttact cagaccatat gtttggctgt gtgtgtggta agggggtaag atatctctag 180

gaggaggcaa ctctactcga aagtacgaaa tggccgacgc caatccgacc tgccattgtg 240

caggacgcgg gtcgagatcc ggaagtgccc tccttctgag gttgttggtt tgttccagtt 300

ggcataagaa attaataatc gtttagtcaa tggtacagta ttaataattc ctgacaaaac 360

cgtgccctac aagcagtgtg agttccgggg gtagcactag tggtcgtggt ggtagggccc 420

acagagagta tttccggaac cgactgatcc tttgattcca atattccaag atttgggttt 480

gtggttgtgt gtaatgcatg aggttatttt gtcaacctcc ttccttctct acggttgttg 540

ctcctcctcc tcttccttcc ctttgttcat ttaaatacac acacaccact ccccttccca 600

cccacaagaa ctttttttt ttttactttc ttcttctctt actttcttct tcttctttca 660

ttacctgtta gattatattt atcaacctca attggtttat tatctaccac ccatacaacc 720

cccacccctg ccacatccac caaca 745

<210> 119
<211> 614
<212> DNA
<213> Candida viswanathii

<400> 119
cggaagttgt ttaccgacct gaccgtaaat ttgctgctga aagaaacgtg tcaaacaaga 60

ccaattggct caattgaccc tgtggaaatg ctttgttgac caccaatgct tccaccaaac 120

gttacttttt ttttgcaatc ggatggtatg ggtctggggt tcacctgttt tgtaaagcta 180

cagaaggtgg catatttctc tgatcaggtg ttttttttt cggctgctgc tgctcgtggt 240

ggtgtagtgg tagtggtgtg tgtgtgtgtg tgtgcgtgcg tgtggaagga cgcttttgc 300

tctctgactc ctcccaatca gaagttgcta tagtggtgaa acaacaatgg atgataatgc 360

cccgggcggt gcgtgtccga cacaaaccac tacatttttt agctgggagc ctactgccac 420

tacgacccac ccacccatgg tcaacaaaaa aattctgaca aattataaaa taacccttga 480

attccccctt ggaaaaattt ttggtatttc tctctctctt ttcctttccc tcttcttttt 540

ctctccatca atcaattgac gttcagtaac tcaattaatt acatcacatc cctcaattaa 600

agaatttaat cgat 614

<210> 120
<211> 500
<212> DNA
<213> Candida viswanathii

<400> 120
ctagcaaagg cttgatcaga gaaagcaaca aaaaaaaaaa ctctaatact ccagaataca 60

ctcctttaga aacacacaac aaacaagcct agactaccat ggactacgat gaagacgatt 120

tagattacat ttctcaagga gaagaggaag agtttgacga aaacaagttg aacaacgaag 180

agtacgactt gttgcatgac atgcttccgg agttgaagac aaaattgaaa gattacaatg 240

atgagatccc agattacgat ttaaaggaag cgttatacta caactatttc gagatagacc 300

ctaccattga agaattgaag acgaaattca aaaagagtac gtatatacaa ctaacatcaa 360

cgcctttcta gtttctgttc tgtctccaat gcttctcctg gtttcttcat ggttctctct 420

gtaccaacaa ggaaaaaaaa aaaaatctgg caaaaaaaac caaaccaacc aatgttctta 480

ctcaccaacg ccctacaatc 500

<210> 121
<211> 536
<212> DNA
<213> Candida viswanathii

<400> 121
gaagatgaag cgtatgagta ttatgagtac tgtcggacgt tggaaggtgg cagagttaag 60

cccgagaaag caaggaagga gtgggagatg atgagtgatg cggccaaaga ggatgtgaag 120

gctgcgtatc tgtttttgat agctggtggt agccgaatag aggaaggcaa gcttgttcat 180

attggatgat gatggtagat ggtggctgcc aaagtggttg taaatagaaa aaagtgggtt 240

tgggtctgtt gatagttagt ggtggcggct gtctgtgatt acgtcagcaa gtagcacctc 300

ggcagttaaa acagcagcaa cagaaaaaaa atgtgtgaaa gtttgattcc cccacagtct 360

accacaccca gagttccatt tatccataat atcacaagca atagaaaaat aaaaaattat 420

caacaaatca caacgaaaag attctgcaaa attattttca cttcttcttt tgacttcctc 480

ttcttcttgt taggttcttt ccatattttc cccttaaacc catacacaac gcagcc 536

<210> 122
<211> 8006
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 122
cgaattccag cacactggcg gccgttacta gtggatccga gctcggtacc aagcttgatg 60

catagcttga gtattctata gtgtcaccta aatagcttgg cgtaatcatg gtcatagctg 120

tttcctgtgt gaaattgtta tccgctcaca attccacaca acatacgagc cggaagcata 180

aagtgtaaag cctggggtgc ctaatgagtg agctaactca cattaattgc gttgcgctca 240

ctgcccgctt tccagtcggg aaacctgtcg tgccagctgc attaatgaat cggccaacgc 300

gcggggagag gcggtttgcg tattgggcgc tcttccgctt cctcgctcac tgactcgctg 360

cgctcggtcg ttcggctgcg gcgagcggta tcagctcact caaaggcggt aatacggtta 420

tccacagaat caggggataa cgcaggaaag aacatgtgag caaaaggcca gcaaaagccc 480

aggaaccgta aaaaggccgc gttgctggcg tttttccata ggctccgccc ccctgacgag 540

catcacaaaa atcgacgctc aagtcagagg tggcgaaacc cgacaggact ataaagatac 600

caggcgtttc cccctggaag ctccctcgtg cgctctcctg ttccgaccct gccgcttacc 660

ggatacctgt ccgcctttct cccttcggga agcgtggcgc tttctcatag ctcacgctgt 720

aggtatctca gttcggtgta ggtcgttcgc tccaagctgg gctgtgtgca cgaaccccc 780

gttcagcccg accgctgcgc cttatccggt aactatcgtc ttgagtccaa cccggtaaga 840

cacgacttat cgccactggc agcagccact ggtaacagga ttagcagagc gaggtatgta 900

ggcggtgcta cagagttctt gaagtggtgg cctaactacg gctacactag aaggacagta 960

tttggtatct gcgctctgct gaagccagtt accttcggaa aaagagttgg tagctcttga 1020

tccggcaaac aaaccaccgc tggtagcggt ggtttttttg tttgcaagca gcagattacg 1080

cgcagaaaaa aaggatctca agaagatcct ttgatctttt ctacggggtc tgacgctcag 1140

tggaacgaaa actcacgtta agggattttg gtcatgagat tatcaaaaag gatcttcacc 1200

tagatccttt taaattaaaa atgaagtttt agcacgtgtc agtcctgctc ctcggccacg 1260

aagtgcacgc agttgccggc cgggtcgcgc agggcgaact cccgccccca cggctgctcg 1320

ccgatctcgg tcatggccgg cccggaggcg tcccggaagt tcgtggacac gacctccgac 1380

cactcggcgt acagctcgtc caggccgcgc acccacaccc aggccagggt gttgtccggc 1440

accacctggt cctggaccgc gctgatgaac agggtcacgt cgtcccggac cacaccggcg 1500

aagtcgtcct ccacgaagtc ccgggagaac ccgagccggt cggtccagaa ctcgaccgct 1560

ccggcgacgt cgcgcgcggt gagcaccgga acggcactgg tcaacttggc catggtggcc 1620

ctcctcacgt gctattattg aagcatttat cagggttatt gtctcatgag cggatacata 1680

tttgaatgta tttagaaaaa taaacaaata ggggttccgc gcacatttcc ccgaaaagtg 1740

ccacctgtat gcggtgtgaa ataccgcaca gatgcgtaag gagaaaatac cgcatcagga 1800

aattgtaagc gttaataatt cagaagaact cgtcaagaag gcgatagaag gcgatgcgct 1860

gcgaatcggg agcggcgata ccgtaaagca cgaggaagcg gtcagcccat tcgccgccaa 1920

gctcttcagc aatatcacgg gtagccaacg ctatgtcctg atagcggtcc gccacaccca 1980

gccggccaca gtcgatgaat ccagaaaagc ggccattttc caccatgata ttcggcaagc 2040

aggcatcgcc atgggtcacg acgagatcct cgccgtcggg catgctcgcc ttgagcctgg 2100

cgaacagttc ggctggcgcg agcccctgat gctcttcgtc cagatcatcc tgatcgacaa 2160

gaccggcttc catccgagta cgtgctcgct cgatgcgatg tttcgcttgg tggtcgaatg 2220

ggcaggtagc cggatcaagc gtatgcagcc gccgcattgc atcagccatg atggatactt 2280

tctcggcagg agcaaggtga gatgacagga gatcctgccc cggcacttcg cccaatagca 2340

gccagtccct tcccgcttca gtgacaacgt cgagcacagc tgcgcaagga acgcccgtcg 2400

tggccagcca cgatagccgc gctgcctcgt cttgcagttc attcagggca ccggacaggt 2460

cggtcttgac aaaaagaacc gggcgcccct gcgctgacag ccggaacacg gcggcatcag 2520

agcagccgat tgtctgttgt gcccagtcat agccgaatag cctctccacc caagcggccg 2580

gagaacctgc gtgcaatcca tcttgttcaa tcatgcgaaa cgatcctcat cctgtctctt 2640

gatcagagct tgatcccctg cgccatcaga tccttggcgg cgagaaagcc atccagttta 2700

ctttgcaggg cttcccaacc ttaccagagg gcgccccagc tggcaattcc ggttcgcttg 2760

ctgtccataa aaccgcccag tctagctatc gccatgtaag cccactgcaa gctacctgct 2820

ttctctttgc gcttgcgttt tcccttgtcc agatagccca gtagctgaca ttcatccggg 2880

gtcagcaccg tttctgcgga ctggctttct acgtgaaaag gatctaggtg aagatccttt 2940

ttgataatct catgcctgac atttatattc cccagaacat caggttaatg gcgtttttga 3000

tgtcattttc gcggtggctg agatcagcca cttcttcccc gataacggag accggcacac 3060

tggccatatc ggtggtcatc atgcgccagc tttcatcccc gatatgcacc accgggtaaa 3120

gttcacggga gactttatct gacagcagac gtgcactggc cagggggatc accatccgtc 3180

gccccggcgt gtcaataata tcactctgta catccacaaa cagacgataa cggctctctc 3240

ttttataggt gtaaacctta aactgccgta cgtataggct gcgcaactgt tgggaagggc 3300

gatcggtgcg ggcctcttcg ctattacgcc agctggcgaa aggggatgt gctgcaaggc 3360

gattaagttg ggtaacgcca gggttttccc agtcacgacg ttgtaaaacg acggccagtg 3420

aattgtaata cgactcacta tagggcgaat tgggccccat gctcgagagt gctccgaaaa 3480

tacctcaagg gtggttagag ttacatgata aacaacaggc caaaaaaaga gagggcgttt 3540

ctcatagagg ggtttttctt tcgttcaaat ctaggggggt tgagaggcgg aagtcctggc 3600

ggaacagaag tgagctgggt aatcatacat ctcaagtaat gttgcatacc ggagagaaga 3660

aaaaacttgg ccctcttctc tggctgagga gtaaactaga ggacggcagg tcggatagtt 3720

ttttttgac tgtcgtagct attgcctatg gccagcacca gcaccatcac aaactccagc 3780

tctcgttgac acacctgccc gcgagcatgc ggcccatgca gcggaaaaga gacctctttg 3840

acggaaatat gtcttagggg gtgttttctt actgggggggt tggtaggact tcggaaacac 3900

ggagtcctct tcttcggaac gccccccctgg ttgccggccg gttcgtcgct tgatcccccct 3960

gcaggaaaat cagaggctac tccggagatg gcacattatc acgtgggcag tcccatctct 4020

ctgacaatgg cactgacaat gccatgtcat tctagaagtc gctcgttcca ttgctagtcg 4080

tgtaggtgtc ttactcagac catatgtttg gctgtgtgtg tggtaagggg gtaagatatc 4140

tctaggagga ggcaactcta ctcgaaagta cgaaatggcc gacgccaatc cgacctgcca 4200

ttgtgcagga cgcgggtcga gatccggaag tgccctcctt ctgaggttgt tggtttgttc 4260

cagttggcat aagaaattaa taatcgttta gtcaatggta cagtattaat aattcctgac 4320

aaaaccgtgc cctacaagca gtgtgagttc cgggggtagc actagtggtc gtggtggtag 4380

ggcccacaga gagtatttcc ggaaccgact gatcctttga ttccaatatt ccaagatttg 4440

ggtttgtggt tgtgtgtaat gcatgaggtt attttgtcaa cctccttcct tctctacggt 4500

tgttgctcct cctcctcttc cttccctttg ttcatttaaa tacacacaca ccactcccct 4560

tcccacccac aagaactttt tttttttta ctttcttctt ctcttacttt cttcttcttc 4620

tttcattacc tgttagatta tatttatcaa cctcaattgg tttattatct accacccata 4680

caacccccac ccctgccaca tccaccaaca atggacgacg ttgattctgc tttagccgat 4740

aatgttaaat ccttcgctgc cggtggtttt ggtggtattt gtgccgtgtt gaccggtcat 4800

ccattcgact tggtcaaggt cagattgcaa actgggttgt acaaatcgtc agtgcaatgt 4860

gtcaaagaaa caatagccaa agacgggttg tttgggttat acagaggtgt cttgcctcca 4920

ttgttgggtg tcaccccaat gtttgctgtt tccttctggg gttacgacgt tggtaagaaa 4980

ttggtttctt ccttcactgg caaatcagtc gacaagtttg aaatcaaaga catttcgact 5040

gccggtttca tcagtgccat cccaaccacc ttggttgccg ctccattcga aagagtcaag 5100

gtcatgatgc agatccaaga aggtgccaag tccaagtcca tgggtgccgt tgttgctgaa 5160

atgtacagaa ccggtggtat cagatccatc ttcaagggta ctgtggctac tttggccaga 5220

gatggtccag gttctgcctt gtatttcgct acttatgaat gggtcaagaa ggaattgact 5280

gcccctggtg aagacttgtc cttgtttgct attaccactg ctggtggttt cgctggtatt 5340

gccatgtggt tgggtgtttt cccaattgac accatcaagt ctacccagca gtcttccaac 5400

gtcaaggtct ccattgtgca ggccactaag aacatctacg ctaaaggtgg tatcaaggct 5460

ttcttccctg gtgttggtcc tgccttggcc agagctttcc cagccaacgc tgctaccttc 5520

ctaggggttg aattggccag aaagttcttg gacaaagtta tataagagtg actcttttga 5580

taagagtcgc aaatttgatt tcataagtat atattcatta tgtaaagtag taaatggaaa 5640

attcattaaa aaaaaagcaa atttccgttg tatgcatact ccgaacacaa aactagcccc 5700

ggaaaaaccc ttagttgata gttgcgaatt taggtcgacc ttggacgcgt gcggccgcca 5760

gtgtgatgga tatctgcaga attcgcccct cccgaaatat tacaattgga gctccaataa 5820

atattgtaat aaataggtct atatacatac actaagcttc taggacgtca ttgtagtctt 5880

cgaagttgtc tgctagttta gttctcatga tttcgaaaac caataacgca atggatgtag 5940

cagggatggt ggttagtgcg ttcctgacaa acccagagta cgccgcctca aaccacgtca 6000

```
cattcgccct ttgcttcatc cgcatcactt gcttgaaggt atccacgtac gagttgtaat     6060

acaccttgaa gaacggcttc gtctcgacgg gtacaacgag aattgtattg aattgatcaa     6120

gaacatgatc ttggtgttac agaacatcaa gttcttggac cagactgaga atgcacagat     6180

atacaaggcg tcatgtgata aaatggatga gatttatcca caattgaaga aagagtttat     6240

ggaaagtggt caaccagaag ctaaacagga agaagcaaac gaagaggtga aacaagaaga     6300

agaaggtaaa taagtatttt gtattatata acaaacaaag taaggaatac agatttatac     6360

aataaattgc catactagtc acgtgagata tctcatccat tccccaactc ccaagaaaaa     6420

aaaaaagtga aaaaaaaaat caaacccaaa gatcaacctc cccatcatca tcgtcatcaa     6480

acccccagct caattcgcaa tggttagcac aaaaacatac acagaaaggg catcagcaca     6540

cccctccaag gttgcccaac gtttattccg cttaatggag tccaaaaaga ccaacctctg     6600

cgcctcgatc gacgtgacca caaccgccga gttcctttcg ctcatcgaca agctcggtcc     6660

ccacatctgt ctcgtgaaga cgcacatcga tatcatctca gacttcagct acgagggcac     6720

gattgagccg ttgcttgtgc ttgcagagcg ccacgggttc ttgatattcg aggacaggaa     6780

gtttgctgat atcggaaaca ccgtgatgtt gcagtacacc tcgggggtat accggatcgc     6840

ggcgtggagt gacatcacga acgcgcacgg agtgactggg aagggcgtcg ttgaagggtt     6900

gaaacgcggt gcggaggggg tagaaaagga aaggggcgtg ttgatgttgg cggagttgtc     6960

gagtaaaggc tcgttggcgc atggtgaata tacccgtgag acgatcgaga ttgcgaagag     7020

tgatcgggag ttcgtgattg ggttcatcgc gcagcgggac atgggggtga gagaagaagg     7080

gtttgattgg atcatcatga cgcctggtgt ggggttggat gataaaggcg atgcgttggg     7140

ccagcagtat aggactgttg atgaggtggt tctgactggt accgatgtga ttattgtcgg     7200

gagagggttg tttggaaaag gaagagaccc tgaggtggag ggaaagagat acagggatgc     7260

tggatggaag gcatacttga agagaactgg tcagttagaa taaatattgt aataaatagg     7320

tctatataca tacactaagc ttctaggacg tcattgtagt cttcgaagtt gtctgctagt     7380

ttagttctca tgatttcgaa aaccaataac gcaatggatg tagcagggat ggtggttagt     7440

gcgttcctga caaacccaga gtacgccgcc tcaaaccacg tcacattcgc cctttgcttc     7500

atccgcatca cttgcttgaa ggtatccacg tacgagttgt aatacacaca ctagttggta     7560

gttctgggga ctaccggcag caattccatc cgtgcgttat ctataatccg gatcttgtga     7620

aaaaaaagga aaaattctat atatagaaaa gaaattagat agaagcctgg cgagtgtgtt     7680

tctagaagac ttccggattc agatgtcgca tggctgcgat atcacaaaaa aaaaagagag     7740

actccaacat tacacaaaac gtggtcttcc gctcccatat ctgattgcgg attttagagt     7800

gacactcttc cctcgttgat tatcaatttc gcaactgtag gatgggctgg aggtattgtt     7860

gatgcatgca ccacttatca ttgtcggatc cactaaccgc gtataatacc cacacacaca     7920
```

cacacacata cacaaacgct tcgtcacatc tctatcggta tgtatgggga gagatagaac 7980

aggaaagtca tgaggccggc caaggg 8006

<210> 123
<211> 7743
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 123
taaacgttgg gcaaccttgg aggggtgtgc tgatgccctt tctgtgtatg tttttgtgct 60

aaccattgcg aattgagctg ggggtttgat gacgatgatg atggggaggt tgatctttgg 120

gtttgatttt ttttttcact ttttttttttt cttgggagtt ggggaatgga tgagatatct 180

cacgtgacta gtatggcaat ttattgtata aatctgtatt ccttactttg tttgttatat 240

aatacaaaat acttatttac cttcttcttc ttgtttcacc tcttcgtttg cttcttcctg 300

tttagcttct ggttgaccac tttccataaa ctctttcttc aattgtggat aaatctcatc 360

cattttatca catgacgcct tgtatatctg tgcattctca gtctggtcca agaacttgat 420

gttctgtaac accaagatca tgttcttgat caattcaata caattctcgt tgtacccgtc 480

gcatatggtc gacctaaatt cgcaactatc aactaagggt ttttccgggg ctagttttgt 540

gttcggagta tgcatacaac ggaaatttgc tttttttta atgaattttc catttactac 600

tttacataat gaatatatac ttatgaaatc aaatttgcga ctcttatcaa aagagtcact 660

ctcttctatt cttaactgga caagatttca gcagcttctt cggacttttc acctctttct 720

ctgacttcaa ggtctggacg gttcaacatg tcttccaaag ccttggagta tggagccttg 780

gtgttttctg gtgggttcaa agccttgaca gtttcgaagt agtgttcgta gacgttacca 840

tcatatctac caatagcagc attggtcatc atgtcagtca agttgaaacc atcagtcaaa 900

ccaacaacgt ttggtctaac cttggcacac aaggctggga agtgctcgga ggaaatcttt 960

ctgctgacat ctggggtaat gataccgtac tgcaagaagg tggcaccaaa gttggtcaag 1020

atccagtcgg catacaagaa acccaaaggt ctcaagactt cagaccattc tggggtgata 1080

cgcttgaaga aaccgtcaat caagaatctg tgagcgttca acttggaaac caaaaccaaa 1140

cttggaccgg taatgtcgaa acccttgtct tcagcaacag caccaatgtc tctggccaag 1200

ccggaaaggg cgtggtcgaa agccttaacc aacttggctg ggtcgtccaa gtcggcaacg 1260

ctggagagaa ccaatccctt ttgttctggc tccttcaaca agtctctaac cattggcttg 1320

gcaacgttca tggccaagat gttgttgtca ccttcccagg tacattggac aacccagtcg 1380

gagtaggctt taccgaaacc gttgtaagac gagtaaccgt gaccaccaca agcttgacga 1440

gcttcgtcaa tggcttcagc agtcaaccag gtacaggtgg acttcaaaca accggaagcg 1500

acgaacaatt tcttggattc gacaatggca gcgtcaatgg cttccttttc accagcactg 1560

acagcttcgt ccaacttgtc gttggttgcg ttcatggttt gttctaagta caaggcacct 1620

tgggagaaca agtaagcggc agccaagaat gggaacaatc tcttttgatg caatgggtag 1680

tcaatcaact tggtttcaat ggtatcggtg tccttcttct tgaattgtct tctgtggatg 1740

gcgtatctca aggcaatggt gatgaatcta ctggtcattc tgtaggagtc catcatcatg 1800

gtgactctac caccaatcaa agccgagtaa gacaattgtt cagatggtgg catggtgact 1860

tcacccagac gggaaacctt acagtacttt tgcaacataa agaatcttgg gattctgacg 1920

ttggagaact ggatccaacc gttatcgata ccgtctctac ccatcttggc accaatgtca 1980

ccaacagtca cacctggctc gaggtcgtgg ttggagtctc tcaatgggac aacaaaggtc 2040

ttgacaccgt agtcctttcc tttgaccttc aatctggcgt agacggagca gtgggtggcg 2100

gagtgcgcag caccaccgat ccaccacttg gtggcaccaa tgtgtggggt gttgatgaca 2160

aactcgtcag tgtcttcgtc aaaagtggcg gtggtttcaa taccttggac gttggaaccg 2220

tgggccaact cagtcatacc aaaacaacca tagatacctc tcaacttgtc gataccctta 2280

tttatggtcc agtagaagaa ctgggagttg gtaccgttac cacggacaca ggaaacaaac 2340

aaacccaagt tgacaccgat tctggttccg acttgtgggt caaagacacc gaggatcgac 2400

aatctctggg cctgttggtc tgggtactcg tgctcaaagt atctggagag tctggcaatc 2460

ttcttggcgg tgacttctct ttgttggtct ttggtcaagt tgtagtatga agcgtcgacc 2520

ttcaagatag ggtctttttg catttgttcg accatgttgc tgatgatctc ggatctttct 2580

tgggaacctt ccaagaagta gttcaactcc tttgggttga acttggtgag ttctctttct 2640

ttttgaagtt cggtaggcat tatgtcgtga agatttgaac aatgttagtg gggaataact 2700

attcaaaaaa aagggatggt aaataaatat aacaaaaaaa aaggaaaagc agaagagaag 2760

aagttgtttt tagaatctgg ggagatttta gggaaatgga gttgatttta tatcaatggt 2820

ggaaaaagtc acaggagggg gaagcacaga gaaaaaaaag aagaggcaag cgtgtatcgg 2880

tggacaaaaa agggaaatgt taatcggtgt gtgtgcaaaa attcacaagt gtaattttga 2940

ttatctttcg cgggggggcca gccagtagct aagacacaac tcagacaatt tttactccac 3000

aattcaaccg aataaccgtg tgtatgcata acacctaagt tgcatacaac tagcgagcta 3060

cagacattgc cagagtcgat cagtagcatt tgcaacaacc agattataat tctaccagaa 3120

ccatctctcc ctctctctct ctgttgtttt accccaacga tatttctccc gaaatattac 3180

aattggagct cctgcaggca tgatgatctg ccatgccgaa ctagacgaag ccgttcttca 3240

aggtgtatta caactcgtac gtggatacct tcaagcaagt gatgcggatg aagcaaaggg 3300

```
cgaatgtgac gtggtttgag gcggcgtact ctgggtttgt caggaacgca ctaaccacca     3360

tccctgctac atccattgcg ttattggttt tcgaaatcat gagaactaaa ctagcagaca     3420

acttcgaaga ctacaatgac gtcctagaag cttagtgtat gtatatagac ctatttatta     3480

caatatttat tctaactgac cagttctctt caagtatgcc ttccatccag catccctgta     3540

tctctttccc tccacctcag ggtctcttcc ttttccaaac aaccctctcc cgacaataat     3600

cacatcggta ccagtcagaa ccacctcatc aacagtccta tactgctggc ccaacgcatc     3660

gcctttatca tccaacccca caccaggcgt catgatgatc caatcaaacc cttcttctct     3720

acccccccatg tcccgctgcg cgatgaaccc aatcacgaac tcccgatcac tcttcgcaat     3780

ctcgatcgtc tcacgggtat attcaccatg cgccaacgag cctttactcg acaactccgc     3840

caacatcaac acgccccttt ccttttctac cccctccgca ccgcgtttca acccttcaac     3900

gacgcccttc ccagtcactc cgtgcgcgtt cgtgatgtca ctccacgccg cgatccggta     3960

taccccccgag gtgtactgca acatcacggt gtttccgata tcagcaaact tcctgtcctc     4020

gaatatcaag aacccgtggc gctctgcaag cacaagcaac ggctcaatcg tgccctcgta     4080

gctgaagtct gagatgatat cgatgtgcgt cttcacgaga cagatgtggg gaccgagctt     4140

gtcgatgagc gaaaggaact cggcggttgt ggtcacgtcg atcgaggcgc agaggttggt     4200

ctttttggac tccattaagc ggaaaagggc gaattctgca gatatccatc acactggcgg     4260

ccgctcgagc atgcatctag agggcccaat tcgccctata gtgagtcgta ttacaattca     4320

ctggccgtcg ttttacaacg tcgtgactgg gaaaaccctg gcgttaccca acttaatcgc     4380

cttgcagcac atccccctttt cgccagctgg cgtaatagcg aagaggcccg caccgatcgc     4440

ccttcccaac agttgcgcag cctatacgta cggcagttta aggtttacac ctataaaaga     4500

gagagccgtt atcgtctgtt tgtggatgta cagagtgata ttattgacac gccggggcga     4560

cggatggtga tccccctggc cagtgcacgt ctgctgtcag ataaagtctc ccgtgaactt     4620

tacccggtgg tgcatatcgg ggatgaaagc tggcgcatga tgaccaccga tatggccagt     4680

gtgccggtct ccgttatcgg ggaagaagtg gctgatctca gccaccgcga aaatgacatc     4740

aaaaacgcca ttaacctgat gttctgggga atataaatgt caggcatgag attatcaaaa     4800

aggatcttca cctagatcct tttcacgtag aaagccagtc cgcagaaacg gtgctgaccc     4860

cggatgaatg tcagctactg ggctatctgg acaagggaaa acgcaagcgc aaagagaaag     4920

caggtagctt gcagtgggct tacatggcga tagctagact gggcggtttt atggacagca     4980

agcgaaccgg aattgccagc tggggcgccc tctggtaagg ttgggaagcc ctgcaaagta     5040

aactggatgg ctttctcgcc gccaaggatc tgatggcgca ggggatcaag ctctgatcaa     5100

gagacaggat gaggatcgtt tcgcatgatt gaacaagatg gattgcacgc aggttctccg     5160

gccgcttggg tggagaggct attcggctat gactgggcac aacagacaat cggctgctct     5220
```

gatgccgccg tgttccggct gtcagcgcag gggcgcccgg ttctttttgt caagaccgac 5280
ctgtccggtg ccctgaatga actgcaagac gaggcagcgc ggctatcgtg gctggccacg 5340
acgggcgttc cttgcgcagc tgtgctcgac gttgtcactg aagcgggaag ggactggctg 5400
ctattgggcg aagtgccggg gcaggatctc ctgtcatctc accttgctcc tgccgagaaa 5460
gtatccatca tggctgatgc aatgcggcgg ctgcatacgc ttgatccggc tacctgccca 5520
ttcgaccacc aagcgaaaca tcgcatcgag cgagcacgta ctcggatgga agccggtctt 5580
gtcgatcagg atgatctgga cgaagagcat caggggctcg cgccagccga actgttcgcc 5640
aggctcaagg cgagcatgcc cgacggcgag gatctcgtcg tgacccatgg cgatgcctgc 5700
ttgccgaata tcatggtgga aaatggccgc ttttctggat tcatcgactg tggccggctg 5760
ggtgtggcgg accgctatca ggacatagcg ttggctaccc gtgatattgc tgaagagctt 5820
ggcggcgaat gggctgaccg cttcctcgtg ctttacggta tcgccgctcc cgattcgcag 5880
cgcatcgcct tctatcgcct tcttgacgag ttcttctgaa ttattaacgc ttacaatttc 5940
ctgatgcggt attttctcct tacgcatctg tgcggtattt cacaccgcat acaggtggca 6000
cttttcgggg aaatgtgcgc ggaacccctа tttgtttatt tttctaaata cattcaaata 6060
tgtatccgct catgagacaa taaccctgat aaatgcttca ataatagcac gtgaggaggg 6120
ccaccatggc caagttgacc agtgccgttc cggtgctcac cgcgcgcgac gtcgccggag 6180
cggtcgagtt ctggaccgac cggctcgggt tctcccggga cttcgtggag gacgacttcg 6240
ccggtgtggt ccgggacgac gtgaccctgt tcatcagcgc ggtccaggac caggtggtgc 6300
cggacaacac cctggcctgg gtgtgggtgc gcggcctgga cgagctgtac gccgagtggt 6360
cggaggtcgt gtccacgaac ttccgggacg cctccgggcc ggccatgacc gagatcggcg 6420
agcagccgtg ggggcgggag ttcgccctgc gcgacccggc cggcaactgc gtgcacttcg 6480
tggccgagga gcaggactga cacgtgctaa aacttcattt ttaatttaaa aggatctagg 6540
tgaagatcct ttttgataat ctcatgacca aaatccctta acgtgagttt tcgttccact 6600
gagcgtcaga ccccgtagaa aagatcaaag gatcttcttg agatcctttt tttctgcgcg 6660
taatctgctg cttgcaaaca aaaaaaccac cgctaccagc ggtggtttgt ttgccggatc 6720
aagagctacc aactcttttt ccgaaggtaa ctggcttcag cagagcgcag ataccaaata 6780
ctgtccttct agtgtagccg tagttaggcc accacttcaa gaactctgta gcaccgccta 6840
catacctcgc tctgctaatc ctgttaccag tggctgctgc cagtggcgat aagtcgtgtc 6900
ttaccgggtt ggactcaaga cgatagttac cggataaggc gcagcggtcg ggctgaacgg 6960
ggggttcgtg cacacagccc agcttggagc gaacgaccta caccgaactg agatacctac 7020
agcgtgagct atgagaaagc gccacgcttc ccgaagggag aaaggcggac aggtatccgg 7080

taagcggcag ggtcggaaca ggagagcgca cgagggagct tccaggggga aacgcctggt 7140

atctttatag tcctgtcggg tttcgccacc tctgacttga gcgtcgattt ttgtgatgct 7200

cgtcaggggg gcggagccta tggaaaaacg ccagcaacgc ggccttttta cggttcctgg 7260

gcttttgctg gccttttgct cacatgttct ttcctgcgtt atcccctgat tctgtggata 7320

accgtattac cgcctttgag tgagctgata ccgctcgccg cagccgaacg accgagcgca 7380

gcgagtcagt gagcgaggaa gcggaagagc gcccaatacg caaaccgcct ctccccgcgc 7440

gttggccgat tcattaatgc agctggcacg acaggtttcc cgactggaaa gcgggcagtg 7500

agcgcaacgc aattaatgtg agttagctca ctcattaggc accccaggct ttacacttta 7560

tgcttccggc tcgtatgttg tgtggaattg tgagcggata acaatttcac acaggaaaca 7620

gctatgacca tgattacgcc aagctattta ggtgacacta tagaatactc aagctatgca 7680

tcaagcttgg taccgagctc ggatccacta gtaacggccg ccagtgtgct ggaattcgcc 7740

ctt 7743

<210> 124
<211> 6826
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 124
ttccgcttaa tggagtccaa aaagaccaac ctctgcgcct cgatcgacgt gaccacaacc 60

gccgagttcc tttcgctcat cgacaagctc ggtccccaca tctgtctcgt gaagacgcac 120

atcgatatca tctcagactt cagctacgag ggcacgattg agccgttgct tgtgcttgca 180

gagcgccacg ggttcttgat attcgaggac aggaagtttg ctgatatcgg aaacaccgtg 240

atgttgcagt acacctcggg ggtataccgg atcgcggcgt ggagtgacat cacgaacgcg 300

cacggagtga ctgggaaggg cgtcgttgaa gggttgaaac gcggtgcgga gggggtagaa 360

aaggaaaggg gcgtgttgat gttggcggag ttgtcgagta aaggctcgtt ggcgcatggt 420

gaatataccc gtgagacgat cgagattgcg aagagtgatc gggagttcgt gattgggttc 480

atcgcgcagc gggacatggg gggtagagaa gaagggtttg attggatcat catgacgcct 540

ggtgtggggt tggatgataa aggcgatgcg ttgggccagc agtataggac tgttgatgag 600

gtggttctga ctggtaccga tgtgattatt gtcgggagag ggttgtttgg aaaaggaaga 660

gaccctgagg tggagggaaa gagatacagg gatgctggat ggaaggcata cttgaagaga 720

actggtcagt tagaataaat attgtaataa ataggtctat atacatacac taagcttcta 780

ggacgtcatt gtagtcttcg aagttgtctg ctagtttagt tctcatgatt tcgaaaacca 840

ataacgcaat ggatgtagca gggatggtgg ttagtgcgtt cctgacaaac ccagagtacg 900

ccgcctcaaa ccacgtcaca ttcgcccttt gcttcatccg catcacttgc ttgaaggtat 960

ccacgtacga gttgtaatac accttgaaga acggcttcgt ctagttcggc atggcagatc 1020

atcatgcctg cagaaaatca gaggctactc cggagatggc acattatcac gtgggcagtc 1080

ccatctctct gacaatggca ctgacaatgc catgtcattc tagaagtcgc tcgttccatt 1140

gctagtcgtg taggtgtctt actcagacca tatgtttggc tgtgtgtgtg gtaaggggt 1200

aagatatctc taggaggagg caactctact cgaaagtacg aaatggccga cgccaatccg 1260

acctgccatt gtgcaggacg cgggtcgaga tccggaagtg ccctccttct gaggttgttg 1320

gtttgttcca gttggcataa gaaattaata atcgtttagt caatggtaca gtattaataa 1380

ttcctgacaa aaccgtgccc tacaagcagt gtgagttccg ggggtagcac tagtggtcgt 1440

ggtggtaggg cccacagaga gtatttccgg aaccgactga tcctttgatt ccaatattcc 1500

aagatttggg tttgtggttg tgtgtaatgc atgaggttat tttgtcaacc tccttccttc 1560

tctacggttg ttgctcctcc tcctcttcct tccctttgtt catttaaata cacacacacc 1620

actccccttc ccacccacaa gaacttttt ttttttact ttcttcttct cttactttct 1680

tcttcttctt tcattacctg ttagattata tttatcaacc tcaattggtt tattatctac 1740

cacccataca acccccaccc ctgccacatc caccaacaat ggacgacgtt gattctgctt 1800

tagccgataa tgttaaatcc ttcgctgccg gtggttttgg tggtatttgt gccgtgttga 1860

ccggtcatcc attcgacttg gtcaaggtca gattgcaaac tgggttgtac aaatcgtcag 1920

tgcaatgtgt caaagaaaca atagccaaag acgggttgtt tgggttatac agaggtgtct 1980

tgcctccatt gttgggtgtc accccaatgt ttgctgtttc cttctggggt tacgacgttg 2040

gtaagaaatt ggtttcttcc ttcactggca aatcagtcga caagtttgaa atcaaagaca 2100

tttcgactgc cggtttcatc agtgccatcc caaccacctt ggttgccgct ccattcgaaa 2160

gagtcaaggt catgatgcag atccaagaag gtgccaagtc caagtccatg ggtgccgttg 2220

ttgctgaaat gtacagaacc ggtggtatca gatccatctt caagggtact gtggctactt 2280

tggccagaga tggtccaggt tctgccttgt atttcgctac ttatgaatgg gtcaagaagg 2340

aattgactgc ccctggtgaa gacttgtcct tgtttgctat taccactgct ggtggtttcg 2400

ctggtattgc catgtggttg ggtgttttcc caattgacac catcaagtct acccagcagt 2460

cttccaacgt caaggtctcc attgtgcagg ccactaagaa catctacgct aaaggtggta 2520

tcaaggcttt cttccctggt gttggtcctg ccttggccag agctttccca gccaacgctg 2580

ctaccttcct aggggttgaa ttggccagaa agttcttgga caaagttata taataagaat 2640

agaagagagt gactcttttg ataagagtcg caaatttgat ttcataagta tatattcatt 2700

atgtaaagta gtaaatggaa aattcattaa aaaaaaagca aatttccgtt gtatgcatac 2760

tccgaacaca aaactagccc cggaaaaacc cttagttgat agttgcgaat ttaggtcgac 2820
catatgcgac gggtacaacg agaattgtat tgaattgatc aagaacatga tcttggtgtt 2880
acagaacatc aagttcttgg accagactga gaatgcacag atatacaagg cgtcatgtga 2940
taaaatggat gagatttatc cacaattgaa gaaagagttt atggaaagtg gtcaaccaga 3000
agctaaacag gaagaagcaa acgaagaggt gaaacaagaa gaagaaggta aataagtatt 3060
ttgtattata taacaaacaa agtaaggaat acagatttat acaataaatt gccatactag 3120
tcacgtgaga tatctcatcc attccccaac tcccaagaaa aaaaaaaagt gaaaaaaaaa 3180
atcaaaccca aagatcaacc tccccatcat catcgtcatc aaaccccccag ctcaattcgc 3240
aatggttagc acaaaaacat acacagaaag ggcatcagca cacccctcca aggttgccca 3300
acgtttaaag ggcgaattcc agcacactgg cggccgttac tagtggatcc gagctcggta 3360
ccaagcttga tgcatagctt gagtattcta tagtgtcacc taaatagctt ggcgtaatca 3420
tggtcatagc tgtttcctgt gtgaaattgt tatccgctca caattccaca caacatacga 3480
gccggaagca taaagtgtaa agcctggggt gcctaatgag tgagctaact cacattaatt 3540
gcgttgcgct cactgcccgc tttccagtcg ggaaacctgt cgtgccagct gcattaatga 3600
atcggccaac gcgcggggag aggcggtttg cgtattgggc gctcttccgc ttcctcgctc 3660
actgactcgc tgcgctcggt cgttcggctg cggcgagcgg tatcagctca ctcaaaggcg 3720
gtaatacggt tatccacaga atcaggggat aacgcaggaa agaacatgtg agcaaaaggc 3780
cagcaaaagc ccaggaaccg taaaaaggcc gcgttgctgg cgtttttcca taggctccgc 3840
cccctgacg agcatcacaa aaatcgacgc tcaagtcaga ggtggcgaaa cccgacagga 3900
ctataaagat accaggcgtt tccccctgga agctccctcg tgcgctctcc tgttccgacc 3960
ctgccgctta ccggatacct gtccgccttt ctcccttcgg gaagcgtggc gctttctcat 4020
agctcacgct gtaggtatct cagttcggtg taggtcgttc gctccaagct gggctgtgtg 4080
cacgaacccc ccgttcagcc cgaccgctgc gccttatccg gtaactatcg tcttgagtcc 4140
aacccggtaa gacacgactt atcgccactg gcagcagcca ctggtaacag gattagcaga 4200
gcgaggtatg taggcggtgc tacagagttc ttgaagtggt ggcctaacta cggctacact 4260
agaaggacag tatttggtat ctgcgctctg ctgaagccag ttaccttcgg aaaaagagtt 4320
ggtagctctt gatccggcaa acaaaccacc gctggtagcg gtggtttttt tgtttgcaag 4380
cagcagatta cgcgcagaaa aaaaggatct caagaagatc ctttgatctt ttctacgggg 4440
tctgacgctc agtggaacga aaactcacgt taagggattt tggtcatgag attatcaaaa 4500
aggatcttca cctagatcct tttaaattaa aaatgaagtt ttagcacgtg tcagtcctgc 4560
tcctcggcca cgaagtgcac gcagttgccg gccgggtcgc gcagggcgaa ctcccgcccc 4620

cacggctgct cgccgatctc ggtcatggcc ggcccggagg cgtcccggaa gttcgtggac 4680

acgacctccg accactcggc gtacagctcg tccaggccgc gcacccacac ccaggccagg 4740

gtgttgtccg gcaccacctg gtcctggacc gcgctgatga acagggtcac gtcgtcccgg 4800

accacaccgg cgaagtcgtc ctccacgaag tcccgggaga acccgagccg gtcggtccag 4860

aactcgaccg ctccggcgac gtcgcgcgcg gtgagcaccg gaacggcact ggtcaacttg 4920

gccatggtgg ccctcctcac gtgctattat tgaagcattt atcagggtta ttgtctcatg 4980

agcggataca tatttgaatg tatttagaaa aataaacaaa taggggttcc gcgcacattt 5040

ccccgaaaag tgccacctgt atgcggtgtg aaataccgca cagatgcgta aggagaaaat 5100

accgcatcag gaaattgtaa gcgttaataa ttcagaagaa ctcgtcaaga aggcgataga 5160

aggcgatgcg ctgcgaatcg ggagcggcga taccgtaaag cacgaggaag cggtcagccc 5220

attcgccgcc aagctcttca gcaatatcac gggtagccaa cgctatgtcc tgatagcggt 5280

ccgccacacc cagccggcca cagtcgatga atccagaaaa gcggccattt tccaccatga 5340

tattcggcaa gcaggcatcg ccatgggtca cgacgagatc ctcgccgtcg ggcatgctcg 5400

ccttgagcct ggcgaacagt tcggctggcg cgagcccctg atgctcttcg tccagatcat 5460

cctgatcgac aagaccggct tccatccgag tacgtgctcg ctcgatgcga tgtttcgctt 5520

ggtggtcgaa tgggcaggta gccggatcaa gcgtatgcag ccgccgcatt gcatcagcca 5580

tgatggatac tttctcggca ggagcaaggt gagatgacag gagatcctgc cccggcactt 5640

cgcccaatag cagccagtcc cttcccgctt cagtgacaac gtcgagcaca gctgcgcaag 5700

gaacgcccgt cgtggccagc cacgatagcc gcgctgcctc gtcttgcagt tcattcaggg 5760

caccggacag gtcggtcttg acaaaaagaa ccgggcgccc ctgcgctgac agccggaaca 5820

cggcggcatc agagcagccg attgtctgtt gtgcccagtc atagccgaat agcctctcca 5880

cccaagcggc cggagaacct gcgtgcaatc catcttgttc aatcatgcga aacgatcctc 5940

atcctgtctc ttgatcagag cttgatcccc tgcgccatca gatccttggc ggcgagaaag 6000

ccatccagtt tactttgcag ggcttcccaa ccttaccaga gggcgcccca gctggcaatt 6060

ccggttcgct tgctgtccat aaaaccgccc agtctagcta tcgccatgta agcccactgc 6120

aagctacctg ctttctcttt gcgcttgcgt tttcccttgt ccagatagcc cagtagctga 6180

cattcatccg gggtcagcac cgtttctgcg gactggcttt ctacgtgaaa aggatctagg 6240

tgaagatcct ttttgataat ctcatgcctg acatttatat tccccagaac atcaggttaa 6300

tggcgttttt gatgtcattt tcgcggtggc tgagatcagc cacttcttcc ccgataacgg 6360

agaccggcac actggccata tcggtggtca tcatgcgcca gctttcatcc ccgatatgca 6420

ccaccgggta aagttcacgg gagactttat ctgacagcag acgtgcactg gccaggggga 6480

tcaccatccg tcgccccggc gtgtcaataa tatcactctg tacatccaca aacagacgat 6540

aacggctctc tcttttatag gtgtaaacct taaactgccg tacgtatagg ctgcgcaact 6600

gttgggaagg gcgatcggtg cgggcctctt cgctattacg ccagctggcg aaaggggat 6660

gtgctgcaag gcgattaagt tgggtaacgc cagggttttc ccagtcacga cgttgtaaaa 6720

cgacggccag tgaattgtaa tacgactcac tatagggcga attgggccct ctagatgcat 6780

gctcgagcgg ccgccagtgt gatggatatc tgcagaattc gcccttt 6826

<210> 125
<211> 446
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 125
tcatgacttt cctgttctat ctctccccat acataccgat agagatgtga cgaagcgttt 60

gtgtatgtgt gtgtgtgtgt gtgggtatta tacgcggtta gtggatccga caatgataag 120

tggtgcatgc atcaacaata cctccagccc atcctacagt tgcgaaattg ataatcaacg 180

agggaagagt gtcactctaa aatccgcaat cagatatggg agcggaagac cacgttttgt 240

gtaatgttgg agtctctctt ttttttttgt gatatcgcag ccatgcgaca tctgaatccg 300

gaagtcttct agaaacacac tcgccaggct tctatctaat ttcttttcta tatatagaat 360

ttttcctttt ttttcacaag atccggatta tagataacgc acggatggaa ttgctgccgg 420

tagtccccag aactaccaac tagtgt 446

<210> 126
<211> 500
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 126
ggatcaagcg acgaaccggc cggcaaccag gggggcgttc cgaagaagag gactccgtgt 60

ttccgaagtc ctaccaaccc cccagtaaga aaacaccccc taagacatat ttccgtcaaa 120

gaggtctctt ttccgctgca tgggccgcat gctcgcgggc aggtgtgtca acgagagctg 180

gagtttgtga tggtgctggt gctggccata ggcaatagct acgacagtca aaaaaaaact 240

atccgacctg ccgtcctcta gtttactcct cagccagaga agagggccaa gtttttttctt 300

ctctccggta tgcaacatta cttgagatgt atgattaccc agctcacttc tgttccgcca 360

ggacttccgc ctctcaaccc ccctagattt gaacgaaaga aaaacccctc tatgagaaac 420

gccctctctt tttttggcct gttgtttatc atgtaactct aaccaccctt gaggtatttt 480

cggagcactc tcgagcatgg 500

<210> 127
<211> 9522
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<220>
<221> modified_base
<222> (3519)..(3519)
<223> a, c, t, g, unknown or other

<220>
<221> modified_base
<222> (3663)..(3663)
<223> a, c, t, g, unknown or other

<220>
<221> modified_base
<222> (3919)..(3919)
<223> a, c, t, g, unknown or other

<400> 127
tttccatagg ctccgccccc ctgacgagca tcacaaaaat cgacgctcaa gtcagaggtg 60

gcgaaacccg acaggactat aaagatacca ggcgtttccc cctggaagct ccctcgtgcg 120

ctctcctgtt ccgaccctgc cgcttaccgg atacctgtcc gcctttctcc cttcgggaag 180

cgtggcgctt tctcatagct cacgctgtag gtatctcagt tcggtgtagg tcgttcgctc 240

caagctgggc tgtgtgcacg aaccccccgt tcagcccgac cgctgcgcct tatccggtaa 300

ctatcgtctt gagtccaacc cggtaagaca cgacttatcg ccactggcag cagccactgg 360

taacaggatt agcagagcga ggtatgtagg cggtgctaca gagttcttga agtggtggcc 420

taactacggc tacactagaa ggacagtatt tggtatctgc gctctgctga agccagttac 480

cttcggaaaa agagttggta gctcttgatc cggcaaacaa accaccgctg gtagcggtgg 540

ttttttttgtt tgcaagcagc agattacgcg cagaaaaaaa ggatctcaag aagatccttt 600

gatcttttct acggggtctg acgctcagtg gaacgaaaac tcacgttaag ggattttggt 660

catgagatta tcaaaaagga tcttcaccta gatcctttta aattaaaaat gaagttttaa 720

atcaatctaa agtatatatg agtaaacttg gtctgacagt taccaatgct taatcagtga 780

ggcacctatc tcagcgatct gtctatttcg ttcatccata gttgcctgac tccccgtcgt 840

gtagataact acgatacggg agggcttacc atctggcccc agtgctgcaa tgataccgcg 900

agacccacgc tcaccggctc cagatttatc agcaataaac cagccagccg gaagggccga 960

gcgcagaagt ggtcctgcaa ctttatccgc ctccatccag tctattaatt gttgccggga 1020

agctagagta agtagttcgc cagttaatag tttgcgcaac gttgttgcca ttgctacagg 1080

catcgtggtg tcacgctcgt cgtttggtat ggcttcattc agctccggtt cccaacgatc 1140

aaggcgagtt acatgatccc ccatgttgtg caaaaaagcg gttagctcct tcggtcctcc 1200

gatcgttgtc agaagtaagt tggccgcagt gttatcactc atggttatgg cagcactgca 1260

taattctctt actgtcatgc catccgtaag atgcttttct gtgactggtg agtactcaac 1320

caagtcattc tgagaatagt gtatgcggcg accgagttgc tcttgcccgg cgtcaatacg 1380

ggataatacc gcgccacata gcagaacttt aaaagtgctc atcattggaa aacgttcttc 1440

ggggcgaaaa ctctcaagga tcttaccgct gttgagatcc agttcgatgt aacccactcg 1500

tgcacccaac tgatcttcag catcttttac tttcaccagc gtttctgggt gagcaaaaac 1560

aggaaggcaa aatgccgcaa aaaagggaat aagggcgaca cggaaatgtt gaatactcat 1620

actcttcctt tttcaatatt attgaagcat ttatcagggt tattgtctca tgagcggata 1680

catatttgaa tgtatttaga aaaataaaca aataggggtt ccgcgcacat ttccccgaaa 1740

agtgccacct gacgtctaag aaaccattat tatcatgaca ttaacctata aaaataggcg 1800

tatcacgagg ccctttcgtc tcgcgcgttt cggtgatgac ggtgaaaacc tctgacacat 1860

gcagctcccg gagacggtca cagcttgtct gtaagcggat gccgggagca gacaagcccg 1920

tcagggcgcg tcagcgggtg ttggcgggtg tcggggctgg cttaactatg cggcatcaga 1980

gcagattgta ctgagagtgc accatatgcg gtgtgaaata ccgcacagat gcgtaaggag 2040

aaaataccgc atcaggcgcc attcgccatt caggctgcgc aactgttggg aagggcgatc 2100

ggtgcgggcc tcttcgctat tacgccagct ggcgaaaggg ggatgtgctg caaggcgatt 2160

aagttgggta acgccagggt tttcccagtc acgacgttgt aaaacgacgg ccagtgaatt 2220

ccggttgaat tgtggagtaa aaattgtctg agttgtgtct tagctactgg ctggcccccc 2280

gcgaaagata atcaaaatta cacttgtgaa tttttgcaca cacaccgatt aacatttccc 2340

ttttttgtcc accgatacac gcttgcctct tctttttttt ctctgtgctt ccccctcctg 2400

tgacttttc caccattgat ataaaatcaa ctccatttcc ctaaaatctc cccagattct 2460

aaaaacaact tcttctcttc tgcttttcct tttttttgt tatatttatt taccatccct 2520

tttttttgaa tagttattcc ccactaacat tgttcaaatc ttcacgacat aatgactttt 2580

acaaagaaaa acgttagtgt atcacaaggt cctgacccta gatcatccat ccaaaaggaa 2640

agagacagct ccaaatggaa ccctcaacaa atgaactact tcttggaagg ctccgtcgaa 2700

agaagtgagt tgatgaaggc tttggcccaa caaatggaaa gagacccaat cttgttcaca 2760

gacggctcct actacgactt gaccaaggac caacaaagag aattgaccgc cgtcaagatc 2820

aacagaatcg ccagatacag agaacaagaa tccatcgaca ctttcaacaa gagattgtcc 2880

ttgattggta tctttgaccc acaggtcggt accagaattg gtgtcaacct cggtttgttc 2940

ctttcttgta tcagaggtaa cggtaccact tcccaattga actactgggc taacgaaaag 3000

gaaaccgctg acgttaaagg tatctacggt tgtttcggta tgacccccgg ggatctgacg 3060

ggtacaacga gaattgtatt gaattgatca agaacatgat cttggtgtta cagaacatca 3120

agttcttgga ccagactgag aatgcacaga tatacaaggc gtcatgtgat aaaatggatg 3180

agatttatcc acaattgaag aaagagttta tggaaagtgg tcaaccagaa gctaaacagg 3240

aagaagcaaa cgaagaggtg aaacaagaag aagaaggtaa ataagtattt tgtattatat 3300

aacaaacaaa gtaaggaata cagatttata caataaattg ccatactagt cacgtgagat 3360

atctcatcca ttccccaact cccaagaaaa aaaaaaagtg aaaaaaaaaa tcaaacccaa 3420

agatcaacct ccccatcatc atcgtcatca aacccccagc tcaattcgca atggttagca 3480

caaaaacata cacagaaagg gcatcagcac acccctccna ggttgcccaa cgtttattcc 3540

gcttaatgga gtccaaaaag accaacctct gcgcctcgat cgacgtgacc acaaccgccg 3600

agttcctttc gctcatcgac aagctcggtc cccacatctg tctcgtgaag acgcacatcg 3660

atntcatctc agacttcagc tacgagggca cgattgagcc gttgcttgtg cttgcagagc 3720

gccacgggtt cttgatattc gaggacagga agtttgctga tatcggaaac accgtgatgt 3780

tgcagtacac ctcgggggta taccggatcg cggcgtggag tgacatcacg aacgcgcacg 3840

gagtgactgg gaagggcgtc gttgaagggt tgaaacgcgg tgcggagggg gtagaaaagg 3900

aaaggggcgt gttgatgtng gcggagttgt cgagtaaagg ctcgttggcg catggtgaat 3960

atacccgtga gacgatcgag attgcgaaga gtgatcggga gttcgtgatt gggttcatcg 4020

cgcagcggga catggggggt agagaagaag ggtttgattg gatcatcatg acgcctggtg 4080

tggggttgga tgataaaggc gatgcgttgg gccagcagta taggactgtt gatgaggtgg 4140

ttctgactgg taccgatgtg attattgtcg ggagagggtt gtttggaaaa ggaagagacc 4200

ctgaggtgga gggaaagaga tacagggatg ctggatggaa ggcatacttg aagagaactg 4260

gtcagttaga ataaatattg taataaatag gtctatatac atacactaag cttctaggac 4320

gtcattgtag tcttcgaagt tgtctgctag tttagttctc atgatttcga aaaccaataa 4380

cgcaatggat gtagcaggga tggtggttag tgcgttcctg acaaacccag agtacgccgc 4440

ctcaaaccac gtcacattcg ccctttgctt catccgcatc acttgcttga aggtatccac 4500

gtacgagttg taatacacct tgaagaacgg cttcgtctac ggtcgacgac gggtacaacg 4560

agaattgtat tgaattgatc aagaacatga tcttggtgtt acagaacatc aagttcttgg 4620

accagactga gaatgcacag atatacaagg cgtcatgtga taaaatggat gagatttatc 4680

cacaattgaa gaaagagttt atggaaagtg gtcaaccaga agctaaacag gaagaagcaa 4740

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 267 of 3077

acgaagaggt gaaacaagaa gaagaaggta aataagtatt ttgtattata taacaaacaa 4800

agtaaggaat acagatttat acaataaatt gccatactag tcacgtgaga tatctcatcc 4860

attccccaac tcccaagaaa aaaaaaaagt gaaaaaaaaa atcaaaccca aagatcaacc 4920

tccccatcat catcgtcatc aaacccccag ctcaattcgc agagctcggt acccggggat 4980

cctctagagt cgacctgcag gcatgctaat gtttatatcg ttgacggtga aaaatttaa 5040

cttacttgta atatttacga tttggacttg aagtccaaaa tgaagtataa atattggttg 5100

aaaattgaaa atgatttaca agccatatac aatagttcca atccaaattg gagatattgt 5160

actaaatatc caggcagaac caacaacaac aatgacgaca aagtgcaagc caaagagtag 5220

aaatcattaa tctttcatat tgtcattttt tttattatta ttcagtacat atatacatat 5280

gtgtcaacta tatatttcat gaaatataca tacgtcaaat aaaacatatc ccaatcacta 5340

aaaaaaaaaa aaatattaac caaaactaat cttaacttta attaaatgtt catattttcc 5400

ttaccagaca tcttcttggt atctattttg aactttccaa gatttaacca attcagcagc 5460

tttatcttcg tggatatctc tacttttggc aacaatcttg gcaattgcag cttgaacatc 5520

tctagccatt ctactggcat caccacaaac ataaataatt gctccactag ataataactc 5580

atcaacaaga gcactatttt ctaaaatctt atcttgaaca taaactttct tagttggatc 5640

ttgtcttgag aaggcattaa acatttcgaa attttctccc aatactgagg catattcact 5700

ccattcttgt ttgtacaaga aatcttgttc ggaatttcta catccataaa acaatacagt 5760

cttaccaaca ttaacaccat ttttaacttg ttgaactctt tctctaacaa aacctctcaa 5820

tggtgcaaca ccggtacctg gaccaatcaa aataactggg gtagtgctat tctttggtaa 5880

tttgaaatta gatcttctaa cgtgaactgg caatctgaac ttgctaaatt tgcctcttgg 5940

accattcaaa tcataatgaa ccattggggt ttcaccagtt ttgttttgtt caatttcaat 6000

attcttcaac aagttggtga caacaccagt aactggtctt ccatcagctt cttcttcggc 6060

ttcaacaaca gcagtaacat taatggtttg cttttcactt aatgaggaag aagaaatgga 6120

gtaataacga ggagttaagt gttggacatt ttcaataagg aattcgaatg gaacatcgga 6180

ccatggtctg ttgttggaag caaataataa agcatcggca atgttgaatt ttctacgggt 6240

gactttacta gcaaattctt gcttatcacc accaattcta gtaaatgact ttttagtttc 6300

ttcatcaggg gcaaatccag caatagataa gaaaaattgt ctagaaacag gacctgaaat 6360

ttccaaatgg tgtctaataa cagctccata agtgattgga ttagggaatg ggatggaata 6420

agtggaatcc aaagctttca attcaataac agtatcaagt ttgtcttcta aaccaaaaca 6480

tttggcaaat tgcttaatgt tttcatcaga gttagatggc cagattgcaa gatgatcacc 6540

ggtggtatat ttcaagtttg attcagaaat atcaaattca acatgaacac aatgtctgtc 6600

ttcagaagta aacaattctt tagttttaac aattctagcc aaaaatggat gagtatgatc 6660

aaatggacct ttagttaagt caacaccttc agatttaatg tatttgacat ttggttcacc 6720
caaggagaca tctggatcat tgccagataa atcatctctt tcagtcaatt taacatttgg 6780
ttcgtatttc aactcttttt cttcaaaatt caaatcattc tttaaggaat caaacacgtt 6840
atccttccag gccaagaaat cttcatctaa agtaccagta ccatcgtcac cttcaccgta 6900
ttcagcaaat ctgtcaccac ctttttctcc caacaatctg tcaaattttc taccaatagc 6960
attgaagaat tcataagttg aattacccaa accaaaaaca gtatatttca aagtactcaa 7020
ggtgtcagct tcttcagtca accaagtgtg aaattcgtca gcattatcgg ttggttcacc 7080
ttcaccataa gtagcaacaa taaagaaaac caagatatct tcagtaatat ctccgaagtt 7140
ttcgaaatca taatcagcga aatcagcaac catggttttc aaaccaaatc ttgaatgcaa 7200
ttctcttgac aatttgttgg cataatcttc tgctgtacct gtttgggatc caaataataa 7260
taacgtattt ttattgttct tcttcaaagc ttgtaagata tctcttgaat taccatctcc 7320
actatcagta ttaaggaacc cggtatcttg ttgttggtca agaaactgat tctttgcaaa 7380
ataagctgca attgcaacca ccaatgttat aataacatat aaatctaact tatctaatgc 7440
catgacttga gcgattgaaa caatagtctt agttgataga gcagtaatgg atcaagaaaa 7500
gaaaaagaat aaaaggtaga aataaaaaaa aaacaaagaa aaattggaaa tttgaaatgt 7560
ttaaaatctt tttttttatt ttttctctcg ttcagttcgt ttgaagattg tcgtatgaaa 7620
ttttttattc ttttttttttt ttgttcctct tctgatgttc aattacggga cccaacgtcg 7680
tttatgtatt acgatactgg gaccatactg actttgaaat atgtatcatt tctgacagtc 7740
aggtaatgtg taccaactta gaccaaattt catctctata attattattg tatatataac 7800
tgtccttttct tttcgtctat tagagtttta tctcaggagg atttatcggt gcaatttcga 7860
tatcggatca tttattcact aagagctaga gtgagacaaa agcagaatgc tacaccgacc 7920
accgctacca ccaccattat ttttggggtt ttactctacg caacccccccc catcgaagca 7980
ttgaatagac atgacagctc ttccgcttcc tcgctcactg actcgctgcg ctcggtcgtt 8040
cggctgcggc gagcggtatc agctcactca aaggcggtaa tacggttatc cacagaatca 8100
ggggataacg caggaaagaa catggacgct gccgttgaaa agaacgacac caaggctatc 8160
tttaagtcta ttgacgacat gaagtcattg tttgttgact ctggttcctt gaagtccact 8220
gccacttggt tgggtgctga agccattgac caatgtagac aagcctgtgg tggtcacggt 8280
tactcgtcct acaacggctt cggtaaagcc tacaacgatt gggttgtcca atgtacttgg 8340
gaaggtgaca acaatgtctt ggccatgagt gttggtaagc caattgtcaa gcaagttatc 8400
agcattgaag atgccggcaa gaccgtcaga ggttccaccg ctttcttgaa ccaattgaag 8460
gactacactg gttccaacag ctccaaggtt gttttgaaca ctgttgctga cttggacgac 8520

atcaagactg tcatcaaggc tattgaagtt gccatcatca gattgtccca agaagctgct 8580

tctattgtca agaaggaatc tttcgactat gtcggcgctg aattggttca actctccaag 8640

ttgaaggctc accactactt gttgactgaa tacatcagaa gaattgacac ctttgaccaa 8700

aaggacttgg ttccatactt gatcaccctc ggtaagttgt acgctgccac tattgtcttg 8760

gacagatttg ccggtgtctt cttgactttc aacgttgcct ccaccgaagc catcactgct 8820

ttggcctctg tgcaaattcc aaagttgtgt gctgaagtca gaccaaacgt tgttgcttac 8880

accgactcct tccaacaatc cgacatgatt gtcaattctg ctattggtag atacgatggt 8940

gacatctatg agaactactt tgacttggtc aagttgcaga acccaccatc caagaccaag 9000

gctccttact ctgatgcttt ggaagccatg ttgaacagac caaccttgga cgaaagagaa 9060

agatttgaaa agtctgatga aaccgctgct atcttgtcca agaagcttgg cgtaatcatg 9120

gtcatagctg tttcctgtgt gaaattgtta tccgctcaca attccacaca acatacgagc 9180

cggaagcata aagtgtaaag cctggggtgc ctaatgagtg agctaactca cattaattgc 9240

gttgcgctca ctgcccgctt tccagtcggg aaacctgtcg tgccagctgc attaatgaat 9300

cggccaacgc gcggggagag gcggtttgcg tattgggcgc tcttccgctt cctcgctcac 9360

tgactcgctg cgctcggtcg ttcggctgcg gcgagcggta tcagctcact caaaggcggt 9420

aatacggtta tccacagaat caggggataa cgcaggaaag aacatgtgag caaaaggcca 9480

gcaaaaggcc aggaaccgta aaaaggccgc gttgctggcg tt 9522

<210> 128
<211> 5570
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 128
ggccgcgagc actaggtttt gataatttgg ttcttacagt ttatgtattt tgattcttcc 60

ttttttagat actttttttt atattttatt attccttatt gatgtaacga cagtcccact 120

ataattaact taaactttgc tgtaaatcag atgacaagtg tttccctgtt tgcaggggag 180

ctctactagt ttccttaatt catcttggtt atgttgttct tgcaccgcaa ggataagctg 240

ggcccgaatc caggccacgc ataccaacta tcgctgcccg gaaaaaatac aaactagtaa 300

actgcaaaaa accatatata aacaataaga attttctctg tatcattctt cgaaaaacct 360

caccttctct ctctctctct ctctcaccct tcctttcccc tatcacatct tcatcttcac 420

atcaccaact caaaacccct aagcagccaa agtaaagaac agagaatgat tccacaacat 480

cacactatgt tcttcgccgt taattaagaa ttcactggcc gtcgttttac aacgtcgtga 540

```
ctgggaaaac cctggcgtta cccaacttaa tcgccttgca gcacatcccc ctttcgccag      600
ctggcgtaat agcgaagagg cccgcaccga tcgcccttcc caacagttgc gcagcctgaa      660
tggcgaatgg cgcctgatgc ggtattttct ccttacgcat ctgtgcggta tttcacaccg      720
catatggtgc actctcagta caatctgctc tgatgccgca tagttaagcc agccccgaca      780
cccgccaaca cccgctgacg cgccctgacg ggcttgtctg ctcccggcat ccgcttacag      840
acaagctgtg accgtctccg ggagctgcat gtgtcagagg ttttcaccgt catcaccgaa      900
acgcgcgaga cgaaagggcc tcgtgatacg cctatttta taggttaatg tcatgataat      960
aatggtttct tagacgtcag gtggcacttt tcggggaaat gtgcgcggaa cccctatttg     1020
tttatttttc taaatacatt caaatatgta tccgctcatg agacaataac cctgataaat     1080
gcttcaataa tattgaaaaa ggaagagtat gagtattcaa catttccgtg tcgcccttat     1140
tccctttttt gcggcatttt gccttcctgt ttttgctcac ccagaaacgc tggtgaaagt     1200
aaaagatgct gaagatcagt tgggtgcacg agtgggttac atcgaactgg atctcaacag     1260
cggtaagatc cttgagagtt ttcgccccga agaacgtttt ccaatgatga gcacttttaa     1320
agttctgcta tgtggcgcgg tattatcccg tattgacgcc gggcaagagc aactcggtcg     1380
ccgcatacac tattctcaga atgacttggt tgagtactca ccagtcacag aaaagcatct     1440
tacggatggc atgacagtaa gagaattatg cagtgctgcc ataaccatga gtgataacac     1500
tgcggccaac ttacttctga caacgatcgg aggaccgaag gagctaaccg ctttttgca      1560
caacatgggg gatcatgtaa ctcgccttga tcgttgggaa ccggagctga atgaagccat     1620
accaaacgac gagcgtgaca ccacgatgcc tgtagcaatg gcaacaacgt tgcgcaaact     1680
attaactggc gaactactta ctctagcttc ccggcaacaa ttaatagact ggatggaggc     1740
ggataaagtt gcaggaccac ttctgcgctc ggcccttccg gctggctggt ttattgctga     1800
taaatctgga gccggtgagc gtgggtctcg cggtatcatt gcagcactgg ggccagatgg     1860
taagccctcc cgtatcgtag ttatctacac gacggggagt caggcaacta tggatgaacg     1920
aaatagacag atcgctgaga taggtgcctc actgattaag cattggtaac tgtcagacca     1980
agtttactca tatatacttt agattgattt aaaacttcat ttttaattta aaaggatcta     2040
ggtgaagatc ctttttgata atctcatgac caaaatccct taacgtgagt tttcgttcca     2100
ctgagcgtca gaccccgtag aaaagatcaa aggatcttct tgagatcctt tttttctgcg     2160
cgtaatctgc tgcttgcaaa caaaaaaacc accgctacca gcggtggttt gtttgccgga     2220
tcaagagcta ccaactcttt ttccgaaggt aactggcttc agcagagcgc agataccaaa     2280
tactgtcctt ctagtgtagc cgtagttagg ccaccacttc aagaactctg tagcaccgcc     2340
tacatacctc gctctgctaa tcctgttacc agtggctgct gccagtggcg ataagtcgtg     2400
tcttaccggg ttggactcaa gacgatagtt accggataag gcgcagcggt cgggctgaac     2460
```

```
gggggggttcg tgcacacagc ccagcttgga gcgaacgacc tacaccgaac tgagatacct     2520
acagcgtgag ctatgagaaa gcgccacgct tcccgaaggg agaaaggcgg acaggtatcc     2580
ggtaagcggc agggtcggaa caggagagcg cacgagggag cttccagggg gaaacgcctg     2640
gtatctttat agtcctgtcg ggtttcgcca cctctgactt gagcgtcgat ttttgtgatg     2700
ctcgtcaggg gggcggagcc tatggaaaaa cgccagcaac gcggcctttt tacggttcct     2760
ggccttttgc tggccttttg ctcacatgtt ctttcctgcg ttatcccctg attctgtgga     2820
taaccgtatt accgccttttg agtgagctga taccgctcgc cgcagccgaa cgaccgagcg     2880
cagcgagtca gtgagcgagg aagcggaaga gcgcccaata cgcaaaccgc ctctccccgc     2940
gcgttggccg attcattaat gcagctggca cgacaggttt cccgactgga aagcgggcag     3000
tgagcgcaac gcaattaatg tgagttagct cactcattag gcaccccagg ctttacactt     3060
tatgcttccg gctcgtatgt tgtgtggaat tgtgagcgga taacaatttc acacaggaaa     3120
cagctatgac catgattacg ccaagctttt aattaaagat aatcacaggg gtagagacct     3180
tggttatggg ctgattggag gaagaaatgg ggtgatttgg cacttgacag cgcgagagtg     3240
gttaacacct ggtttccctc atttgggttc tgacatttga taagttgaaa gaacaatgca     3300
gaattcacat ggctaatttg gcctcggttc cacaacgcac tcagcattaa aaaaaaaata     3360
cgcaatggca gctcggtcga cgcagcagaa gcgccgacgt accgtcgcgt tgccccgccc     3420
atgcctcgcc gacccctcca ccgccatcgt ttgcccattg tttgtggtag tgcgccgtga     3480
cacaaaaact tgtcctgtca catgctgaag ttacaccaac ataactacta tgggattacg     3540
taatcaaaaa tttcacagtt ttaacaaaaa aaaaatcata caatcaacat tgggacatct     3600
tgccctcccc cacaaaactt gcttctgcat caatcatata taaacatcat gaaataagcc     3660
taaactcact tctttttttt tcatccttcc tacttcttct ttcatagtaa ctacttttt     3720
tttattacca cacttattca ttcataccac gctatcggcc ggccgcatgc ggatccgagc     3780
tcgcggccgc cagtgtgatg gatatctgca gcggccgcta aatattgtaa taaataggtc     3840
tatatacata cactaagctt ctaggacgtc attgtagtct tcgaagttgt ctgctagttt     3900
agttctcatg atttcgaaaa ccaataacgc aatggatgta gcagggatgg tggttagtgc     3960
gttcctgaca aacccagagt acgccgcctc aaaccacgtc acattcgccc tttgcttcat     4020
ccgcatcact tgcttgaagg tatccacgta cgagttgtaa tacaccttga agaacggctt     4080
cgtctcgacg ggtacaacga gaattgtatt gaattgatca agaacatgat cttggtgtta     4140
cagaacatca agttcttgga ccagactgag aatgcacaga tatacaaggc gtcatgtgat     4200
aaaatggatg agatttatcc acaattgaag aaagagttta tggaaagtgg tcaaccagaa     4260
gctaaacagg aagaagcaaa cgaagaggtg aaacaagaag aagaaggtaa ataagtattt     4320
```

tgtattatat aacaaacaaa gtaaggaata cagatttata caataaattg ccatacaagt 4380

cacgtgagat atctcatcca ttccccaact cccaagaaaa aaaaaaagtg aaaaaaaaaa 4440

tcaaacccaa agatcaacct ccccatcatc atcgtcatca aacccccagc tcaattcgca 4500

atggttagca caaaaacata cacagaaagg gcatcagcac acccctccaa ggttgcccaa 4560

cgtttattcc gcttaatgga gtccaaaaag accaacctct gcgcctcgat cgacgtgacc 4620

acaaccgccg agttcctttc gctcatcgac aagctcggtc cccacatctg tctcgtgaag 4680

acgcacatcg atatcatctc agacttcagc tacgagggca cgattgagcc gttgcttgtg 4740

cttgcagagc gccacgggtt cttgatattc gaggacagga agtttgctga tatcggaaac 4800

accgtgatgt tgcagtacac ctcgggggta taccggatcg cggcgtggag tgacatcacg 4860

aacgcgcacg gagtgactgg gaagggcgtc gttgaagggt tgaaacgcgg tgcggagggg 4920

gtagaaaagg aaaggggcgt gttgatgttg gcggagttgt cgagtaaagg ctcgttggcg 4980

catggtgaat atacccgtga gacgatcgag attgcgaaga gtgatcggga gttcgtgatt 5040

gggttcatcg cgcagcggga catggggggt agagaagaag ggtttgattg gatcatcatg 5100

acgcctggtg tggggttgga tgataaaggc gatgcgttgg gccagcagta taggactgtt 5160

gatgaggtgg ttctgactgg taccgatgtg attattgtcg ggagagggtt gtttggaaaa 5220

ggaagagacc ctgaggtgga gggaaagaga tacagggatg ctggatggaa ggcatacttg 5280

aagagaactg gtcagttaga ataaatattg taataaatag gtctatatac atacactaag 5340

cttctaggac gtcattgtag tcttcgaagt tgtctgctag tttagttctc atgatttcga 5400

aaaccaataa cgcaatggat gtagcaggga tggtggttag tgcgttcctg acaaacccag 5460

agtacgccgc ctcaaaccac gtcacattcg ccctttgctt catccgcatc acttgcttga 5520

aggtatccac gtacgagttg taatacacct tgaagaacgg cttcgtctgc 5570

<210> 129
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 129
cacacacata tgcgacgggt acaacgagaa tt 32

<210> 130
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 130
cacacaacgc gtagacgaag ccgttcttca ag 32

<210> 131
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 131
atgatctgcc atgccgaact c 21

<210> 132
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 132
agcgagttcg gcatggcaga tcatcatg 28

<210> 133
<211> 37
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 133
cacacagctc ttctagaatg gctttagaca agttaga 37

<210> 134
<211> 43
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 134
cacacagctc ttcgtttcca aacatcttct tggtatctat ttt 43

<210> 135
<211> 30
<212> DNA
<213> Artificial Sequence

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 274 of 3077

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 135
cacacactgc aggagctcca attgtaatat 30

<210> 136
<211> 48
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 136
cacacatcta gacccgggct cttcttatgt cgtgaagatt tgaacaat 48

<210> 137
<211> 46
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 137
gagcccgggt ctagatgtgt gctcttccga gtgactcttt tgataa 46

<210> 138
<211> 37
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 138
cacacacata tggtcgacct aaattcgcaa ctatcaa 37

<210> 139
<211> 41
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 139
cacacagctc ttccataatg actgcacagg atattatcgc c 41

<210> 140
<211> 52
<212> DNA

<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 140
cacacagctc ttccctctct tctattccta atacatctca atgttggcac cg 52

<210> 141
<211> 42
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 141
cacacagctc ttccataatg gctttagaca agttagattt gt 42

<210> 142
<211> 47
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 142
cacacagctc ttccctctct tctattccta ccaaacatct tcttggt 47

<210> 143
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 143
ttccgcttaa tggagtccaa a 21

<210> 144
<211> 40
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 144
atgatctgcc atgccgaact agacgaagcc gttcttcaag 40

<210> 145

<211> 40
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 145
cttgaagaac ggcttcgtct agttcggcat ggcagatcat 40

<210> 146
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 146
taaacgttgg gcaaccttgg 20

<210> 147
<211> 40
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 147
aaaatcagag gctactccgg agatggcaca ttatcacgtg 40

<210> 148
<211> 41
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 148
tgttggtgga tgtggcaggg gtgggggttg tatgggtggt a 41

<210> 149
<211> 59
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 149
ccggttcgtc gcttgatcca aaatcagagg ctactccgga gatggcacat tatcacgtg 59

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 277 of 3077

<210> 150
<211> 51
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 150
cgtcgtccat tgttggtgga tgtggcaggg gtgggggttg tatgggtggt a 51

<210> 151
<211> 41
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 151
agagtcactc ttatataact ttgtccaaga actttctggc c 41

<210> 152
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 152
caccaacaat ggacgacgtt gattctg 27

<210> 153
<211> 42
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 153
actggcggcc gcacgcgtcc aaggtcgacc taaattcgca ac 42

<210> 154
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 154

aagttatata agagtgactc ttttgataag ag 32

<210> 155
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 155
cttggacgcg tgcggccg 18

<210> 156
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 156
ggatcaagcg acgaaccg 18

<210> 157
<211> 43
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 157
cacacagctc ttccataatg tttaacttta agttgtcgca aca 43

<210> 158
<211> 53
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 158
cacacagctc ttccctctct tctattctta caacttaggc ttagcatcag tca 53

<210> 159
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 279 of 3077

<400> 159
ccaccaaatc cattatgcca 20

<210> 160
<211> 48
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 160
ccgaaacaac cgtagatacc tttaatgggt tacccttgat gacttctg 48

<210> 161
<211> 48
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 161
cagaagtcat caagggtaac ccattaaagg tatctacggt tgtttcgg 48

<210> 162
<211> 40
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 162
gatttcttcg tcgtgggcag ggtaccgagc tctgcgaatt 40

<210> 163
<211> 40
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 163
aattcgcaga gctcggtacc ctgcccacga cgaagaaatc 40

<210> 164
<211> 20
<212> DNA
<213> Artificial Sequence

<220>

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 280 of 3077

<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 164
agtaagccaa accgaaaccg 20

<210> 165
<211> 44
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 165
aattcgcaga gctcggtacc caacaagttg ttgtaggcac tcaa 44

<210> 166
<211> 44
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 166
ttgagtgcct acaacaactt gttgggtacc gagctctgcg aatt 44

<210> 167
<211> 47
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 167
gacgaagaat tccaaaggct tgttaaaggt atctacggtt gtttcgg 47

<210> 168
<211> 47
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 168
ccgaaacaac cgtagatacc tttaacaagc ctttggaatt cttcgtc 47

<210> 169
<211> 54
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 169
aaagaaagaa agaaactata acaatcaatc atgccaattt tgaaaaaacc attc 54

<210> 170
<211> 53
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 170
tcttatcaaa agagtcactc tcttctattc ttacaactta ggcttagcat cag 53

<210> 171
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 171
gaatagaaga gagtgactct tttg 24

<210> 172
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 172
gattgattgt tatagtttct ttctttc 27

<210> 173
<211> 42
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 173
ttgactgatg ctaagtaaga atagaagaga gtgactcttt tg 42

<210> 174
<211> 41

<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 174
tctcttctat tcttacttag catcagtcaa taatccctta g 41

<210> 175
<211> 58
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 175
aaaagaaaga aagaaactat aacaatcaat caatgtttaa ctttaagttg tcgcaaca 58

<210> 176
<211> 56
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 176
tcaaaagagt cactctcttc tattcttact tagcatcagt caataatccc ttagtc 56

<210> 177
<211> 33
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 177
gaatagaaga gagtgactct tttgataaga gtc 33

<210> 178
<211> 33
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 178
tgattgattg ttatagtttc tttctttctt ttg 33

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 283 of 3077

<210> 179
<211> 52
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 179
aaagaaagaa agaaactata acaatcaatc atgtcaactt accagttcca ag 52

<210> 180
<211> 52
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 180
tcttatcaaa agagtcactc tcttctattc ctaatactcg cttaatctca ac 52

<210> 181
<211> 33
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 181
tgattgattg ttatagtttc tttctttctt ttg 33

<210> 182
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 182
tcaacttacc agttccaaga aactttagaa 30

<210> 183
<211> 59
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 183
caaaagaaag aaagaaacta taacaatcaa tcaatgcttt cccgtaccac tttaagagt 59

<210> 184
<211> 51
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 184
ttctaaagtt tcttggaact ggtaagttga ttggtcggtc ttgctgttga a 51

<210> 185
<211> 59
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 185
caaaagaaag aaagaaacta taacaatcaa tcaatgtctg cattaagatc attccaacg 59

<210> 186
<211> 59
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 186
ttctaaagtt tcttggaact ggtaagttga ataggttctg actgagttct ttaaggtgg 59

<210> 187
<211> 59
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 187
caaaagaaag aaagaaacta taacaatcaa tcaatgttta actttaagtt gtcgcaaca 59

<210> 188
<211> 48
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 188
ttctaaagtt tcttggaact ggtaagttga acccttgcg tggctggt 48

<210> 189
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 189
atggacgacg ttgattctgc tttagccgat aatgtt 36

<210> 190
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 190
ttatataact ttgtccaaga actttctggc c 31

<210> 191
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 191
ttgtttcttt gacacattgc 20

<210> 192
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 192
tgctattacc actgctgg 18

<210> 193
<211> 37
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 193
caatgtgtca aagaaacaag gtaccgagct ctgcgaa 37

<210> 194
<211> 41
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 194
cagcagtggt aatagcatta aaggtatcta cggttgtttc g 41

<210> 195
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 195
gtaaccccag aaggaaac 18

<210> 196
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 196
ttcaagggta ctgtggctac 20

<210> 197
<211> 50
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 197
gtttgctgtt tccttctggg gttacgacgg gtacaacgag aattgtattg 50

<210> 198
<211> 43
<212> DNA
<213> Artificial Sequence

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 287 of 3077

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 198
ccaaagtagc cacagtaccc ttgaatgcga attgagctgg ggg 43

<210> 199
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 199
agatcatcat gcctgcagaa aatcagaggc tactcc 36

<210> 200
<211> 43
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 200
gtcactctct tctattctta tataactttg tccaagaact ttc 43

<210> 201
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 201
gaatagaaga gagtgactct tttg 24

<210> 202
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 202
ctgcaggcat gatgatctg 19

<210> 203
<211> 59
<212> DNA

<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 203
aaaagaaaga aagaaactat aacaatcaat caatggctcc atacctcaaa aaattgaga 59

<210> 204
<211> 59
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 204
agagtcactc tcttctattc ctacaacttg gcccgtctac ctttgttcaa cttcatggt 59

<210> 205
<211> 46
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 205
cacacagctc ttccataatg agatgccaag tatctcaacc atcaag 46

<210> 206
<211> 55
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 206
cacacagctc ttccctctct tctattccta gttcaatgat ttaaggaact tttcc 55

<210> 207
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 207
cacacactgc agttgtccaa tgtaataatt tt 32

<210> 208

<211> 56
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 208
cacacatcta gacccgggct cttcttctga ataggcaatt gataaactta cttatc 56

<210> 209
<211> 48
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 209
gagcccgggt ctagatgtgt gctcttccaa agtacggtgt tgttgaca 48

<210> 210
<211> 37
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 210
cacacacata tgaattctgt actggtagag ctaaatt 37

<210> 211
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 211
gctgccagtg ttggtgacga ct 22

<210> 212
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 212
ttggacggcg ttgtatctgg aa 22

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 290 of 3077

<210> 213
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 213
gctgtttctg acttgtcgtt tgt 23

<210> 214
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 214
tgcagcatgc ttcatagtgg tc 22

<210> 215
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 215
gcttttgttg ttgattccaa gag 23

<210> 216
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 216
caagtcagaa acagatgcag ca 22

<210> 217
<211> 54
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 217

gctgtttctg acttggcttt tgttgttgat tccaagagcg aatccacttc caga 54

<210> 218
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 218
tgcagcatgc ttcatagtgg tc 22

<210> 219
<211> 50
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 219
ttcaagttgg ctcaaatcgc aggtgccaga tacaacgccg tccaatctgc 50

<210> 220
<211> 50
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 220
gcagattgga cggcgttgta tctggcacct gcgatttgag ccaacttgaa 50

<210> 221
<211> 33
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 221
ggtgctcgct cattggctga tggtggcttg ttg 33

<210> 222
<211> 33
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 292 of 3077

<400> 222
caacaagcca ccatcagcca atgagcgagc acc 33

<210> 223
<211> 33
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 223
ctctacgcgt tgtactggta gagctaaatt gta 33

<210> 224
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 224
catcatgcct gcaggtcgac t 21

<210> 225
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 225
tcatgacttt cctgttctat ctctcccc 28

<210> 226
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 226
ccatgctcga gagtgctccg aaaa 24

<210> 227
<211> 57
<212> DNA
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 227
aaaagaaaga aagaaactat aacaatcaat caatgccaga atcaactcaa caatctc 57

<210> 228
<211> 55
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 228
gactcttatc aaaagagtca ctctcttcta ttcttatttc ttgcaggcat ggaca 55

<210> 229
<211> 57
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 229
aaaagaaaga aagaaactat aacaatcaat caatgactac agactcaaac acccaca 57

<210> 230
<211> 57
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 230
gactcttatc aaaagagtca ctctcttcta ttcttacttt ctaccgaatt gagcacc 57

<210> 231
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 231
atgcttacac taacttctgg tcccaacc 28

<210> 232
<211> 28
<212> DNA
<213> Artificial Sequence

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 294 of 3077

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 232
gactattgtc tctagcggtc tagatcct 28

<210> 233
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 233
cacatcgagc catgattgag aac 23

<210> 234
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 234
cgtgagcaaa caattggaac g 21

<210> 235
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 235
atgagtcacc caacgccaga 20

<210> 236
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 236
tcagagcttc ggcgactcag 20

<210> 237
<211> 20

<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 237
accatgacgc tgattgcacc 20

<210> 238
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 238
cgagctacaa tttctgcaca cct 23

<210> 239
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 239
atgcccacat tcaactacaa ggac 24

<210> 240
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 240
cgctctgttg caagtatcgt ca 22

<210> 241
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 241
atgtcaacca ccggaatcta caac 24

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 296 of 3077

<210> 242
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 242
ctacaacttt ggagttgtga cctgcttc 28

<210> 243
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 243
cagaccccga ataatggaca aa 22

<210> 244
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 244
cctctctata atttcggggt gatcac 26

<210> 245
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 245
gatcatggag agattacagg ccg 23

<210> 246
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 246
cgttttcccc tacaactttg gc 22

<210> 247
<211> 52
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 247
aaagaaagaa agaaactata acaatcaatc atgagtcacc caacgccaga ag 52

<210> 248
<211> 55
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 248
tcttatcaaa agagtcactc tcttctattc ttacgactca gctatgtaca acctc 55

<210> 249
<211> 37
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 249
cacacagctc ttccataatg gtcgccgatt ctttagt 37

<210> 250
<211> 47
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 250
cacacagctc ttccctctct tctattctta agtggccttc cacaagt 47

<210> 251
<211> 50
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 298 of 3077

<400> 251
ttgatagttg cgaatttagg tcgacaacga tataaacatt agcatgcctg 50

<210> 252
<211> 47
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 252
cattatggtg ttccggttat gaattcggtt gaattgtgga gtaaaaa 47

<210> 253
<211> 52
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 253
cccgaaatat tacaattgga gctcaaggct atctttaagt ctattgacga ca 52

<210> 254
<211> 54
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 254
gtcagatttc acgcctttac gaattcatca gacttttcaa atctttctct ttcg 54

<210> 255
<211> 50
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 255
caggcatgct aatgtttata tcgttgtcga cctaaattcg caactatcaa 50

<210> 256
<211> 52
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 256
tgtcgtcaat agacttaaag atagccttga gctccaattg taatatttcg gg 52

<210> 257
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 257
cattatggtg ttccggttat 20

<210> 258
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 258
gtcagatttc acgcctttac 20

<210> 259
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 259
cctacttcca cagctttaat ctactatcat 30

<210> 260
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 260
tttaagaaaa caactaagag aagccac 27

<210> 261
<211> 23
<212> DNA
<213> Artificial Sequence

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 300 of 3077

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 261
cattgaagag tgttttgcca gaa 23

<210> 262
<211> 47
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 262
ccgaaacaac cgtagatacc tttaaggcaa aaatgcaatc actctgt 47

<210> 263
<211> 47
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 263
acagagtgat tgcatttttg ccttaaaggt atctacggtt gtttcgg 47

<210> 264
<211> 42
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 264
cgaaatatta caattggagc tcggtaccga gctctgcgaa tt 42

<210> 265
<211> 42
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 265
aattcgcaga gctcggtacc gagctccaat tgtaatattt cg 42

<210> 266
<211> 40
<212> DNA

<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 266
cagtggcggc tttaaccttc cgcatatggt cgacctaaat 40

<210> 267
<211> 40
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 267
atttaggtcg accatatgcg gaaggttaaa gccgccactg 40

<210> 268
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 268
gggtccactc gtcatccaac 20

<210> 269
<211> 56
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 269
aaagaaagaa agaaactata acaatcaatc atgattgaac aactcctaga atattg 56

<210> 270
<211> 51
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 270
tcttatcaaa agagtcactc tcttctattc ctagtcaaac ttgacaatag c 51

<210> 271

<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 271
gaatagaaga gagtgactct tttg 24

<210> 272
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 272
gattgattgt tatagtttct ttctttc 27

<210> 273
<211> 37
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 273
cacacagctc ttccataatg gtcaacatat cgaaatt 37

<210> 274
<211> 47
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 274
cacacagctc ttccctctct tctattctta tatttcaaat actttat 47

<210> 275
<211> 38
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 275
cacacagctc ttccataatg acagtggaga atgcaaaa 38

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 303 of 3077

<210> 276
<211> 63
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 276
cacacagctc ttccctctct tctattccta cttcttggcg acactcttct tttcattctc 60

ctt 63

<210> 277
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 277
gacgggttat aacaagcagg aca 23

<210> 278
<211> 50
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 278
ccgaaacaac cgtagatacc tttaatcaag atcttccaga taatggtcaa 50

<210> 279
<211> 50
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 279
ttgaccatta tctggaagat cttgattaaa ggtatctacg gttgtttcgg 50

<210> 280
<211> 45
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 280
caacaaataa caagttcttg cccttggtac cgagctctgc gaatt 45

<210> 281
<211> 45
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 281
aattcgcaga gctcggtacc aagggcaaga acttgttatt tgttg 45

<210> 282
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 282
cttactaaga gatggacggt ttgaa 25

<210> 283
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 283
aattgacggg ttataacaag cagg 24

<210> 284
<211> 47
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 284
aattcgcaga gctcggtacc ggtgatgtac tggtctatgt tttggat 47

<210> 285
<211> 47
<212> DNA
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 285
atccaaaaca tagaccagta catcaccggt accgagctct gcgaatt 47

<210> 286
<211> 52
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 286
ggtcatcgtc atctatatca gcaatattta aaggtatcta cggttgtttc gg 52

<210> 287
<211> 52
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 287
ccgaaacaac cgtagatacc tttaaatatt gctgatatag atgacgatga cc 52

<210> 288
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 288
cacctggtac atcaactctt ccata 25

<210> 289
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 289
cacacagctc ttccataatg tcc 23

<210> 290
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 290
cacacagctc ttccctctct tct 23

<210> 291
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 291
cacacagctc ttccataatg tcag 24

<210> 292
<211> 37
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 292
cacacagctc ttccatcatg tccgccaacg aaaacat 37

<210> 293
<211> 40
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 293
cacacagctc ttccagctca tgatctagtc ttggccttag 40

<210> 294
<211> 55
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 294
cacacagctc ttccatcatg tcagcgaaat ctattcacga agccgacggt aaggc 55

<210> 295
<211> 44

<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 295
cacacagctc ttccagctca cacgcctaaa ggagtagaag cttc 44

<210> 296
<211> 43
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 296
cacacagctc ttccataatg cctaccgaac ttcaaaaaga aag 43

<210> 297
<211> 51
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 297
cacacagctc ttccctctct tctattctta actggacaag atttcagcag c 51

<210> 298
<211> 39
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 298
gtctttgacc cacaagtcgg aaccagaatc ggtgtcaac 39

<210> 299
<211> 39
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 299
gttgacaccg attctggttc cgacttgtgg gtcaaagac 39

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 308 of 3077

<210> 300
<211> 37
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 300
gtaaagccta ctccgacttt gttgtccaat gtacctg 37

<210> 301
<211> 37
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 301
caggtacatt ggacaacaaa gtcggagtag gctttac 37

<210> 302
<211> 53
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 302
gaattcacat ggctaatttg gcctcggttc cacaacgcac tcagcattaa aaa 53

<210> 303
<211> 40
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 303
gagctcccct gcaaacaggg aaacacttgt catctgattt 40

<210> 304
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 304
gatattattc caccttccct tcatt 25

<210> 305
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 305
ccgttaaaca aaaatcagtc tgtaaa 26

<210> 306
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 306
tgccatcctt ggtagtcagt tatt 24

<210> 307
<211> 45
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 307
ccgaaacaac cgtagatacc tttaatggct tgtccttggt gttga 45

<210> 308
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 308
tgtcgccatt caaccagtag at 22

<210> 309
<211> 43
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 310 of 3077

<400> 309
aattcgcaga gctcggtacc gacttgttga agatggacga gga 43

<210> 310
<211> 45
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 310
tcaacaccaa ggacaagcca ttaaaggtat ctacggttgt ttcgg 45

<210> 311
<211> 43
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 311
tcctcgtcca tcttcaacaa gtcggtaccg agctctgcga att 43

<210> 312
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 312
ttgatccact gtcttaagat tgtcaa 26

<210> 313
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 313
aggattttgc tgttggtggc 20

<210> 314
<211> 41
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 314
aattcgcaga gctcggtacc atacgtcggt gaggtctgtc g 41

<210> 315
<211> 47
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 315
attctccgct acttcgttct ggttaaaggt atctacggtt gtttcgg 47

<210> 316
<211> 41
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 316
cgacagacct caccgacgta tggtaccgag ctctgcgaat t 41

<210> 317
<211> 47
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 317
ccgaaacaac cgtagatacc tttaaccaga acgaagtagc ggagaat 47

<210> 318
<211> 40
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 318
aagcttttaa ttaaagataa tcacaggggt agagaccttg 40

<210> 319
<211> 48
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 319
ggatccgcat gcggccggcc gatagcgtgg tatgaatgaa taagtgtg 48

<210> 320
<211> 51
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 320
ggatccgagc tcgcggccgc gagcactagg ttttgataat ttggttctta c 51

<210> 321
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 321
gaattcttaa ttaacggcga agaacatagt gtgatg 36

<210> 322
<211> 355
<212> PRT
<213> Sulfolobus islandicus

<400> 322
Met Arg Arg Thr Leu Lys Ala Ala Ile Leu Gly Ser Thr Gly Leu Val
1 5 10 15

Gly Ile Glu Tyr Val Arg Met Leu Ala Asn His Pro Tyr Ile Lys Val
20 25 30

Gly Tyr Leu Ala Gly Lys Gly Ser Val Gly Lys Pro Tyr Glu Glu Val
35 40 45

Val Arg Trp Gln Thr Val Gly Gln Val Pro Lys Glu Ile Ala Asp Met
50 55 60

Glu Val Lys Pro Thr Asp Pro Lys Leu Met Asp Asp Val Asp Ile Ile
65 70 75 80

Phe Ser Pro Leu Pro Gln Gly Ala Ala Gly Pro Val Glu Glu Glu Phe
85 90 95

```
Ala Lys His Gly Phe Pro Val Ile Ser Asn Ser Pro Asp His Arg Phe
            100                 105                 110

Asp Pro Asp Val Pro Leu Leu Ile Pro Glu Ile Asn Pro His Thr Ile
        115                 120                 125

Ser Leu Ile Asp Lys Gln Arg Glu Arg His Asp Trp Arg Gly Phe Ile
    130                 135                 140

Val Thr Thr Pro Leu Cys Thr Ala Gln Gly Ala Ala Ile Pro Leu Ala
145                 150                 155                 160

Pro Ile Tyr Met Asn Phe Arg Ile Asp Ser Ser Leu Ile Thr Thr Ile
                165                 170                 175

Gln Ser Leu Ser Gly Ala Gly Tyr Pro Gly Ile Pro Ser Leu Asp Val
            180                 185                 190

Val Asp Asn Val Leu Pro Leu Gly Asp Asn Tyr Asp Asn Lys Thr Val
        195                 200                 205

Lys Glu Ile Ser Arg Ile Leu Ser Glu Thr Lys Arg Met Val Asn Asp
    210                 215                 220

Asp Asn Asp Leu Ser Leu Gly Ala Thr Thr His Arg Ile Ser Thr Ile
225                 230                 235                 240

His Gly His Tyr Glu Val Ile Tyr Val Thr Phe Arg Glu Asp Val Ser
                245                 250                 255

Val Glu Lys Ile Arg Glu Thr Leu Asp Ser Phe Arg Gly Glu Pro Gln
            260                 265                 270

Lys Leu Lys Leu Pro Thr Ala Pro Asp Lys Pro Ile Leu Leu Thr Asn
        275                 280                 285

Gln Asp Ala Arg Pro Gln Val Tyr Phe Asp Arg Trp Ala Gly Asp Pro
    290                 295                 300

Pro Gly Met Ser Val Val Val Gly Arg Leu Ser Gln Ile Asn Arg Arg
305                 310                 315                 320

Thr Ile Arg Leu Val Ser Val Val His Asn Thr Val Arg Gly Ala Ala
                325                 330                 335

Gly Gly Gly Ile Leu Ala Ala Glu Leu Leu Val Glu Lys Gly Tyr Ile
            340                 345                 350
```

Asp Lys Arg
355

<210> 323
<211> 1068
<212> DNA
<213> Sulfolobus islandicus

<400> 323
atgaggagaa ctcttaaggc cgccattttg ggatcgactg gattggttgg tatcgaatat 60

gttagaatgt tagcaaacca cccttacata aaggtgggtt acctagcggg taagggctca 120

gtcggcaagc cctacgaaga agttgtcaga tggcaaaccg tcggccaagt tcctaaagaa 180

atcgctgata tggaagtcaa gcctaccgat cccaagttga tggacgacgt cgacattatc 240

ttttcgccgt tgccccaagg cgctgctggt cctgttgaag aggaatttgc caaacatgga 300

tttcctgtca tttccaactc tcctgatcac agattcgacc cagatgtccc tctcctcatc 360

ccagaaatca atccgcatac aatctcgctt attgacaaac aaagggaaag gcatgactgg 420

aggggcttca tcgttaccac tcctttgtgc accgcgcaag gtgctgccat tcctttggcc 480

cctatctata tgaatttcag aatcgactcc agtctcatca ctaccattca atcgctctct 540

ggcgctgggt atccaggtat cccatcatta gacgttgtcg acaacgtctt gccattaggc 600

gataattatg acaacaaaac agttaaagag atttccagga ttttgtccga gacaaagagg 660

atggtgaacg atgacaacga tctatcgttg ggcgccacta cccacagaat ctcgactatt 720

cacgggcatt acgaggtcat ttatgtaacc ttcagagaag acgtttccgt cgaaaagata 780

agagagacgt tagactcatt tagaggtgaa ccccagaagt tgaagttgcc taccgcgcct 840

gacaagccta tacttttgac caaccaagac gcaagaccac aagtctactt tgacaggtgg 900

gccggagacc ctccaggtat gtccgtggtc gtgggtagat tgagtcaaat caacaggaga 960

accattaggc tcgtctcggt cgttcataat accgtcaggg gagctgctgg tggtggaatc 1020

ttggccgctg aactactcgt agagaagggt tacattgaca agaggtaa 1068

<210> 324
<211> 402
<212> PRT
<213> Gerbera hybrida

<400> 324
Met Gly Ser Tyr Ser Ser Asp Asp Val Glu Val Ile Arg Glu Ala Gly
1 5 10 15

Arg Ala Gln Gly Leu Ala Thr Ile Leu Ala Ile Gly Thr Ala Thr Pro
20 25 30

```
Pro Asn Cys Val Ala Gln Ala Asp Tyr Ala Asp Tyr Tyr Phe Arg Val
        35                  40                  45

Thr Lys Ser Glu His Met Val Asp Leu Lys Glu Lys Phe Lys Arg Ile
    50                  55                  60

Cys Glu Lys Thr Ala Ile Lys Lys Arg Tyr Leu Ala Leu Thr Glu Asp
65                  70                  75                  80

Tyr Leu Gln Glu Asn Pro Thr Met Cys Glu Phe Met Ala Pro Ser Leu
                85                  90                  95

Asn Ala Arg Gln Asp Leu Val Val Thr Gly Val Pro Met Leu Gly Lys
            100                 105                 110

Glu Ala Ala Val Lys Ala Ile Asp Glu Trp Gly Leu Pro Lys Ser Lys
        115                 120                 125

Ile Thr His Leu Ile Phe Cys Thr Thr Ala Gly Val Asp Met Pro Gly
    130                 135                 140

Ala Asp Tyr Gln Leu Val Lys Leu Leu Gly Leu Ser Pro Ser Val Lys
145                 150                 155                 160

Arg Tyr Met Leu Tyr Gln Gln Gly Cys Ala Ala Gly Gly Thr Val Leu
                165                 170                 175

Arg Leu Ala Lys Asp Leu Ala Glu Asn Asn Lys Gly Ser Arg Val Leu
            180                 185                 190

Ile Val Cys Ser Glu Ile Thr Ala Ile Leu Phe His Gly Pro Asn Glu
        195                 200                 205

Asn His Leu Asp Ser Leu Val Ala Gln Ala Leu Phe Gly Asp Gly Ala
    210                 215                 220

Ala Ala Leu Ile Val Gly Ser Gly Pro His Leu Ala Val Glu Arg Pro
225                 230                 235                 240

Ile Phe Glu Ile Val Ser Thr Asp Gln Thr Ile Leu Pro Asp Thr Glu
                245                 250                 255

Lys Ala Met Lys Leu His Leu Arg Glu Gly Gly Leu Thr Phe Gln Leu
            260                 265                 270

His Arg Asp Val Pro Leu Met Val Ala Lys Asn Ile Glu Asn Ala Ala
        275                 280                 285
```

Glu Lys Ala Leu Ser Pro Leu Gly Ile Thr Asp Trp Asn Ser Val Phe
290 295 300

Trp Met Val His Pro Gly Gly Arg Ala Ile Leu Asp Gln Val Glu Arg
305 310 315 320

Lys Leu Asn Leu Lys Glu Asp Lys Leu Arg Ala Ser Arg His Val Leu
325 330 335

Ser Glu Tyr Gly Asn Leu Ile Ser Ala Cys Val Leu Phe Ile Ile Asp
340 345 350

Glu Val Arg Lys Arg Ser Met Ala Glu Gly Lys Ser Thr Thr Gly Glu
355 360 365

Gly Leu Asp Cys Gly Val Leu Phe Gly Phe Gly Pro Gly Met Thr Val
370 375 380

Glu Thr Val Val Leu Arg Ser Val Arg Val Thr Ala Ala Val Ala Asn
385 390 395 400

Gly Asn

<210> 325
<211> 1209
<212> DNA
<213> Gerbera hybrida

<400> 325
atgggatcct actcaagtga cgacgtcgaa gtcattcgtg aagccggacg agcgcaaggt 60

ctcgccacca ttttagccat tggcactgct actcctccga attgtgttgc tcaagctgac 120

tatgcggatt attatttccg tgttactaag tctgagcata tggtggactt aaaagaaaaa 180

ttcaaacgga tttgcgaaaa aacggcaata aagaaacggt acctagcctt gacagaggat 240

tatctccaag aaaacccaac gatgtgcgaa tttatggctc caagtctcaa cgctcggcaa 300

gatctagtcg ttacaggcgt tccaatgtta ggcaaagagg ccgcggttaa ggccattgac 360

gagtggggac taccaaaaag taagatcaca cacttgatct tttgtacaac agctggcgtg 420

gatatgccgg gtgctgatta tcaattggtt aaattgttag gtttgagtcc ttcagttaaa 480

cggtatatgc tttaccaaca gggatgcgcc gccggcggca cggttttgcg actagccaag 540

gatttagctg agaacaacaa gggctcacgg gttttaatcg tttgtagtga aatcactgct 600

atcctctttc atggaccgaa tgaaaaccac ttagattcat tggttgctca agctctcttt 660

ggagatggag ctgcggcgtt gattgtcggt tcaggccctc accttgccgt agagcgacca 720

atatttgaaa tcgtctcaac tgaccaaacg atccttcccg atactgaaaa ggcgatgaag 780

ctccacctta gagaaggagg gcttaccttc cagcttcata gagacgtacc gcttatggtt 840

gcgaagaaca tagaaaacgc ggcagaaaaa gcactttccc cactagggat aactgactgg 900

aactcagtgt tttggatggt ccacccaggt ggtcgggcca tacttgatca ggtcgaacgg 960

aaactaaact taaaggagga caagctcagg gcttctaggc atgtcttaag cgagtacgga 1020

aacctcattt ctgcttgcgt ccttttttatc attgatgaag tcaggaagag atccatggca 1080

gaggggaaga gcacgacagg tgagggtctt gactgtggtg tgcttttcgg attcggaccc 1140

ggtatgactg tggaaactgt ggtgttacgt tctgttcggg tgactgctgc agtggccaat 1200

ggaaactaa 1209

<210> 326
<211> 301
<212> PRT
<213> Chronobacter sakazakii

<400> 326

Met Asn Ala Asn Ala Val Lys Ser Ser Gly Gln Glu Ile Glu Leu Gln
1 5 10 15

Ala Leu Arg Asp Ala Leu Gln Thr Arg Leu Asp Glu Leu Leu Pro Pro
20 25 30

Gly Gln Glu Arg Asp Leu Val Cys Ala Ala Met Arg Glu Gly Ala Leu
35 40 45

Thr Pro Gly Lys Arg Val Arg Pro Leu Leu Leu Ile Leu Ala Ala Arg
50 55 60

Asp Leu Gly Cys Asp Ala Ser Gln Pro Ala Leu Met Asp Leu Ala Cys
65 70 75 80

Ala Val Glu Met Val His Ala Ala Ser Leu Met Leu Asp Asp Ile Pro
85 90 95

Cys Met Asp Asn Ala Leu Leu Arg Arg Gly Lys Pro Thr Ile His Arg
100 105 110

Gln Tyr Gly Glu Ser Val Ala Ile Leu Ala Ala Val Ala Leu Leu Ser
115 120 125

Arg Ala Phe Gly Val Val Ala Gln Ala Asn Pro Leu Ser Asp Ser Cys
130 135 140

Lys Thr Gln Ala Val Ser Glu Leu Ser Ser Ala Val Gly Leu Gln Gly
145 150 155 160

```
Leu Val Gln Gly Gln Phe Arg Asp Leu Ser Glu Gly Asn Gln Ala Arg
                165                 170                 175

Ser Ala Glu Ala Ile Leu Ala Thr Asn Asp Leu Lys Thr Ser Val Leu
            180                 185                 190

Phe Asp Ala Thr Leu Gln Ile Ala Ala Ile Ala Ala Gly Ala Ser Ala
        195                 200                 205

Ser Val Arg His Lys Leu Arg Glu Phe Ser Arg His Leu Gly Gln Ala
    210                 215                 220

Phe Gln Leu Leu Asp Asp Leu Ala Asp Gly Leu Asn His Thr Gly Lys
225                 230                 235                 240

Asp Ile Asn Lys Asp Ala Gly Lys Ser Thr Leu Val Ala Met Leu Gly
                245                 250                 255

Pro Glu Ala Val His Gln Arg Leu Arg Asp His Leu Leu Arg Ala Asp
            260                 265                 270

Glu His Leu Thr Gly Ala Cys Ser Arg Gly Ala Ser Thr Arg Arg Phe
        275                 280                 285

Met Tyr Ala Trp Phe Asp Lys Gln Leu Ala Met Phe Gly
    290                 295                 300


<210> 327
<211> 492
<212> PRT
<213> Chronobacter sakazakii

<400> 327
Met Thr Lys Thr Val Val Ile Gly Ser Gly Phe Gly Gly Leu Ala Leu
1               5                   10                  15

Ala Ile Arg Leu Gln Ala Ala Gly Val Pro Thr Leu Leu Leu Glu Gln
            20                  25                  30

Arg Asp Lys Pro Gly Gly Arg Ala Tyr Val Tyr Glu Asp Lys Gly Phe
        35                  40                  45

Thr Phe Asp Ala Gly Pro Thr Val Ile Thr Asp Pro Ser Ala Ile Glu
    50                  55                  60

Glu Leu Phe Thr Leu Ala Gly Lys Asn Ile Ala Asp Tyr Val Asp Leu
65                  70                  75                  80
```

```
Leu Pro Val Thr Pro Phe Tyr Arg Leu Cys Trp Glu Asn Gly Gln Val
                85                  90                  95

Phe Asn Tyr Asp Asn Asp Gln Ala Ser Leu Glu Ala Gln Ile Ala Arg
            100                 105                 110

Phe Asn Pro Arg Asp Val Glu Gly Tyr Arg Gln Phe Leu Ala Tyr Ser
        115                 120                 125

Gln Ala Val Phe Lys Glu Gly Tyr Leu Lys Leu Gly Ala Val Pro Phe
    130                 135                 140

Leu Ser Phe Arg Asp Met Leu Arg Ala Gly Pro Gln Leu Ala Arg Leu
145                 150                 155                 160

Gln Ala Trp Arg Ser Val Tyr Gly Met Val Ser Lys Phe Ile Glu Asn
                165                 170                 175

Asp His Leu Arg Gln Ala Phe Ser Phe His Ser Leu Leu Val Gly Gly
            180                 185                 190

Asn Pro Phe Ala Thr Ser Ser Ile Tyr Thr Leu Ile His Ala Leu Glu
        195                 200                 205

Arg Gln Trp Gly Val Trp Phe Ala Arg Gly Gly Thr Gly Ala Leu Val
    210                 215                 220

Gln Gly Leu Val Lys Leu Phe Thr Asp Leu Gly Gly Glu Ile Glu Leu
225                 230                 235                 240

Asn Ala Lys Val Thr Arg Leu Asp Thr Gln Gly Asp Lys Ile Ser Gly
                245                 250                 255

Val Thr Leu Ala Asp Gly Arg Arg Ile Pro Ala Arg Ala Val Ala Ser
            260                 265                 270

Asn Ala Asp Val Val His Thr Tyr Asn Asn Leu Leu Gly His His Pro
        275                 280                 285

Arg Gly Val Ser Gln Ala Ala Ser Leu Arg Arg Lys Arg Met Ser Asn
    290                 295                 300

Ser Leu Phe Val Leu Tyr Phe Gly Leu Asn His His His Ser Gln Leu
305                 310                 315                 320

Ala His His Thr Val Cys Phe Gly Pro Arg Tyr Lys Gly Leu Ile Glu
                325                 330                 335
```

```
Asp Ile Phe Lys Arg Asp Ser Leu Ala Asp Asp Phe Ser Leu Tyr Leu
            340                 345                 350

His Ala Pro Cys Val Thr Asp Pro Ser Leu Ala Pro Pro Gly Cys Gly
        355                 360                 365

Ser Tyr Tyr Val Leu Ala Pro Val Pro His Leu Gly Thr Ala Asn Leu
    370                 375                 380

Asn Trp Asp Val Glu Gly Pro Arg Leu Arg Asp Arg Ile Phe Glu Tyr
385                 390                 395                 400

Leu Glu Gln His Tyr Met Pro Gly Leu Arg Asp Gln Leu Val Thr His
                405                 410                 415

Arg Met Phe Thr Pro Phe Asp Phe Arg Asp Gln Leu Gly Ala Tyr His
            420                 425                 430

Gly Ser Ala Phe Ser Val Glu Pro Ile Leu Thr Gln Ser Ala Trp Phe
        435                 440                 445

Arg Pro His Asn Arg Asp Ser Arg Ile Asp Asn Leu Tyr Leu Val Gly
    450                 455                 460

Ala Gly Thr His Pro Gly Ala Gly Ile Pro Gly Val Ile Gly Ser Ala
465                 470                 475                 480

Lys Ala Thr Ala Gly Leu Met Leu Glu Gly His Ala
                485                 490


<210> 328
<211> 323
<212> PRT
<213> Chronobacter sakazakii

<400> 328
Met Ser Asp Lys Pro Leu Leu Thr His Ala Thr Glu Thr Ile Glu Ala
1               5                   10                  15

Gly Ser Lys Ser Phe Ala Thr Ala Ser Lys Leu Phe Asp Ala Lys Thr
            20                  25                  30

Arg Arg Ser Ala Leu Met Leu Tyr Ala Trp Cys Arg His Cys Asp Asp
        35                  40                  45

Val Thr Asp Gly Gln Ala Leu Gly Phe Arg Ala Ala Asp Ala Pro Thr
    50                  55                  60
```

```
Asp Thr Pro Gln Ala Arg Ile Ala Leu Leu Arg Ala Leu Thr Leu Glu
65                  70                  75                  80

Ala Tyr Ala Gly Lys Pro Met Arg Glu Pro Asn Phe Ala Ala Phe Gln
                85                  90                  95

Glu Val Ala Leu Ala His Gln Ile Pro Pro Ala Leu Ala Leu Asp His
            100                 105                 110

Leu Glu Gly Phe Ala Met Asp Val Arg Glu Glu Arg Tyr His Thr Phe
        115                 120                 125

Asp Asp Thr Leu Arg Tyr Cys Tyr His Val Ala Gly Val Val Gly Leu
    130                 135                 140

Met Met Ala Arg Val Met Gly Val Arg Asp Glu Ala Val Leu Asp Arg
145                 150                 155                 160

Ala Cys Asp Leu Gly Leu Ala Phe Gln Leu Thr Asn Ile Ala Arg Asp
                165                 170                 175

Ile Val Glu Asp Ala Ala Ile Gly Arg Cys Tyr Leu Pro Glu Ala Trp
            180                 185                 190

Leu Gln Glu Glu Gly Leu Cys Ala Asp Thr Leu Thr Asp Arg Ala His
        195                 200                 205

Arg Pro Ala Leu Ala Arg Leu Ala Ala Arg Leu Val Asp Glu Ala Glu
    210                 215                 220

Pro Tyr Tyr Ala Ser Ala Arg Ala Gly Leu Ala Gly Leu Pro Leu Arg
225                 230                 235                 240

Ser Ala Trp Ala Ile Ala Thr Ala His Gly Val Tyr Arg Glu Ile Gly
                245                 250                 255

Val Lys Val Lys Arg Ala Gly Val Asn Ala Trp Glu Thr Arg Gln Gly
            260                 265                 270

Thr Ser Lys Ala Glu Lys Leu Ala Leu Leu Ala Lys Gly Ala Val Met
        275                 280                 285

Ala Val Ser Ser Arg Gly Ala Ser Ser Ser Pro Arg Pro Ser Ala Leu
    290                 295                 300

Trp Gln Arg Pro Arg Ala Gln Asp Asp Arg Tyr Ala His Ala Ala Pro
305                 310                 315                 320
```

Pro Ala Ala

<210> 329
<211> 376
<212> PRT
<213> Xanthophyllomyces dendrorhous

<400> 329
Met Asp Tyr Ala Asn Ile Leu Thr Ala Ile Pro Leu Glu Phe Thr Pro
1 5 10 15

Gln Asp Asp Ile Val Leu Leu Glu Pro Tyr His Tyr Leu Gly Lys Asn
20 25 30

Pro Gly Lys Glu Ile Arg Ser Gln Leu Ile Glu Ala Phe Asn Tyr Trp
35 40 45

Leu Asp Val Lys Lys Glu Asp Leu Glu Val Ile Gln Asn Val Val Gly
50 55 60

Met Leu His Thr Ala Ser Leu Leu Met Asp Asp Val Glu Asp Ser Ser
65 70 75 80

Val Leu Arg Arg Gly Ser Pro Val Ala His Leu Ile Tyr Gly Ile Pro
85 90 95

Gln Thr Ile Asn Thr Ala Asn Tyr Val Tyr Phe Leu Ala Tyr Gln Glu
100 105 110

Ile Phe Lys Leu Arg Pro Thr Pro Ile Pro Met Pro Val Ile Pro Pro
115 120 125

Ser Ser Ala Ser Leu Gln Ser Ser Val Ser Ser Ala Ser Ser Ser Ser
130 135 140

Ser Ala Ser Ser Glu Asn Gly Gly Thr Ser Thr Pro Asn Ser Gln Ile
145 150 155 160

Pro Phe Ser Lys Asp Thr Tyr Leu Asp Lys Val Ile Thr Asp Glu Met
165 170 175

Leu Ser Leu His Arg Gly Gln Gly Leu Glu Leu Phe Trp Arg Asp Ser
180 185 190

Leu Thr Cys Pro Ser Glu Glu Glu Tyr Val Lys Met Val Leu Gly Lys
195 200 205

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 323 of 3077

```
Thr Gly Gly Leu Phe Arg Ile Ala Val Arg Leu Met Met Ala Lys Ser
    210                 215                 220

Glu Cys Asp Ile Asp Phe Val Gln Leu Val Asn Leu Ile Ser Ile Tyr
225                 230                 235                 240

Phe Gln Ile Arg Asp Asp Tyr Met Asn Leu Gln Ser Ser Glu Tyr Ala
                245                 250                 255

His Asn Lys Asn Phe Ala Glu Asp Leu Thr Glu Gly Lys Phe Ser Phe
            260                 265                 270

Pro Thr Ile His Ser Ile His Ala Asn Pro Ser Ser Arg Leu Val Ile
        275                 280                 285

Asn Thr Leu Gln Lys Lys Ser Thr Ser Pro Glu Ile Leu His His Cys
    290                 295                 300

Val Asn Tyr Met Arg Thr Glu Thr His Ser Phe Glu Tyr Thr Gln Glu
305                 310                 315                 320

Val Leu Asn Thr Leu Ser Gly Ala Leu Glu Arg Glu Leu Gly Arg Leu
                325                 330                 335

Gln Gly Glu Phe Ala Glu Ala Asn Ser Lys Ile Asp Leu Gly Asp Val
            340                 345                 350

Glu Ser Glu Gly Arg Thr Gly Lys Asn Val Lys Leu Glu Ala Ile Leu
        355                 360                 365

Lys Lys Leu Ala Asp Ile Pro Leu
    370                 375


<210> 330
<211> 582
<212> PRT
<213> Xanthophyllomyces dendrorhous

<400> 330
Met Gly Lys Glu Gln Asp Gln Asp Lys Pro Thr Ala Ile Ile Val Gly
1               5                   10                  15

Cys Gly Ile Gly Gly Ile Ala Thr Ala Ala Arg Leu Ala Lys Glu Gly
            20                  25                  30

Phe Gln Val Thr Val Phe Glu Lys Asn Asp Tyr Ser Gly Gly Arg Cys
        35                  40                  45
```

```
Ser Leu Ile Glu Arg Asp Gly Tyr Arg Phe Asp Gln Gly Pro Ser Leu
    50                  55                  60

Leu Leu Leu Pro Asp Leu Phe Lys Gln Thr Phe Glu Asp Leu Gly Glu
65                  70                  75                  80

Lys Met Glu Asp Trp Val Asp Leu Ile Lys Cys Glu Pro Asn Tyr Val
                85                  90                  95

Cys His Phe His Asp Glu Glu Thr Phe Thr Phe Ser Thr Asp Met Ala
            100                 105                 110

Leu Leu Lys Arg Glu Val Glu Arg Phe Glu Gly Lys Asp Gly Phe Asp
        115                 120                 125

Arg Phe Leu Ser Phe Ile Gln Glu Ala His Arg His Tyr Glu Leu Ala
    130                 135                 140

Val Val His Val Leu Gln Lys Asn Phe Pro Gly Phe Ala Ala Phe Leu
145                 150                 155                 160

Arg Leu Gln Phe Ile Gly Gln Ile Leu Ala Leu His Pro Phe Glu Ser
                165                 170                 175

Ile Trp Thr Arg Val Cys Arg Tyr Phe Lys Thr Asp Arg Leu Arg Arg
            180                 185                 190

Val Phe Ser Phe Ala Val Met Tyr Met Gly Gln Ser Pro Tyr Ser Ala
        195                 200                 205

Pro Gly Thr Tyr Ser Leu Leu Gln Tyr Thr Glu Leu Thr Glu Gly Ile
    210                 215                 220

Trp Tyr Pro Arg Gly Gly Phe Trp Gln Val Pro Asn Thr Leu Leu Gln
225                 230                 235                 240

Ile Val Lys Arg Asn Asn Pro Ser Ala Lys Phe Asn Phe Asn Ala Pro
                245                 250                 255

Val Ser Gln Val Leu Leu Ser Pro Ala Lys Asp Arg Ala Thr Gly Val
            260                 265                 270

Arg Leu Glu Ser Gly Glu Glu His His Ala Asp Val Val Ile Val Asn
        275                 280                 285

Ala Asp Leu Val Tyr Ala Ser Glu His Leu Ile Pro Asp Asp Ala Arg
    290                 295                 300
```


LCY Biotechnology Holding, Inc.
Ex. 1002
Page 325 of 3077

```
Asn Lys Ile Gly Gln Leu Gly Glu Val Lys Arg Ser Trp Trp Ala Asp
305                 310                 315                 320

Leu Val Gly Gly Lys Lys Leu Lys Gly Ser Cys Ser Ser Leu Ser Phe
                325                 330                 335

Tyr Trp Ser Met Asp Arg Ile Val Asp Gly Leu Gly Gly His Asn Ile
            340                 345                 350

Phe Leu Ala Glu Asp Phe Lys Gly Ser Phe Asp Thr Ile Phe Glu Glu
        355                 360                 365

Leu Gly Leu Pro Ala Asp Pro Ser Phe Tyr Val Asn Val Pro Ser Arg
    370                 375                 380

Ile Asp Pro Ser Ala Ala Pro Glu Gly Lys Asp Ala Ile Val Ile Leu
385                 390                 395                 400

Val Pro Cys Gly His Ile Asp Ala Ser Asn Pro Gln Asp Tyr Asn Lys
                405                 410                 415

Leu Val Ala Arg Ala Arg Lys Phe Val Ile Gln Thr Leu Ser Ala Lys
            420                 425                 430

Leu Gly Leu Pro Asp Phe Glu Lys Met Ile Val Ala Glu Lys Val His
        435                 440                 445

Asp Ala Pro Ser Trp Glu Lys Glu Phe Asn Leu Lys Asp Gly Ser Ile
    450                 455                 460

Leu Gly Leu Ala His Asn Phe Met Gln Val Leu Gly Phe Arg Pro Ser
465                 470                 475                 480

Thr Arg His Pro Lys Tyr Asp Lys Leu Phe Phe Val Gly Ala Ser Thr
                485                 490                 495

His Pro Gly Thr Gly Val Pro Ile Val Leu Ala Gly Ala Lys Leu Thr
            500                 505                 510

Ala Asn Gln Val Leu Glu Ser Phe Asp Arg Ser Pro Ala Pro Asp Pro
        515                 520                 525

Asn Met Ser Leu Ser Val Pro Tyr Gly Lys Pro Leu Lys Ser Asn Gly
    530                 535                 540

Thr Gly Ile Asp Ser Gln Val Gln Leu Lys Phe Met Asp Leu Glu Arg
545                 550                 555                 560
```

```
Trp Val Tyr Leu Leu Val Leu Leu Ile Gly Ala Val Ile Ala Arg Ser
                565                 570                 575

Val Gly Val Leu Ala Phe
            580
```

<210> 331
<211> 673
<212> PRT
<213> Xanthophyllomyces dendrorhous

<400> 331

```
Met Thr Ala Leu Ala Tyr Tyr Gln Ile His Leu Ile Tyr Thr Leu Pro
1               5                   10                  15

Ile Leu Gly Leu Leu Gly Leu Leu Thr Ser Pro Ile Leu Thr Lys Phe
            20                  25                  30

Asp Ile Tyr Lys Ile Ser Ile Leu Val Phe Ile Ala Phe Ser Ala Thr
        35                  40                  45

Thr Pro Trp Asp Ser Trp Ile Ile Arg Asn Gly Ala Trp Thr Tyr Pro
    50                  55                  60

Ser Ala Glu Ser Gly Gln Gly Val Phe Gly Thr Phe Leu Asp Val Pro
65                  70                  75                  80

Tyr Glu Glu Tyr Ala Phe Phe Val Ile Gln Thr Val Ile Thr Gly Leu
                85                  90                  95

Val Tyr Val Leu Ala Thr Arg His Leu Leu Pro Ser Leu Ala Leu Pro
            100                 105                 110

Lys Thr Arg Ser Ser Ala Leu Ser Leu Ala Leu Lys Ala Leu Ile Pro
        115                 120                 125

Leu Pro Ile Ile Tyr Leu Phe Thr Ala His Pro Ser Pro Ser Pro Asp
    130                 135                 140

Pro Leu Val Thr Asp His Tyr Phe Tyr Met Arg Ala Leu Ser Leu Leu
145                 150                 155                 160

Ile Thr Pro Pro Thr Met Leu Leu Ala Ala Leu Ser Gly Glu Tyr Ala
                165                 170                 175

Phe Asp Trp Lys Ser Gly Arg Ala Lys Ser Thr Ile Ala Ala Ile Met
            180                 185                 190
```

```
Ile Pro Thr Val Tyr Leu Ile Trp Val Asp Tyr Val Ala Val Gly Gln
        195                 200                 205

Asp Ser Trp Ser Ile Asn Asp Glu Lys Ile Val Gly Trp Arg Leu Gly
    210                 215                 220

Gly Val Leu Pro Ile Glu Glu Ala Met Phe Phe Leu Leu Thr Asn Leu
225                 230                 235                 240

Met Ile Val Leu Gly Leu Ser Ala Cys Asp His Thr Gln Ala Leu Tyr
                245                 250                 255

Leu Leu His Gly Arg Thr Ile Tyr Gly Asn Lys Lys Met Pro Ser Ser
            260                 265                 270

Phe Pro Leu Ile Thr Pro Pro Val Leu Ser Leu Phe Phe Ser Ser Arg
        275                 280                 285

Pro Tyr Ser Ser Gln Pro Lys Arg Asp Leu Glu Leu Ala Val Lys Leu
    290                 295                 300

Leu Glu Glu Lys Ser Arg Ser Phe Phe Val Ala Ser Ala Gly Phe Pro
305                 310                 315                 320

Ser Glu Val Arg Glu Arg Leu Val Gly Leu Tyr Ala Phe Cys Arg Val
                325                 330                 335

Thr Asp Asp Leu Ile Asp Ser Pro Glu Val Ser Ser Asn Pro His Ala
            340                 345                 350

Thr Ile Asp Met Val Ser Asp Phe Leu Thr Leu Leu Phe Gly Pro Pro
        355                 360                 365

Leu His Pro Ser Gln Pro Asp Lys Ile Leu Ser Ser Pro Leu Leu Pro
    370                 375                 380

Pro Ser His Pro Ser Arg Pro Thr Gly Met Tyr Pro Leu Pro Pro Pro
385                 390                 395                 400

Pro Ser Leu Ser Pro Ala Glu Leu Val Gln Phe Leu Thr Glu Arg Val
                405                 410                 415

Pro Val Gln Tyr His Phe Ala Phe Arg Leu Leu Ala Lys Leu Gln Gly
            420                 425                 430

Leu Ile Pro Arg Tyr Pro Leu Asp Glu Leu Leu Arg Gly Tyr Thr Thr
        435                 440                 445
```

```
Asp Leu Ile Phe Pro Leu Ser Thr Glu Ala Val Gln Ala Arg Lys Thr
    450                 455                 460

Pro Ile Glu Thr Thr Ala Asp Leu Leu Asp Tyr Gly Leu Cys Val Ala
465                 470                 475                 480

Gly Ser Val Ala Glu Leu Leu Val Tyr Val Ser Trp Ala Ser Ala Pro
                485                 490                 495

Ser Gln Val Pro Ala Thr Ile Glu Glu Arg Glu Ala Val Leu Val Ala
            500                 505                 510

Ser Arg Glu Met Gly Thr Ala Leu Gln Leu Val Asn Ile Ala Arg Asp
        515                 520                 525

Ile Lys Gly Asp Ala Thr Glu Gly Arg Phe Tyr Leu Pro Leu Ser Phe
    530                 535                 540

Phe Gly Leu Arg Asp Glu Ser Lys Leu Ala Ile Pro Thr Asp Trp Thr
545                 550                 555                 560

Glu Pro Arg Pro Gln Asp Phe Asp Lys Leu Leu Ser Leu Ser Pro Ser
                565                 570                 575

Ser Thr Leu Pro Ser Ser Asn Ala Ser Glu Ser Phe Arg Phe Glu Trp
            580                 585                 590

Lys Thr Tyr Ser Leu Pro Leu Val Ala Tyr Ala Glu Asp Leu Ala Lys
        595                 600                 605

His Ser Tyr Lys Gly Ile Asp Arg Leu Pro Thr Glu Val Gln Ala Gly
    610                 615                 620

Met Arg Ala Ala Cys Ala Ser Tyr Leu Leu Ile Gly Arg Glu Ile Lys
625                 630                 635                 640

Val Val Trp Lys Gly Asp Val Gly Glu Arg Arg Thr Val Ala Gly Trp
                645                 650                 655

Arg Arg Val Arg Lys Val Leu Ser Val Val Met Ser Gly Trp Glu Gly
            660                 665                 670

Gln
```

<210> 332

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 329 of 3077

<211> 302
<212> PRT
<213> Pantoea ananatis

<400> 332

```
Met Thr Val Cys Ala Lys Lys His Val His Leu Thr Arg Asp Ala Ala
1               5                  10                  15

Glu Gln Leu Leu Ala Asp Ile Asp Arg Arg Leu Asp Gln Leu Leu Pro
            20                  25                  30

Val Glu Gly Glu Arg Asp Val Val Gly Ala Ala Met Arg Glu Gly Ala
        35                  40                  45

Leu Ala Pro Gly Lys Arg Ile Arg Pro Met Leu Leu Leu Leu Thr Ala
    50                  55                  60

Arg Asp Leu Gly Cys Ala Val Ser His Asp Gly Leu Leu Asp Leu Ala
65                  70                  75                  80

Cys Ala Val Glu Met Val His Ala Ala Ser Leu Ile Leu Asp Asp Met
                85                  90                  95

Pro Cys Met Asp Asp Ala Lys Leu Arg Arg Gly Arg Pro Thr Ile His
            100                 105                 110

Ser His Tyr Gly Glu His Val Ala Ile Leu Ala Ala Val Ala Leu Leu
        115                 120                 125

Ser Lys Ala Phe Gly Val Ile Ala Asp Ala Asp Gly Leu Thr Pro Leu
    130                 135                 140

Ala Lys Asn Arg Ala Val Ser Glu Leu Ser Asn Ala Ile Gly Met Gln
145                 150                 155                 160

Gly Leu Val Gln Gly Gln Phe Lys Asp Leu Ser Glu Gly Asp Lys Pro
                165                 170                 175

Arg Ser Ala Glu Ala Ile Leu Met Thr Asn His Phe Lys Thr Ser Thr
            180                 185                 190

Leu Phe Cys Ala Ser Met Gln Met Ala Ser Ile Val Ala Asn Ala Ser
        195                 200                 205

Ser Glu Ala Arg Asp Cys Leu His Arg Phe Ser Leu Asp Leu Gly Gln
    210                 215                 220

Ala Phe Gln Leu Leu Asp Asp Leu Thr Asp Gly Met Thr Asp Thr Gly
225                 230                 235                 240
```

```
Lys Asp Ser Asn Gln Asp Ala Gly Lys Ser Thr Leu Val Asn Leu Leu
                245                 250                 255

Gly Pro Arg Ala Val Glu Glu Arg Leu Arg Gln His Leu Gln Leu Ala
            260                 265                 270

Ser Glu His Leu Ser Ala Ala Cys Gln His Gly His Ala Thr Gln His
        275                 280                 285

Phe Ile Gln Ala Trp Phe Asp Lys Lys Leu Ala Ala Val Ser
    290                 295                 300
```

<210> 333
<211> 492
<212> PRT
<213> Pantoea ananatis

<400> 333

```
Met Lys Pro Thr Thr Val Ile Gly Ala Gly Phe Gly Gly Leu Ala Leu
1               5                   10                  15

Ala Ile Arg Leu Gln Ala Ala Gly Ile Pro Val Leu Leu Leu Glu Gln
            20                  25                  30

Arg Asp Lys Pro Gly Gly Arg Ala Tyr Val Tyr Glu Asp Gln Gly Phe
        35                  40                  45

Thr Phe Asp Ala Gly Pro Thr Val Ile Thr Asp Pro Ser Ala Ile Glu
    50                  55                  60

Glu Leu Phe Ala Leu Ala Gly Lys Gln Leu Lys Glu Tyr Val Glu Leu
65                  70                  75                  80

Leu Pro Val Thr Pro Phe Tyr Arg Leu Cys Trp Glu Ser Gly Lys Val
                85                  90                  95

Phe Asn Tyr Asp Asn Asp Gln Thr Arg Leu Glu Ala Gln Ile Gln Gln
            100                 105                 110

Phe Asn Pro Arg Asp Val Glu Gly Tyr Arg Gln Phe Leu Asp Tyr Ser
        115                 120                 125

Arg Ala Val Phe Lys Glu Gly Tyr Leu Lys Leu Gly Thr Val Pro Phe
    130                 135                 140

Leu Ser Phe Arg Asp Met Leu Arg Ala Ala Pro Gln Leu Ala Lys Leu
145                 150                 155                 160
```

```
Gln Ala Trp Arg Ser Val Tyr Ser Lys Val Ala Ser Tyr Ile Glu Asp
                165                 170                 175

Glu His Leu Arg Gln Ala Phe Ser Phe His Ser Leu Leu Val Gly Gly
            180                 185                 190

Asn Pro Phe Ala Thr Ser Ser Ile Tyr Thr Leu Ile His Ala Leu Glu
        195                 200                 205

Arg Glu Trp Gly Val Trp Phe Pro Arg Gly Gly Thr Gly Ala Leu Val
    210                 215                 220

Gln Gly Met Ile Lys Leu Phe Gln Asp Leu Gly Gly Glu Val Val Leu
225                 230                 235                 240

Asn Ala Arg Val Ser His Met Glu Thr Thr Gly Asn Lys Ile Glu Ala
                245                 250                 255

Val His Leu Glu Asp Gly Arg Arg Phe Leu Thr Gln Ala Val Ala Ser
            260                 265                 270

Asn Ala Asp Val Val His Thr Tyr Arg Asp Leu Leu Ser Gln His Pro
        275                 280                 285

Ala Ala Val Lys Gln Ser Asn Lys Leu Gln Thr Lys Arg Met Ser Asn
    290                 295                 300

Ser Leu Phe Val Leu Tyr Phe Gly Leu Asn His His His Asp Gln Leu
305                 310                 315                 320

Ala His His Thr Val Cys Phe Gly Pro Arg Tyr Arg Glu Leu Ile Asp
                325                 330                 335

Glu Ile Phe Asn His Asp Gly Leu Ala Glu Asp Phe Ser Leu Tyr Leu
            340                 345                 350

His Ala Pro Cys Val Thr Asp Ser Ser Leu Ala Pro Glu Gly Cys Gly
        355                 360                 365

Ser Tyr Tyr Val Leu Ala Pro Val Pro His Leu Gly Thr Ala Asn Leu
    370                 375                 380

Asp Trp Thr Val Glu Gly Pro Lys Leu Arg Asp Arg Ile Phe Ala Tyr
385                 390                 395                 400

Leu Glu Gln His Tyr Met Pro Gly Leu Arg Ser Gln Leu Val Thr His
                405                 410                 415
```

Arg Met Phe Thr Pro Phe Asp Phe Arg Asp Gln Leu Asn Ala Tyr His
420 425 430

Gly Ser Ala Phe Ser Val Glu Pro Val Leu Thr Gln Ser Ala Trp Phe
435 440 445

Arg Pro His Asn Arg Asp Lys Thr Ile Thr Asn Leu Tyr Leu Val Gly
450 455 460

Ala Gly Thr His Pro Gly Ala Gly Ile Pro Gly Val Ile Gly Ser Ala
465 470 475 480

Lys Ala Thr Ala Gly Leu Met Leu Glu Asp Leu Ile
485 490

<210> 334
<211> 309
<212> PRT
<213> Pantoea ananatis

<400> 334
Met Asn Asn Pro Ser Leu Leu Asn His Ala Val Glu Thr Met Ala Val
1 5 10 15

Gly Ser Lys Ser Phe Ala Thr Ala Ser Lys Leu Phe Asp Ala Lys Thr
20 25 30

Arg Arg Ser Val Leu Met Leu Tyr Ala Trp Cys Arg His Cys Asp Asp
35 40 45

Val Ile Asp Asp Gln Thr Leu Gly Phe Gln Ala Arg Gln Pro Ala Leu
50 55 60

Gln Thr Pro Glu Gln Arg Leu Met Gln Leu Glu Met Lys Thr Arg Gln
65 70 75 80

Ala Tyr Ala Gly Ser Gln Met His Glu Pro Ala Phe Ala Ala Phe Gln
85 90 95

Glu Val Ala Met Ala His Asp Ile Ala Pro Ala Tyr Ala Phe Asp His
100 105 110

Leu Glu Gly Phe Ala Met Asp Val Arg Glu Ala Gln Tyr Ser Gln Leu
115 120 125

Asp Asp Thr Leu Arg Tyr Cys Tyr His Val Ala Gly Val Val Gly Leu
130 135 140

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 333 of 3077

```
Met Met Ala Gln Ile Met Gly Val Arg Asp Asn Ala Thr Leu Asp Arg
145                 150                 155                 160

Ala Cys Asp Leu Gly Leu Ala Phe Gln Leu Thr Asn Ile Ala Arg Asp
                165                 170                 175

Ile Val Asp Asp Ala His Ala Gly Arg Cys Tyr Leu Pro Ala Ser Trp
            180                 185                 190

Leu Glu His Glu Gly Leu Asn Lys Glu Asn Tyr Ala Ala Pro Glu Asn
        195                 200                 205

Arg Gln Ala Leu Ser Arg Ile Ala Arg Arg Leu Val Gln Glu Ala Glu
    210                 215                 220

Pro Tyr Tyr Leu Ser Ala Thr Ala Gly Leu Ala Gly Leu Pro Leu Arg
225                 230                 235                 240

Ser Ala Trp Ala Ile Ala Thr Ala Lys Gln Val Tyr Arg Lys Ile Gly
                245                 250                 255

Val Lys Val Glu Gln Ala Gly Gln Gln Ala Trp Asp Gln Arg Gln Ser
            260                 265                 270

Thr Thr Thr Pro Glu Lys Leu Thr Leu Leu Leu Ala Ala Ser Gly Gln
        275                 280                 285

Ala Leu Thr Ser Arg Met Arg Ala His Pro Pro Arg Pro Ala His Leu
    290                 295                 300

Trp Gln Arg Pro Leu
305


<210> 335
<211> 746
<212> PRT
<213> Xanthophyllomyces dendrorhous

<400> 335
Met Ala Thr Leu Ser Asp Leu Val Ile Leu Leu Leu Gly Ala Leu Leu
1               5                   10                  15

Ala Leu Gly Phe Tyr Asn Lys Asp Arg Leu Leu Gly Ser Ser Ser Ser
            20                  25                  30

Ser Ala Ser Thr Thr Ser Gly Ser Ser Ala Ala Thr Ala Asn Gly Ser
        35                  40                  45
```


LCY Biotechnology Holding, Inc.
Ex. 1002
Page 334 of 3077

```
Lys Pro Thr Tyr Ser Asn Gly Asn Gly Asn Ala Phe Lys Gly Asp Pro
    50                  55                  60

Arg Asp Phe Val Ala Arg Met Lys Asp Gln Lys Lys Arg Leu Ala Val
65                  70                  75                  80

Phe Tyr Gly Ser Gln Thr Gly Thr Ala Glu Glu Tyr Ala Thr Arg Ile
                85                  90                  95

Ala Lys Glu Ala Lys Ser Arg Phe Gly Val Ser Ser Leu Val Cys Asp
            100                 105                 110

Ile Glu Glu Tyr Asp Phe Glu Lys Leu Asp Gln Val Pro Glu Asp Cys
        115                 120                 125

Ala Ile Val Phe Cys Met Ala Thr Tyr Gly Glu Gly Glu Pro Thr Asp
    130                 135                 140

Asn Ala Val Gln Phe Ile Glu Met Ile Ser Gln Asp Asp Pro Glu Phe
145                 150                 155                 160

Ser Glu Gly Ser Thr Leu Asp Gly Leu Lys Tyr Val Val Phe Gly Leu
                165                 170                 175

Gly Asn Lys Thr Tyr Glu Gln Tyr Asn Val Val Gly Arg Gln Leu Asp
            180                 185                 190

Ala Arg Leu Thr Ala Leu Gly Ala Thr Arg Val Gly Glu Arg Gly Glu
        195                 200                 205

Gly Asp Asp Asp Lys Ser Met Glu Glu Asp Tyr Leu Ala Trp Lys Asp
    210                 215                 220

Asp Met Phe Ala Ala Leu Ala Thr Thr Leu Ser Phe Glu Glu Gly Ala
225                 230                 235                 240

Ser Gly Glu Thr Pro Asp Phe Val Val Thr Glu Val Pro Asn His Pro
                245                 250                 255

Ile Glu Lys Val Phe Gln Gly Glu Leu Ser Ser Arg Ala Leu Leu Gly
            260                 265                 270

Ser Lys Gly Val His Asp Ala Lys Asn Pro Tyr Ala Ser Pro Val Leu
        275                 280                 285

Ala Cys Arg Glu Leu Phe Thr Gly Gly Asp Arg Asn Cys Ile His Leu
    290                 295                 300
```


LCY Biotechnology Holding, Inc.
Ex. 1002
Page 335 of 3077

```
Glu Phe Asp Ile Thr Gly Ser Gly Ile Thr Tyr Gln Thr Gly Asp His
305                 310                 315                 320

Val Ala Val Trp Pro Ser Asn Pro Asp Val Glu Val Glu Arg Leu Leu
                325                 330                 335

Ala Val Leu Gly Leu Thr Ser Pro Glu Lys Arg Arg Met Ile Ile Gln
            340                 345                 350

Val Val Ser Leu Asp Pro Thr Leu Ala Lys Val Pro Phe Pro Thr Pro
        355                 360                 365

Thr Thr Tyr Asp Ala Val Phe Arg His Tyr Leu Asp Ile Ser Ala Val
    370                 375                 380

Ala Ser Arg Gln Thr Leu Ala Val Leu Ala Lys Tyr Ala Pro Ser Glu
385                 390                 395                 400

Gln Ala Ala Glu Phe Leu Thr Arg Leu Gly Thr Asp Lys Gln Ala Tyr
                405                 410                 415

His Thr Glu Val Val Gly Gly His Leu Arg Leu Ala Glu Val Leu Gln
            420                 425                 430

Leu Ala Ala Gly Asn Asp Ile Thr Val Met Pro Thr Ala Glu Asn Thr
        435                 440                 445

Thr Val Trp Asn Ile Pro Phe Asp His Val Val Ser Asp Val Ser Arg
    450                 455                 460

Leu Gln Pro Arg Phe Tyr Ser Ile Ser Ser Ser Pro Lys Leu His Pro
465                 470                 475                 480

Asn Ser Ile His Val Thr Ala Val Ile Leu Lys Tyr Glu Ser Gln Ala
                485                 490                 495

Thr Asp Arg His Pro Ala Arg Trp Val Phe Gly Leu Gly Thr Asn Tyr
            500                 505                 510

Leu Leu Asn Val Lys Gln Ala Ala Asn Asn Glu Thr Thr Pro Met Ile
        515                 520                 525

Ser Asp Gly Gln Asp Asp Val Pro Glu His Val Ser Ala Pro Lys Tyr
    530                 535                 540

Thr Leu Glu Gly Pro Arg Gly Ser Tyr Lys His Asp Asp Gln Leu Phe
545                 550                 555                 560
```

```
Lys Val Pro Ile His Val Arg Arg Ser Thr Phe Arg Leu Pro Thr Ser
                565                 570                 575

Pro Lys Ile Pro Val Ile Met Ile Gly Pro Gly Thr Gly Val Ala Pro
            580                 585                 590

Phe Arg Gly Phe Ile Gln Glu Arg Ile Ala Leu Ala Arg Arg Ser Ile
        595                 600                 605

Ala Lys Asn Gly Pro Asp Ala Leu Ala Asp Trp Ala Pro Ile Tyr Leu
    610                 615                 620

Phe Tyr Gly Ser Arg Asp Glu Gln Asp Phe Leu Tyr Ala Glu Glu Trp
625                 630                 635                 640

Pro Ala Tyr Glu Ala Glu Leu Gln Gly Lys Phe Lys Ile His Val Ala
                645                 650                 655

Phe Ser Arg Ser Gly Pro Arg Lys Pro Asp Gly Ser Lys Ile Tyr Val
            660                 665                 670

Gln Asp Leu Leu Trp Asp Gln Lys Glu Val Ile Lys Ser Ala Ile Val
        675                 680                 685

Glu Lys Arg Ala Ser Val Tyr Ile Cys Gly Asp Gly Arg Asn Met Ser
    690                 695                 700

Lys Asp Val Glu Gln Lys Leu Ala Ala Met Leu Ala Glu Ser Lys Asn
705                 710                 715                 720

Gly Ser Ala Ala Val Glu Gly Ala Ala Glu Val Lys Ser Leu Lys Glu
                725                 730                 735

Arg Ser Arg Leu Leu Met Asp Val Trp Ser
            740                 745


<210> 336
<211> 557
<212> PRT
<213> Xanthophyllomyces dendrorhous

<400> 336
Met Phe Ile Leu Val Leu Leu Thr Gly Ala Leu Gly Leu Ala Ala Phe
1               5                   10                  15

Ser Trp Ala Ser Ile Ala Phe Phe Ser Leu Tyr Leu Ala Pro Arg Arg
            20                  25                  30
```

```
Ser Ser Leu Tyr Asn Leu Gln Gly Pro Asn His Thr Asn Tyr Phe Thr
        35                  40                  45

Gly Asn Phe Leu Asp Ile Leu Ser Ala Arg Thr Gly Glu Glu His Ala
    50                  55                  60

Lys Tyr Arg Glu Lys Tyr Gly Ser Thr Leu Arg Phe Ala Gly Ile Ala
65                  70                  75                  80

Gly Ala Pro Val Leu Asn Ser Thr Asp Pro Lys Val Phe Asn His Val
                85                  90                  95

Met Lys Glu Ala Tyr Asp Tyr Pro Lys Pro Gly Met Ala Ala Arg Val
            100                 105                 110

Leu Arg Ile Ala Thr Gly Asp Gly Val Val Thr Ala Glu Gly Glu Ala
        115                 120                 125

His Lys Arg His Arg Arg Ile Met Ile Pro Ser Leu Ser Ala Gln Ala
    130                 135                 140

Val Lys Ser Met Val Pro Ile Phe Leu Glu Lys Gly Met Glu Leu Val
145                 150                 155                 160

Asp Lys Met Met Glu Asp Ala Ala Glu Lys Asp Met Ala Val Gly Glu
                165                 170                 175

Ser Ala Gly Glu Lys Lys Ala Thr Arg Leu Glu Thr Glu Gly Val Asp
            180                 185                 190

Val Lys Asp Trp Val Gly Arg Ala Thr Leu Asp Val Met Ala Leu Ala
        195                 200                 205

Gly Phe Asp Tyr Lys Ser Asp Ser Leu Gln Asn Lys Thr Asn Glu Leu
    210                 215                 220

Tyr Val Ala Phe Val Gly Leu Thr Asp Gly Phe Ala Pro Thr Leu Asp
225                 230                 235                 240

Ser Phe Lys Ala Ile Met Trp Asp Phe Val Pro Tyr Phe Arg Thr Met
                245                 250                 255

Lys Arg Arg His Glu Ile Pro Leu Thr Gln Gly Leu Ala Val Ser Arg
            260                 265                 270

Arg Val Gly Ile Glu Leu Met Glu Gln Lys Lys Gln Ala Val Leu Gly
        275                 280                 285
```

```
Ser Ala Ser Asp Gln Ala Val Asp Lys Lys Asp Val Gln Gly Arg Asp
    290                 295                 300

Ile Leu Ser Leu Leu Val Arg Ala Asn Ile Ala Ala Asn Leu Pro Glu
305                 310                 315                 320

Ser Gln Lys Leu Ser Asp Glu Glu Val Leu Ala Gln Ile Ser Asn Leu
                325                 330                 335

Leu Phe Ala Gly Tyr Glu Thr Ser Ser Thr Val Leu Thr Trp Met Phe
            340                 345                 350

His Arg Leu Ser Glu Asp Lys Ala Val Gln Asp Lys Leu Arg Glu Glu
        355                 360                 365

Ile Cys Gln Ile Asp Thr Asp Met Pro Thr Leu Asp Glu Leu Asn Ala
    370                 375                 380

Leu Pro Tyr Leu Glu Ala Phe Val Lys Glu Ser Leu Arg Leu Asp Pro
385                 390                 395                 400

Pro Ser Pro Tyr Ala Asn Arg Glu Cys Leu Lys Asp Glu Asp Phe Ile
                405                 410                 415

Pro Leu Ala Glu Pro Val Ile Gly Arg Asp Gly Ser Val Ile Asn Glu
            420                 425                 430

Val Arg Ile Thr Lys Gly Thr Met Val Met Leu Pro Leu Phe Asn Ile
        435                 440                 445

Asn Arg Ser Lys Phe Ile Tyr Gly Glu Asp Ala Glu Glu Phe Arg Pro
    450                 455                 460

Glu Arg Trp Leu Glu Asp Val Thr Asp Ser Leu Asn Ser Ile Glu Ala
465                 470                 475                 480

Pro Tyr Gly His Gln Ala Ser Phe Ile Ser Gly Pro Arg Ala Cys Phe
                485                 490                 495

Gly Trp Arg Phe Ala Val Ala Glu Met Lys Ala Phe Leu Phe Val Thr
            500                 505                 510

Leu Arg Arg Val Gln Phe Glu Pro Ile Ile Ser His Pro Glu Tyr Glu
        515                 520                 525

His Ile Thr Leu Ile Ile Ser Arg Pro Arg Ile Val Gly Arg Glu Lys
```

530 535 540

Glu Gly Tyr Gln Met Arg Leu Gln Val Lys Pro Val Glu
545 550 555

<210> 337
<211> 403
<212> PRT
<213> Candida viswanathii

<400> 337
Met Thr Leu Pro Pro Val Tyr Ile Val Ser Thr Ala Arg Thr Pro Ile
1 5 10 15

Gly Ser Phe Gln Gly Ser Leu Ser Ser Leu Thr Tyr Ser Asp Leu Gly
20 25 30

Ala His Ala Val Lys Ala Ala Leu Ala Lys Val Pro Gln Ile Lys Pro
35 40 45

Gln Asp Val Asp Glu Ile Val Phe Gly Gly Val Leu Gln Ala Asn Val
50 55 60

Gly Gln Ala Pro Ala Arg Gln Val Ala Leu Lys Ala Gly Leu Pro Asp
65 70 75 80

Ser Ile Ile Ala Ser Thr Ile Asn Lys Val Cys Ala Ser Gly Met Lys
85 90 95

Ala Val Ile Ile Gly Ala Gln Asn Ile Ile Cys Gly Thr Ser Asp Ile
100 105 110

Val Val Val Gly Gly Ala Glu Ser Met Ser Asn Thr Pro Tyr Tyr Leu
115 120 125

Pro Ser Ala Arg Ser Gly Ala Arg Tyr Gly Asp Ala Val Met Val Asp
130 135 140

Gly Val Gln Lys Asp Gly Leu Leu Asp Val Tyr Glu Glu Lys Leu Met
145 150 155 160

Gly Val Ala Ala Glu Lys Cys Ala Lys Asp His Gly Phe Ser Arg Glu
165 170 175

Asp Gln Asp Asn Phe Ala Ile Asn Ser Tyr Lys Lys Ala Gly Lys Ala
180 185 190

Leu Ser Glu Gly Lys Phe Lys Ser Glu Ile Ala Pro Val Thr Ile Lys
195 200 205

Gly Phe Arg Gly Lys Pro Asp Thr Val Ile Glu Asn Asp Glu Glu Ile
210 215 220

Gly Lys Phe Asn Glu Asp Arg Leu Lys Ser Ala Arg Thr Val Phe Gln
225 230 235 240

Lys Glu Asn Gly Thr Val Thr Ala Pro Asn Ala Ser Lys Leu Asn Asp
245 250 255

Gly Gly Ala Ala Leu Val Leu Val Ser Glu Ala Lys Leu Lys Gln Leu
260 265 270

Gly Leu Lys Pro Leu Ala Lys Ile Ser Gly Trp Gly Glu Ala Ala Arg
275 280 285

Thr Pro Phe Asp Phe Thr Ile Ala Pro Ala Leu Ala Val Pro Lys Ala
290 295 300

Val Lys His Ala Gly Leu Thr Val Asp Arg Val Asp Phe Phe Glu Leu
305 310 315 320

Asn Glu Ala Phe Ser Val Val Gly Leu Ala Asn Ala Glu Leu Val Lys
325 330 335

Ile Pro Leu Glu Lys Leu Asn Val Tyr Gly Gly Ala Val Ala Met Gly
340 345 350

His Pro Leu Gly Cys Ser Gly Ala Arg Ile Ile Val Thr Leu Leu Ser
355 360 365

Val Leu Thr Gln Glu Gly Gly Arg Phe Gly Ala Ala Gly Val Cys Asn
370 375 380

Gly Gly Gly Gly Ala Ser Ala Ile Val Ile Glu Lys Ile Asp Ser Asp
385 390 395 400

Ala Lys Leu

<210> 338
<211> 449
<212> PRT
<213> Candida viswanathii

<400> 338
Met Thr Asn Ala Pro Gln Asn Ile Gly Ile Lys Gly Ile Glu Val Tyr
1 5 10 15

```
Ile Pro Gly Gln Ala Val Asn Gln Ser Asp Leu Glu Lys Phe Asp Gly
            20                  25                  30

Ile Pro Ala Gly Lys Tyr Thr Ile Gly Leu Gly Gln Thr Asn Met Ala
        35                  40                  45

Phe Val Asn Asp Arg Glu Asp Ile Tyr Ser Ile Ala Leu Thr Val Val
    50                  55                  60

Ser Arg Leu Ile Lys His Tyr Asn Val Asp Thr Asn Asn Val Gly Arg
65                  70                  75                  80

Leu Glu Val Gly Thr Glu Thr Leu Leu Asp Lys Ser Lys Ser Val Lys
                85                  90                  95

Ser Val Leu Met Gln Leu Phe Pro Asp Asn Asn Asp Ile Glu Gly Ile
            100                 105                 110

Asp Thr Val Asn Ala Cys Tyr Gly Gly Thr Ser Ala Val Ile Asn Ala
        115                 120                 125

Ile Asn Trp Ile Glu Ser Ser Ser Trp Asp Gly Arg Asp Ala Ile Val
    130                 135                 140

Val Ala Gly Asp Ile Ala Ile Tyr Asp Lys Gly Ala Ala Arg Pro Thr
145                 150                 155                 160

Gly Gly Ala Gly Ser Ile Ala Met Leu Ile Gly Pro Asp Ala Pro Ile
                165                 170                 175

Val Phe Asp Ser Val Arg Gly Ser Phe Met Glu His Ala Tyr Asp Phe
            180                 185                 190

Tyr Lys Pro Asp Phe Thr Ser Glu Tyr Pro Val Val Asp Gly His Phe
        195                 200                 205

Ser Leu Ser Cys Tyr Val Lys Ala Val Asp Asn Cys Tyr Lys Asn Tyr
    210                 215                 220

Ser Lys Lys Val Thr Gly Asn Val Asp Lys Thr Val Gly Val Tyr Asp
225                 230                 235                 240

His Phe Asp Tyr Asn Ala Phe His Val Pro Thr Cys Lys Leu Val Thr
                245                 250                 255

Lys Ser Tyr Ala Arg Leu Leu Tyr Asn Asp Tyr Lys Ser Asp Pro Ser
            260                 265                 270
```

```
Lys Phe Ala Asp Leu Ile Asp Glu Ser Thr Arg Gln His Ile Asp Ser
        275                 280                 285

Leu Ser Tyr Glu Ala Ser Leu Thr Asp Lys Val Leu Glu Lys Thr Phe
    290                 295                 300

Val Thr Leu Ala Lys Glu Glu Thr Lys Lys Arg Val Gln Pro Ala Leu
305                 310                 315                 320

Gln Val Pro Thr Asn Thr Gly Asn Met Tyr Thr Ala Ser Ser Trp Val
                325                 330                 335

Ser Leu Ala Ser Leu Leu Tyr Tyr Val Gly Ala Glu Asn Leu Lys Glu
            340                 345                 350

Lys Arg Ile Gly Leu Phe Ser Tyr Gly Ser Gly Leu Ala Ser Thr Leu
        355                 360                 365

Leu Ser Val Thr Val Val Gly Asp Val Ser Pro Ile Thr Lys Val Leu
    370                 375                 380

Asp Phe Asp Tyr Lys Leu Gly Glu Gly Arg Lys Ile Gln Ser Pro Glu
385                 390                 395                 400

Glu Tyr Leu Ala Ala Ile Glu Leu Arg Glu Lys Ala His Leu Gln Lys
                405                 410                 415

Ser Phe Lys Pro Gln Gly Ser Leu Asp Asn Leu Ser Gln Gly Thr Tyr
            420                 425                 430

Tyr Leu Thr Glu Val Asp Asp Lys Phe Arg Arg Ser Tyr Ala Ile Lys
        435                 440                 445

Glu
```

<210> 339
<211> 1039
<212> PRT
<213> Candida viswanathii

```
<400> 339
Met Leu Ser Phe Ile Thr Glu Ala Ser Gly Arg Ile Ala Gln Thr Ala
1               5                   10                  15

Ala His Arg Pro Ile His Phe Met Val Ile Pro Ala Leu Leu Ala Ser
            20                  25                  30
```

```
Ile Ala Tyr Leu Ser Ile Ile Asp Asp Tyr Val Pro Glu His Ile Arg
        35                  40                  45

Ala Gln Tyr Met Ser Gly Val Ser Tyr Phe His Pro Gln Gly Ser Ala
    50                  55                  60

Ser Asp Leu Asp Lys Trp Ile Glu Ile His Asp Thr Thr Gln Tyr Ala
65                  70                  75                  80

Asn Ala Asn Gln Ile Ser Val Ile Pro Leu Arg Phe Arg Arg Phe His
                85                  90                  95

Asp Ser Ile Pro Gln Ile Ala Asn Ala Ile Lys Ile Ser Asn Asn Glu
            100                 105                 110

Gln Ile Leu Ile Val Pro Ser Asp Lys Ala Glu Ser Thr Val Glu Gly
        115                 120                 125

Leu Ser Glu Ile Thr Glu Asn Gly Ile Thr Trp Arg Ala Arg Asn Asn
    130                 135                 140

Asp Lys Leu Ser Lys Tyr Tyr Asp Tyr Ala Arg Tyr Gly Leu Leu Arg
145                 150                 155                 160

Ile Gln Asp Ala Ile His Asn Ala Asp Asn Phe Asp Ile Leu Leu Val
                165                 170                 175

Phe Val Ala Tyr Leu Gly Met Trp Tyr Ser Leu Ile Lys Val Phe Ile
            180                 185                 190

Asp Met Arg Arg Ile Gly Ser Lys Phe Trp Leu Ala Phe Gly Thr Leu
        195                 200                 205

Thr Ser Ser Thr Phe Ser Phe Leu Phe Ala Leu Val Ile Ser Asn Lys
    210                 215                 220

Phe Leu Asp Ala Lys Val Ser Leu Arg Ser Leu Ser Glu Cys Ile Pro
225                 230                 235                 240

Phe Leu Val Ala Ile Ile Gly Phe Lys His Lys Val Ala Ile Thr Thr
                245                 250                 255

Ser Val Ala Gln Ser Ser Thr Ser Ser Pro Glu Asp Val Pro His Val
            260                 265                 270

Val Gly Lys Ala Val Ser Asp Gln Cys Leu Phe Ile Leu Arg Asp His
        275                 280                 285
```


LCY Biotechnology Holding, Inc.
Ex. 1002
Page 344 of 3077

```
Leu Val Val Ile Ile Gly Phe Leu Ala Cys Ala Ala Tyr Gly Asn Glu
    290                 295                 300

Leu Lys Gly Leu Arg Asn Phe Cys Ile Leu Gly Ala Leu Ile Leu Ser
305                 310                 315                 320

Phe Asp Ile Val Ser Val Tyr Thr Phe Phe Ser Ala Ile Leu Ala Leu
                325                 330                 335

Lys Val Glu Ile Asn Arg Ala Arg Arg Thr Gln Asp Leu Gln His Val
            340                 345                 350

Leu Glu Glu Asp Gly Ile Ser Ser Leu Val Ala Ala Arg Val Ala Glu
        355                 360                 365

Arg Asn Ala Thr Ile Glu His Pro Asn Glu Thr Asn Phe Phe Ser Ser
    370                 375                 380

Asn Asn Ser Ser Ile Val Tyr Phe Lys Val Ile Met Ser Leu Gly Phe
385                 390                 395                 400

Phe Ala Phe His Ala Phe Trp Leu Gly Ser Ser Trp Leu Tyr Asn Thr
                405                 410                 415

Ser Asp Gly Gly Ser His Gly Ser Phe Ser Phe Leu Ser Asn Ile Pro
            420                 425                 430

Leu Leu Thr Gln Asp Ile Ser Asn Ser Ile Pro Ile Gly Arg Arg Gly
        435                 440                 445

Thr Val Val Thr Ile Leu Pro Thr Thr Phe Tyr Met Pro Ser Gly Ile
    450                 455                 460

Ile Ile Gln Phe Glu Asp Met Val Tyr Leu Ala Leu Ser Lys Ile Ser
465                 470                 475                 480

Ser Ala Ile Arg Asp Ser Leu Ile Ser Lys Cys Ile Val Phe Ala Leu
                485                 490                 495

Thr Ile Ser Ile Val Thr Asn Ile Tyr Phe Leu Asn Ala Ala Arg Phe
            500                 505                 510

Gln Val Ser Ala Thr Arg Lys Leu Ile Asp Gln Glu Met Ser Arg Pro
        515                 520                 525

Lys Gln Ala Ala Ala Ala Ala Ala Ala Ala Pro Ser Ala Ala Lys Ser
```

530 535 540

Val Ala Pro Glu Glu Asp Glu Asp Glu Thr Ser Ser Glu Glu Leu Glu
545 550 555 560

Ile Lys Ala Pro Val Lys Pro Leu Ser Leu Glu Glu Cys Thr Arg Ile
565 570 575

Leu Lys Glu Gly Lys Val Lys Thr Leu Ser Asn Ala Glu Val Ser Ser
580 585 590

Leu Val Val Gly Gly Lys Leu Pro Leu Tyr Ala Leu Glu Lys Gln Leu
595 600 605

Gly Asp Asn Lys Arg Ala Val Ala Val Arg Arg Lys Ala Ile Ala Lys
610 615 620

Leu Ala Asn Ala Pro Val Leu Glu Thr Asn Arg Leu Pro Tyr Ser His
625 630 635 640

Tyr Asp Tyr Asp Arg Val Phe Gly Ala Cys Cys Glu Asn Val Ile Gly
645 650 655

Tyr Met Pro Ile Pro Val Gly Val Ala Gly Pro Leu Val Ile Asp Gly
660 665 670

Lys Pro Tyr His Ile Pro Met Ala Thr Thr Glu Gly Cys Leu Val Ala
675 680 685

Ser Thr Met Arg Gly Cys Lys Ala Ile Asn Ala Gly Gly Gly Val Glu
690 695 700

Thr Val Leu Thr Lys Asp Gly Met Thr Arg Gly Pro Cys Val Arg Phe
705 710 715 720

Pro Thr Leu Lys Arg Ala Gly Ala Ala Lys Leu Trp Ile Asp Ser Glu
725 730 735

Glu Gly Gln Ile Thr Ile Lys Lys Ala Phe Asn Ser Thr Ser Arg Phe
740 745 750

Ala Arg Leu Gln His Ile Gln Thr Ala Leu Ala Gly Thr Ser Leu Phe
755 760 765

Ile Arg Phe Arg Thr Thr Thr Gly Asp Ala Met Gly Met Asn Met Ile
770 775 780

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 346 of 3077

```
Ser Lys Gly Val Glu Tyr Ser Leu Lys His Met Val Glu Glu Cys Gly
785                 790                 795                 800

Trp Asp Asp Met Glu Val Ile Ser Val Ser Gly Asn Tyr Cys Thr Asp
                805                 810                 815

Lys Lys Pro Ala Ala Ile Asn Trp Ile Glu Gly Arg Gly Lys Ser Ile
            820                 825                 830

Val Ala Ala Ala Thr Ile Pro Ala Asp Val Val Thr Lys Val Leu Lys
        835                 840                 845

Ser Asp Val Asp Ala Leu Val Glu Leu Asn Val Ser Lys Asn Leu Val
    850                 855                 860

Gly Ser Ala Met Ala Gly Ser Val Gly Gly Phe Asn Ala His Ala Ala
865                 870                 875                 880

Asn Leu Val Thr Ala Val Tyr Leu Ala Cys Gly Gln Asp Pro Ala Gln
                885                 890                 895

Asn Val Glu Ser Ser Asn Cys Ile Thr Leu Met Glu Lys Asp Lys Gln
            900                 905                 910

Thr Gly Asp Leu Val Ile Ser Val Ser Met Pro Ser Ile Glu Val Gly
        915                 920                 925

Thr Ile Gly Gly Gly Thr Ile Leu Glu Pro Gln Ala Ala Met Leu Asp
    930                 935                 940

Leu Leu Gly Val Arg Gly Pro His Pro Thr Asn Pro Gly Asp Asn Ala
945                 950                 955                 960

Arg Gln Leu Ala Lys Ile Val Ala Ser Ala Val Leu Ala Ala Glu Leu
                965                 970                 975

Ser Leu Cys Ser Ala Leu Ala Ala Gly His Leu Val Gln Ser His Met
            980                 985                 990

Gln His Asn Arg Lys Gly Ala Thr  Pro Ala Ala Ala Pro  Ala Ile Ser
        995                 1000                1005

Asn Gly  Ser Ala Lys Gly Thr  Lys Thr Asn Gly Ser  Ile Asn Gly
    1010                1015                1020

Lys Asp  Leu Lys Arg Leu Lys  Asp Gly Ser Val Thr  Cys Ile Lys
    1025                1030                1035
```


LCY Biotechnology Holding, Inc.
Ex. 1002
Page 347 of 3077

Ser

<210> 340
<211> 432
<212> PRT
<213> Candida viswanathii

<400> 340
Met Ser Val Ser Pro Phe Val Val Ser Ala Pro Gly Lys Val Ile Ile
1 5 10 15

Phe Gly Glu His Ser Ala Val Tyr Gly Lys Pro Ala Ile Ala Ala Ala
20 25 30

Leu Ser Leu Arg Cys Tyr Leu Leu Val Ser Pro Ser Val Asp Asp Ala
35 40 45

Asn Thr Ile Arg Leu Gln Phe Pro Asp Ile Gln Leu Asp His Ser Trp
50 55 60

Asp Ile Asn Asp Ile Pro Trp Asp Glu Ile Lys Pro Phe Val Lys Tyr
65 70 75 80

Asp Ala Asn Asn Lys Pro Leu Thr Ser Ser Glu Leu Val Pro Glu Ile
85 90 95

Val Asp Lys Leu Ser Pro Leu Leu Thr Asp Phe Asp Asn Lys Met His
100 105 110

Tyr Tyr Ala Cys Phe Cys Phe Leu Tyr Leu Tyr Val Asn Leu Cys Thr
115 120 125

Ala Glu Thr Pro Gly Thr Thr Phe Ile Val Arg Ser Thr Leu Pro Ile
130 135 140

Gly Ala Gly Leu Gly Ser Ser Ala Ser Thr Ser Val Cys Leu Ser Ser
145 150 155 160

Ala Leu Ala Leu Leu Gly Gly Trp Ile Ser Glu Pro Ser Ile Ser Ala
165 170 175

Thr Asp Lys Ile Leu Asn Glu Asp Ile Pro Asp Leu Glu Phe Val Asp
180 185 190

Lys Trp Ser Leu Ile Gly Glu Lys Cys Phe His Gly Asn Pro Ser Gly
195 200 205

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 348 of 3077

```
Ile Asp Asn Ala Val Ala Thr Phe Gly Gly Ala Val Met Phe Gln Arg
    210                 215                 220

Thr Ser Ala Pro Glu Gln Pro Ser Ile Arg Thr Asn Met Arg Asn Phe
225                 230                 235                 240

Pro Ala Ile Lys Leu Leu Leu Thr Asn Thr Lys Val Pro Lys Ser Thr
                245                 250                 255

Ala Asp Leu Val Ala Gly Val Gly Lys Leu Asn Ala Glu Phe Asn Pro
            260                 265                 270

Ile Thr Thr Ser Ile Leu Thr Ala Met Glu His Leu Ser Gln Glu Ala
        275                 280                 285

Tyr Lys Val Met Ile Thr Pro Gly Phe Gly Lys Asp Glu Thr Asn Thr
    290                 295                 300

Leu Arg Lys Leu Val Asn Ile Asn His Gly Leu Leu Val Ala Leu Gly
305                 310                 315                 320

Val Ser His Pro Ser Leu Glu Thr Val Lys Ile Ile Gly Asp Arg His
                325                 330                 335

Lys Ile Gly Ala Thr Lys Leu Thr Gly Ala Gly Gly Gly Gly Cys Ala
            340                 345                 350

Ile Thr Leu Val Asn Asp Asn Val Glu Glu Ser Ala Ile Gln Glu Ala
        355                 360                 365

Ile Lys Glu Phe Ala Gly Glu Gly Tyr Glu Ser Phe Glu Thr Ser Leu
    370                 375                 380

Gly Gly Lys Gly Val Gly Val Leu Phe Ser Gly Asp Val Ser Val Asp
385                 390                 395                 400

Gly Lys Phe Ser Gln Ser Val Phe Cys Asn Tyr Pro Asp Arg Ala Ser
                405                 410                 415

Ile Glu Asp Ala Leu Gly Met Ile Asn Val Lys Glu Trp Lys Phe Trp
            420                 425                 430


<210> 341
<211> 420
<212> PRT
<213> Candida viswanathii

<400> 341
Met Ser Lys Ala Phe Ser Ala Pro Gly Lys Ala Leu Leu Ala Gly Gly
```

```
1               5                   10                  15

Tyr Leu Val Leu Glu Pro Thr Tyr Asp Ala Tyr Val Thr Ala Leu Ser
            20                  25                  30

Ser Arg Met His Ala Ile Ile Thr Pro Gln Lys Pro Ala Ser Ile Ser
        35                  40                  45

Lys Ile Lys Ile Ser Ser Pro Gln Phe Ala Asn Gly Glu Trp Glu Tyr
    50                  55                  60

His Val Thr Ser Asn Glu Lys Pro Lys Asp Ile Lys Ser Arg Ser Asn
65                  70                  75                  80

Pro Phe Leu Glu Ala Thr Ile Phe Ile Val Leu Ser Tyr Ile Gln Pro
                85                  90                  95

Thr Glu Pro Phe Asp Leu Asp Leu Val Ile Tyr Ser Asp Pro Gly Tyr
            100                 105                 110

His Ser Gln Glu His Thr Thr Gln Lys Val Ser Ser Asn Gly Lys Lys
        115                 120                 125

Lys Phe Leu Tyr His Ser Arg Ala Ile Asn Asp Val Glu Lys Thr Gly
    130                 135                 140

Leu Gly Ser Ser Ala Gly Leu Val Ser Val Val Thr Thr Ser Leu Leu
145                 150                 155                 160

Ser Tyr Phe Ile Pro Gly Ile Glu Glu Ser Asn Lys Asp Met Leu His
                165                 170                 175

Asn Val Ala Gln Ile Ala His Cys Phe Ala Gln Lys Lys Ile Gly Ser
            180                 185                 190

Gly Phe Asp Val Ala Thr Ala Ile Tyr Gly Ser Ile Val Tyr Arg Arg
        195                 200                 205

Phe Gln Pro Ser Leu Ile Asn Asp Val Phe Glu Ile Leu Glu Glu Thr
    210                 215                 220

Pro Gly Arg Phe Pro Gly Ala Leu Lys Ser Leu Val Glu Ser Asn Trp
225                 230                 235                 240

Glu Phe Lys His Glu Arg Cys Val Leu Pro Pro Lys Ile Lys Leu Leu
                245                 250                 255
```


LCY Biotechnology Holding, Inc.
Ex. 1002
Page 350 of 3077

```
Met Gly Asp Ile Lys Gly Gly Ser Glu Thr Pro Lys Leu Val Ser Lys
            260                 265                 270

Ile Leu Gln Trp Lys Lys Asp Lys Pro Glu Glu Ser Gly Leu Val Tyr
        275                 280                 285

Asp Gln Leu Asn Ser Ala Asn Val Ala Phe Met Lys Lys Ile Ser Thr
    290                 295                 300

Leu Asn Glu Ser Ser Gln Val Gln Glu Ile Asp Glu Leu Ser Asp Tyr
305                 310                 315                 320

Ile Ser Ala Val Arg Lys Gly Leu Gln Glu Leu Thr Glu Lys Ser Lys
                325                 330                 335

Val Pro Val Glu Pro Pro Val Gln Thr Glu Leu Leu Asp Arg Ile Ala
            340                 345                 350

Lys Leu Pro Gly Cys Leu Gly Gly Val Val Pro Gly Ala Gly Gly Tyr
        355                 360                 365

Asp Ala Ile Ala Val Leu Val Leu Glu Lys Glu Val Glu Asn Phe Arg
    370                 375                 380

Asn Lys Thr Leu Glu Asn Pro Glu Tyr Tyr His Asn Val Tyr Trp Val
385                 390                 395                 400

Asp Leu Glu Glu Glu Thr Glu Gly Val Val Val Glu Asn Trp Glu Asp
                405                 410                 415

Tyr Ile Gly Leu
            420


<210> 342
<211> 369
<212> PRT
<213> Candida viswanathii

<400> 342
Met Tyr Ser Ala Ser Val Thr Ala Pro Val Asn Ile Ala Thr Leu Lys
1               5                   10                  15

Tyr Trp Gly Lys Arg Asp Lys Thr Leu Asn Leu Pro Thr Asn Ser Ser
            20                  25                  30

Ile Ser Val Thr Leu Ser Gln Asp Asp Leu Arg Thr Leu Thr Thr Ala
        35                  40                  45

Ala Ala Ser Thr Thr Phe Asp Lys Asp Gln Leu Trp Leu Asn Gly Lys
```

```
    50                  55                  60

Leu Glu Ser Leu Asp Thr Pro Arg Thr Gln Ala Cys Leu Val Asp Leu
65                  70                  75                  80

Arg Lys Leu Arg Ala Asp Val Glu Gln Ala Asn Ala Asp Leu Pro Lys
                85                  90                  95

Leu Ser Thr Met Lys Leu His Ile Val Ser Glu Asn Asn Phe Pro Thr
            100                 105                 110

Ala Ala Gly Leu Ala Ser Ser Ala Ala Gly Phe Ala Ala Leu Ile Thr
        115                 120                 125

Ala Ile Ala Lys Leu Tyr Glu Leu Pro Gln Asp Met Ser Glu Leu Ser
    130                 135                 140

Lys Ile Ala Arg Lys Gly Ser Gly Ser Ala Cys Arg Ser Leu Phe Gly
145                 150                 155                 160

Gly Phe Val Ala Trp Glu Met Gly Asp Ala Ala Asp Gly Gln Asp Ser
                165                 170                 175

Lys Ala Val Glu Val Ala Pro Leu Asp His Trp Pro Ser Met Arg Ala
            180                 185                 190

Val Ile Leu Val Val Ser Asp Asp Lys Lys Asp Thr Pro Ser Thr Thr
        195                 200                 205

Gly Met Gln Ala Thr Val Gln Thr Ser Asp Leu Phe Ala His Arg Val
    210                 215                 220

Thr Lys Val Val Pro Gln Cys Phe Glu Glu Met Lys Lys Ala Ile Val
225                 230                 235                 240

Ala Lys Asp Phe Pro Lys Phe Thr Glu Leu Thr Met Lys Asp Ser Asn
                245                 250                 255

Ser Phe His Ala Val Cys Leu Asp Ser Tyr Pro Pro Ile Phe Tyr Leu
            260                 265                 270

Asn Asp Thr Ser Lys Lys Ile Ile Lys Leu Val Glu Ala Ile Asn Lys
        275                 280                 285

His Ala Thr Ile Ala Ala Tyr Thr Phe Asp Ala Gly Pro Asn Ala Val
    290                 295                 300
```


LCY Biotechnology Holding, Inc.
Ex. 1002
Page 352 of 3077

```
Ile Tyr Tyr Asp Ala Ala Asn Glu Asp Glu Val Leu Ser Gln Leu Tyr
305                 310                 315                 320

Lys Ser Phe Gly His Val Gln Gly Trp Lys Lys Ala Tyr Thr Ala Glu
                325                 330                 335

Thr Ala Val Ala Gly Val Ser Arg Ile Ile Gln Thr Ser Ile Gly Gln
            340                 345                 350

Gly Pro Gln Val Thr Asn Glu Ser Leu Ile Asn Glu Ser Gly Leu Pro
        355                 360                 365

Asn
```

<210> 343
<211> 286
<212> PRT
<213> Candida viswanathii

<400> 343

```
Met Ser Ser Glu Tyr Ala Lys Leu Val Ala Ser Phe Ser Pro Asn Asp
1               5                   10                  15

Ile Leu Ala Lys Trp Pro Glu Val Thr Pro Leu Lys Lys Ile Ser Gly
            20                  25                  30

Ile Pro Arg Ser Ala Glu Ser Asp Ser Ser Asn Gly Ser His Asn Asn
        35                  40                  45

Thr Glu Leu Phe Asn Gly His Asp Glu Glu Gln Ile Arg Leu Met Glu
    50                  55                  60

Glu Leu Cys Ile Val Leu Asp Tyr Asp Asp Lys Pro Val Gly Ala Gly
65                  70                  75                  80

Thr Lys Lys Leu Cys His Ile Met Asp Asn Ile Asn Glu Gly Leu Leu
                85                  90                  95

His Arg Ala Phe Ser Val Phe Leu Phe Asn Glu Asp Gly Lys Leu Leu
            100                 105                 110

Leu Gln Gln Arg Ala Asp Glu Lys Ile Thr Phe Pro Ala Met Trp Thr
        115                 120                 125

Asn Thr Cys Cys Ser His Pro Leu Cys Val Pro Ser Glu Leu Gly Val
    130                 135                 140

Asp Ala Asp Ala Lys Asp Val Asn Asn Leu Asp Asn Ala Val His Gly
```

145 150 155 160

Ala Lys Val Ala Ala Gln Arg Lys Leu Asp His Glu Leu Gly Ile Pro
165 170 175

Phe Ser Asp Thr Pro Leu Asp Gln Phe Thr Tyr Leu Thr Arg Ile His
180 185 190

Tyr Lys Ser Ala Ser Gly Ala Glu Asp Ser Lys Trp Gly Glu His Glu
195 200 205

Ile Asp Tyr Ile Leu Ile Leu Lys Thr Lys Asn Asp Ile Thr Ile Asn
210 215 220

Ala Asn Tyr Asn Glu Val Lys Asp Tyr Lys Tyr Val Asp Ala Lys Glu
225 230 235 240

Leu Gln Glu Met Phe Glu Asp Lys Asp Leu Val Phe Thr Pro Trp Phe
245 250 255

Lys Leu Ile Cys Gln Ser Phe Leu Phe Lys Trp Trp Asn Asn Leu Ser
260 265 270

Asp Leu Glu Lys Tyr Gln Asp Thr Glu Ile His Arg Leu Leu
275 280 285

<210> 344
<211> 351
<212> PRT
<213> Candida viswanathii

<400> 344
Met Ser Asp Lys Ala Ala Ala Arg Glu Arg Phe Leu Ser Val Phe Glu
1 5 10 15

Cys Ala Val Glu Glu Leu Lys Glu Val Leu Val Ser His Lys Met Pro
20 25 30

Gln Glu Ala Ile Asp Trp Phe Val Lys Asn Leu Asn Tyr Asn Thr Pro
35 40 45

Gly Gly Lys Leu Asn Arg Gly Leu Ser Val Val Asp Thr Tyr Ala Ile
50 55 60

Leu Asn Asn Thr Thr Ala Asp Lys Leu Asn Asp Glu Gln Tyr Lys Lys
65 70 75 80

Val Ala Leu Leu Gly Trp Ser Ile Glu Leu Leu Gln Ala Tyr Phe Leu
85 90 95

```
Val Ala Asp Asp Met Met Asp Gln Ser Lys Thr Arg Arg Gly Gln Lys
            100                 105                 110

Cys Trp Tyr Leu Val Glu Gly Val Gly Asn Ile Ala Ile Asn Asp Ser
        115                 120                 125

Phe Met Leu Glu Gly Ala Ile Tyr Val Leu Leu Lys Lys His Phe Arg
    130                 135                 140

Gln Asp Pro Tyr Tyr Val Asp Leu Leu Asp Leu Phe His Glu Val Thr
145                 150                 155                 160

Phe Gln Thr Glu Leu Gly Gln Leu Leu Asp Leu Val Thr Ala Asp Glu
                165                 170                 175

Glu Val Val Asp Leu Asp Lys Phe Ser Leu Asp Lys His Ser Phe Ile
            180                 185                 190

Val Ile Phe Lys Thr Ala Tyr Tyr Ser Phe Tyr Leu Pro Val Ala Leu
        195                 200                 205

Ala Met Tyr Met Ser Gly Ile Ser Ser Glu Glu Asp Leu Lys Gln Val
    210                 215                 220

Arg Asp Ile Leu Ile Pro Leu Gly Glu Tyr Phe Gln Ile Gln Asp Asp
225                 230                 235                 240

Phe Leu Asp Cys Phe Gly Thr Pro Glu Gln Ile Gly Lys Ile Gly Thr
                245                 250                 255

Asp Ile Lys Asp Asn Lys Cys Ser Trp Val Val Asn Gln Ala Leu Leu
            260                 265                 270

His Ala Thr Pro Glu Gln Arg Lys Leu Leu Asp Asp Asn Tyr Gly Lys
        275                 280                 285

Lys Asp Asp Glu Ser Glu Gln Arg Cys Lys Asp Leu Phe Lys Ser Met
    290                 295                 300

Gly Ile Glu Lys Ile Tyr His Asp Tyr Glu Glu Ser Ile Val Ala Lys
305                 310                 315                 320

Leu Arg Glu Gln Ile Asp Lys Val Asp Glu Ser Arg Gly Leu Lys Lys
                325                 330                 335

Asp Val Leu Thr Ala Phe Leu Gly Lys Val Tyr Lys Arg Ser Lys
```

340 345 350

<210> 345
<211> 332
<212> PRT
<213> Candida viswanathii

<400> 345

```
Met Ser Phe Asn Ile Asp Ser Leu Ile Gln Pro Gly Ala Ala Tyr Asp
1               5                  10                  15

Pro Ser Met Thr Asp Ala Ile Met Lys Pro Tyr Arg Tyr Ile Ser Glu
            20                  25                  30

Val Pro Ser Asn Asn His Asn Val Arg Thr Arg Phe Leu Leu Ala Phe
        35                  40                  45

Asn Glu Leu Phe Tyr Gly Met Lys Asn Glu Asp Leu Leu His Arg Ile
    50                  55                  60

Ser His Ile Ile Ser Val Phe His Asn Ser Ser Leu Leu Ile Asp Asp
65                  70                  75                  80

Ile Glu Asp Asp Ser Gln Leu Arg Arg Gly Met Pro Val Ala His Val
                85                  90                  95

Lys Tyr Gly Val Pro Leu Thr Ile Asn Cys Gly Asn Met Met Tyr Phe
            100                 105                 110

Val Ala Val Gln Lys Ala Ile Asp Leu Ala Gly Glu Ala Gly Ser Ala
        115                 120                 125

Glu Leu Lys Phe Glu Thr Ser Gln Ile Leu Val Asp Glu Met Met Asn
    130                 135                 140

Val His His Gly Gln Gly Leu Asp Ile Tyr Trp Arg Asp Tyr Leu Lys
145                 150                 155                 160

Asp Leu Glu His Leu Pro Glu Ile Glu Asp Tyr Leu Gly Met Val Lys
                165                 170                 175

Asp Lys Thr Gly Ser Leu Phe Arg Leu Ala Ile Lys Leu Leu Ser Leu
            180                 185                 190

His Ser Asp Val Gly Glu Asp Asn Gly Leu Val Ala Ile Ala Asn Leu
        195                 200                 205

Leu Gly Ile Ile Tyr Gln Ile Arg Asp Asp Tyr Leu Asn Leu Val Asp
    210                 215                 220
```

```
Ile Lys Tyr Ser Ala Met Lys Gly Val Thr Cys Glu Asp Leu Ile Glu
225                 230                 235                 240

Gly Lys Leu Ser Leu Pro Ile Leu His Cys Leu Arg Thr Thr Thr Asn
                245                 250                 255

Ser Pro Val His Glu Ile Leu Tyr Asn Tyr Asn Thr Ser Ala Glu Arg
            260                 265                 270

Ala Lys Gln Asn Ala Leu Ile Glu Glu Cys Leu Thr Tyr Met Lys Asn
        275                 280                 285

Lys Ser Arg Ser Leu Gln Tyr Thr Leu Asp Leu Ile Lys Thr Leu Glu
    290                 295                 300

Arg Lys Ile Lys Ala Met Met Thr Lys Tyr Pro Asn Ser Glu Asn Ser
305                 310                 315                 320

Gly Leu Ile Lys Ile Ile Asp Arg Leu Cys Asp Leu
                325                 330


<210> 346
<211> 284
<212> PRT
<213> Candida viswanathii

<400> 346
Met Asp Asp Val Asp Ser Ala Leu Ala Asp Asn Val Lys Ser Phe Ala
1               5                   10                  15

Ala Gly Gly Phe Gly Gly Ile Cys Ala Val Leu Thr Gly His Pro Phe
            20                  25                  30

Asp Leu Val Lys Val Arg Leu Gln Thr Gly Leu Tyr Lys Ser Ser Val
        35                  40                  45

Gln Cys Val Lys Glu Thr Ile Ala Lys Asp Gly Leu Phe Gly Leu Tyr
    50                  55                  60

Arg Gly Val Leu Pro Pro Leu Leu Gly Val Thr Pro Met Phe Ala Val
65                  70                  75                  80

Ser Phe Trp Gly Tyr Asp Val Gly Lys Lys Leu Val Ser Ser Phe Thr
                85                  90                  95

Gly Lys Ser Val Asp Lys Phe Glu Ile Lys Asp Ile Ser Thr Ala Gly
            100                 105                 110
```


LCY Biotechnology Holding, Inc.
Ex. 1002
Page 357 of 3077

```
Phe Ile Ser Ala Ile Pro Thr Thr Leu Val Ala Ala Pro Phe Glu Arg
        115                 120                 125

Val Lys Val Met Met Gln Ile Gln Glu Gly Ala Lys Ser Lys Ser Met
    130                 135                 140

Gly Ala Val Val Ala Glu Met Tyr Arg Thr Gly Gly Ile Arg Ser Ile
145                 150                 155                 160

Phe Lys Gly Thr Val Ala Thr Leu Ala Arg Asp Gly Pro Gly Ser Ala
                165                 170                 175

Leu Tyr Phe Ala Thr Tyr Glu Trp Val Lys Lys Glu Leu Thr Ala Pro
            180                 185                 190

Gly Glu Asp Leu Ser Leu Phe Ala Ile Thr Thr Ala Gly Gly Phe Ala
        195                 200                 205

Gly Ile Ala Met Trp Leu Gly Val Phe Pro Ile Asp Thr Ile Lys Ser
    210                 215                 220

Thr Gln Gln Ser Ser Asn Val Lys Val Ser Ile Val Gln Ala Thr Lys
225                 230                 235                 240

Asn Ile Tyr Ala Lys Gly Gly Ile Lys Ala Phe Phe Pro Gly Val Gly
                245                 250                 255

Pro Ala Leu Ala Arg Ala Phe Pro Ala Asn Ala Ala Thr Phe Leu Gly
            260                 265                 270

Val Glu Leu Ala Arg Lys Phe Leu Asp Lys Val Ile
        275                 280


<210> 347
<211> 696
<212> PRT
<213> Candida viswanathii

<400> 347
Met Gly Ala Pro Leu Thr Val Ala Val Gly Glu Ala Lys Pro Gly Glu
1               5                   10                  15

Thr Ala Pro Arg Arg Lys Ala Ala Gln Lys Met Ala Ser Val Glu Arg
            20                  25                  30

Pro Thr Asp Ser Lys Ala Thr Thr Leu Pro Asp Phe Ile Glu Glu Cys
        35                  40                  45
```


LCY Biotechnology Holding, Inc.
Ex. 1002
Page 358 of 3077

```
Phe Ala Arg Asn Gly Thr Arg Asp Ala Met Ala Trp Arg Asp Leu Val
    50                  55                  60

Glu Ile His Val Glu Thr Lys Gln Val Thr Lys Ile Ile Asp Gly Glu
65                  70                  75                  80

Gln Lys Lys Val Asp Lys Asp Trp Ile Tyr Tyr Glu Met Gly Pro Tyr
                85                  90                  95

Asn Tyr Ile Ser Tyr Pro Lys Leu Leu Thr Leu Val Lys Asn Tyr Ser
            100                 105                 110

Lys Gly Leu Leu Glu Leu Gly Leu Ala Pro Asp Gln Glu Ser Lys Leu
        115                 120                 125

Met Ile Phe Ala Ser Thr Ser His Lys Trp Met Gln Thr Phe Leu Ala
    130                 135                 140

Ser Ser Phe Gln Gly Ile Pro Val Val Thr Ala Tyr Asp Thr Leu Gly
145                 150                 155                 160

Glu Ser Gly Leu Thr His Ser Leu Val Gln Thr Glu Ser Asp Ala Val
                165                 170                 175

Phe Thr Asp Asn Gln Leu Leu Ser Ser Leu Ile Arg Pro Leu Glu Lys
            180                 185                 190

Ala Thr Ser Val Lys Tyr Val Ile His Gly Glu Lys Ile Asp Pro Asn
        195                 200                 205

Asp Lys Arg Gln Gly Gly Lys Ile Tyr Gln Asp Ala Glu Lys Ala Lys
    210                 215                 220

Glu Lys Ile Leu Gln Ile Arg Pro Asp Ile Lys Phe Ile Ser Phe Asp
225                 230                 235                 240

Glu Val Val Ala Leu Gly Glu Gln Ser Ser Lys Glu Leu His Phe Pro
                245                 250                 255

Lys Pro Glu Asp Pro Ile Cys Ile Met Tyr Thr Ser Gly Ser Thr Gly
            260                 265                 270

Ala Pro Lys Gly Val Val Ile Thr Asn Ala Asn Ile Val Ala Ala Val
        275                 280                 285

Gly Gly Ile Ser Thr Asn Ala Thr Arg Asp Leu Val Arg Thr Val Asp
    290                 295                 300
```


LCY Biotechnology Holding, Inc.
Ex. 1002
Page 359 of 3077

```
Arg Val Ile Ala Phe Leu Pro Leu Ala His Ile Phe Glu Leu Ala Phe
305                 310                 315                 320

Glu Leu Val Thr Phe Trp Trp Gly Ala Pro Leu Gly Tyr Ala Asn Val
                325                 330                 335

Lys Thr Leu Thr Glu Ala Ser Cys Arg Asn Cys Gln Pro Asp Leu Ile
            340                 345                 350

Glu Phe Lys Pro Thr Ile Met Val Gly Val Ala Ala Val Trp Glu Ser
        355                 360                 365

Val Arg Lys Gly Val Leu Ser Lys Leu Lys Gln Ala Ser Pro Ile Gln
    370                 375                 380

Gln Lys Ile Phe Trp Ala Ala Phe Asn Ala Lys Ser Thr Leu Asn Arg
385                 390                 395                 400

Tyr Gly Leu Pro Gly Gly Gly Leu Phe Asp Ala Val Phe Lys Lys Val
                405                 410                 415

Lys Ala Ala Thr Gly Gly Gln Leu Arg Tyr Val Leu Asn Gly Gly Ser
            420                 425                 430

Pro Ile Ser Val Asp Ala Gln Val Phe Ile Ser Thr Leu Leu Ala Pro
        435                 440                 445

Met Leu Leu Gly Tyr Gly Leu Thr Glu Thr Cys Ala Asn Thr Thr Ile
    450                 455                 460

Val Glu His Thr Arg Phe Gln Ile Gly Thr Leu Gly Thr Leu Val Gly
465                 470                 475                 480

Ser Val Thr Ala Lys Leu Val Asp Val Ala Asp Ala Gly Tyr Tyr Ala
                485                 490                 495

Lys Asn Asn Gln Gly Glu Ile Trp Leu Lys Gly Gly Pro Val Val Lys
            500                 505                 510

Glu Tyr Tyr Lys Asn Glu Glu Glu Thr Lys Ala Ala Phe Thr Glu Asp
        515                 520                 525

Gly Trp Phe Lys Thr Gly Asp Ile Gly Glu Trp Thr Ala Asp Gly Gly
    530                 535                 540

Leu Asn Ile Ile Asp Arg Lys Lys Asn Leu Val Lys Thr Leu Asn Gly
545                 550                 555                 560
```

```
Glu Tyr Ile Ala Leu Glu Lys Leu Glu Ser Ile Tyr Arg Ser Asn His
                565                 570                 575

Leu Ile Leu Asn Leu Cys Val Tyr Ala Asp Gln Thr Lys Val Lys Pro
            580                 585                 590

Ile Ala Ile Val Leu Pro Ile Glu Ala Asn Leu Lys Ser Met Leu Lys
        595                 600                 605

Asp Glu Lys Ile Ile Pro Asp Ala Asp Ser Gln Glu Leu Ser Ser Leu
    610                 615                 620

Val His Asn Lys Lys Val Ala Gln Ala Val Leu Arg His Leu Leu Gln
625                 630                 635                 640

Thr Gly Lys Gln Gln Gly Leu Lys Gly Ile Glu Leu Leu Gln Asn Val
                645                 650                 655

Val Leu Leu Asp Asp Glu Trp Thr Pro Gln Asn Gly Phe Val Thr Ser
            660                 665                 670

Ala Gln Lys Leu Gln Arg Lys Lys Ile Leu Glu Ser Cys Lys Lys Glu
        675                 680                 685

Val Glu Glu Ala Tyr Lys Ser Ser
    690                 695


<210> 348
<211> 650
<212> PRT
<213> Candida viswanathii

<400> 348
Met Ser Gly Leu Glu Ile Ala Ala Ala Ala Ile Leu Gly Ser Gln Leu
1               5                   10                  15

Leu Glu Ala Lys Tyr Leu Ile Ala Asp Asp Val Ser Leu Ala Lys Thr
            20                  25                  30

Val Ala Val Asn Ala Leu Pro Tyr Leu Trp Lys Ala Ser Arg Gly Lys
        35                  40                  45

Ala Ser Tyr Trp Tyr Phe Phe Glu Gln Ser Val Phe Lys Asn Pro Asn
    50                  55                  60

Asn Lys Ala Leu Ala Phe Pro Arg Pro Arg Lys Asn Ala Pro Thr Pro
65                  70                  75                  80
```

```
Lys Thr Asp Ala Glu Gly Phe Gln Ile Tyr Asp Asp Gln Phe Asp Leu
                85                  90                  95

Glu Glu Tyr Thr Tyr Lys Glu Leu Tyr Asp Met Val Leu Lys Tyr Ser
            100                 105                 110

Tyr Ile Leu Lys Asn Glu Tyr Gly Val Thr Ala Asn Asp Thr Ile Gly
        115                 120                 125

Val Ser Cys Met Asn Lys Pro Leu Phe Ile Val Leu Trp Leu Ala Leu
    130                 135                 140

Trp Asn Ile Gly Ala Leu Pro Ala Phe Leu Asn Phe Asn Thr Lys Asp
145                 150                 155                 160

Lys Pro Leu Ile His Cys Leu Lys Ile Val Asn Ala Ser Gln Val Phe
                165                 170                 175

Val Asp Pro Asp Cys Asp Ser Pro Ile Arg Asp Thr Glu Ala Gln Ile
            180                 185                 190

Arg Glu Glu Leu Pro His Val Gln Ile Asn Tyr Ile Asp Glu Phe Ala
        195                 200                 205

Leu Phe Asp Arg Leu Arg Leu Lys Ser Thr Pro Lys His Arg Ala Glu
    210                 215                 220

Asp Lys Thr Arg Arg Pro Thr Asp Thr Asp Ser Ser Ala Cys Ala Leu
225                 230                 235                 240

Ile Tyr Thr Ser Gly Thr Thr Gly Leu Pro Lys Ala Gly Ile Met Ser
                245                 250                 255

Trp Arg Lys Ala Phe Met Ala Ser Val Phe Phe Gly His Ile Met Lys
            260                 265                 270

Ile Asp Ser Lys Ser Asn Val Leu Thr Ala Met Pro Leu Tyr His Ser
        275                 280                 285

Thr Ala Ala Met Leu Gly Leu Cys Pro Thr Leu Ile Val Gly Gly Cys
    290                 295                 300

Val Ser Val Ser Gln Lys Phe Ser Ala Thr Ser Phe Trp Thr Gln Ala
305                 310                 315                 320

Arg Leu Cys Gly Ala Thr His Val Gln Tyr Val Gly Glu Val Cys Arg
                325                 330                 335
```

```
Tyr Leu Leu Asn Ser Lys Pro His Pro Asp Gln Asp Arg His Asn Val
            340                 345                 350

Arg Ile Ala Tyr Gly Asn Gly Leu Arg Pro Asp Ile Trp Ser Glu Phe
        355                 360                 365

Lys Arg Arg Phe His Ile Glu Gly Ile Gly Glu Phe Tyr Ala Ala Thr
    370                 375                 380

Glu Ser Pro Ile Ala Thr Thr Asn Leu Gln Tyr Gly Glu Tyr Gly Val
385                 390                 395                 400

Gly Ala Cys Arg Lys Tyr Gly Ser Leu Ile Ser Leu Leu Leu Ser Thr
                405                 410                 415

Gln Gln Lys Leu Ala Lys Met Asp Pro Glu Asp Glu Ser Glu Ile Tyr
            420                 425                 430

Lys Asp Pro Lys Thr Gly Phe Cys Thr Glu Ala Ala Tyr Asn Glu Pro
        435                 440                 445

Gly Glu Leu Leu Met Arg Ile Leu Asn Pro Asn Asp Val Gln Lys Ser
    450                 455                 460

Phe Gln Gly Tyr Tyr Gly Asn Lys Ser Ala Thr Asn Ser Lys Ile Leu
465                 470                 475                 480

Thr Asn Val Phe Lys Lys Gly Asp Ala Trp Tyr Arg Ser Gly Asp Leu
                485                 490                 495

Leu Lys Met Asp Glu Asp Lys Leu Leu Tyr Phe Val Asp Arg Leu Gly
            500                 505                 510

Asp Thr Phe Arg Trp Lys Ser Glu Asn Val Ser Ala Thr Glu Val Glu
        515                 520                 525

Asn Glu Leu Met Gly Ser Lys Ala Leu Lys Gln Ser Val Val Val Gly
    530                 535                 540

Val Lys Val Pro Asn His Glu Gly Arg Ala Cys Phe Ala Val Cys Glu
545                 550                 555                 560

Ala Lys Asp Glu Leu Ser His Glu Glu Ile Leu Lys Leu Ile His Ser
                565                 570                 575

His Val Thr Lys Ser Leu Pro Val Tyr Ala Gln Pro Ala Phe Ile Lys
```

580 585 590

Ile Gly Thr Ile Glu Ala Ser His Asn His Lys Val Pro Lys Asn Gln
595 600 605

Phe Lys Asn Gln Lys Leu Pro Lys Gly Glu Asp Gly Lys Asp Leu Ile
610 615 620

Tyr Trp Leu Asn Gly Asp Lys Tyr Gln Glu Leu Thr Glu Asp Asp Trp
625 630 635 640

Ser Leu Ile Cys Thr Gly Lys Ala Lys Leu
645 650

<210> 349
<211> 374
<212> PRT
<213> Candida viswanathii

<400> 349
Met Ser Val Lys Thr Lys Thr Ile Thr Ile Leu Pro Gly Asp His Val
1 5 10 15

Gly Thr Glu Ile Val Ala Glu Ala Ile Lys Ser Leu His Ala Ile Glu
20 25 30

Ser Ser Thr Pro Tyr Gln Lys Val His Phe Glu Phe Lys His His Leu
35 40 45

Ile Gly Gly Ala Ala Ile Asp Ala Thr Gly Val Pro Leu Pro Asp Asp
50 55 60

Ala Leu Ala Ala Ala Lys Ser Ser Asp Ala Val Leu Leu Gly Ala Val
65 70 75 80

Gly Gly Pro Lys Trp Gly Thr Gly Ala Val Arg Pro Glu Gln Gly Leu
85 90 95

Leu Lys Ile Arg Lys Glu Leu Asn Leu Tyr Ala Asn Ile Arg Pro Cys
100 105 110

Asn Phe Ala Ser Asp Ser Leu Leu Glu Leu Ser Pro Leu Arg Pro Glu
115 120 125

Val Val Lys Gly Thr Asn Leu Ile Ile Val Arg Glu Leu Val Gly Gly
130 135 140

Ile Tyr Phe Gly Glu Arg Gln Glu Gln Glu Glu Ser Glu Asp Gly Lys
145 150 155 160

```
Ser Ala Trp Asp Thr Glu Lys Tyr Thr Val Asp Glu Val Ala Arg Ile
                165                 170                 175

Thr Arg Met Ala Ala Phe Met Ala Leu Gln His Thr Pro Pro Leu Pro
            180                 185                 190

Ile Trp Ser Leu Asp Lys Ala Asn Val Leu Ala Ser Ser Arg Leu Trp
        195                 200                 205

Arg Lys Thr Val Asp Lys Ile Ile Ser Glu Glu Phe Pro Thr Leu Ala
    210                 215                 220

Val Gln His Gln Leu Ile Asp Ser Ala Ala Met Ile Leu Ile Gln Asn
225                 230                 235                 240

Pro Thr Lys Leu Asn Gly Ile Ile Ile Thr Ser Asn Met Phe Gly Asp
                245                 250                 255

Ile Ile Ser Asp Glu Ala Ser Val Ile Pro Gly Ser Leu Gly Leu Leu
            260                 265                 270

Pro Ser Ala Ser Leu Ala Ser Leu Pro Asp Thr Asn Thr Ala Phe Gly
        275                 280                 285

Leu Tyr Glu Pro Cys His Gly Ser Ala Pro Asp Leu Pro Glu Asn Lys
    290                 295                 300

Val Asn Pro Ile Ala Thr Ile Leu Ser Val Ala Ser Met Leu Arg Leu
305                 310                 315                 320

Ser Leu Asp Cys Val Lys Glu Ala Glu Ala Leu Glu Gln Ala Val Lys
                325                 330                 335

Glu Val Leu Asp Lys Gly Ile Arg Thr Ala Asp Leu Arg Gly Ser Ser
            340                 345                 350

Thr Thr Thr Glu Val Gly Asp Ala Val Ala Glu Thr Val Ser Arg Ile
        355                 360                 365

Leu Lys Glu Ala Lys Ala
    370


<210> 350
<211> 162
<212> PRT
<213> Agrobacterium aurantiacum
```


LCY Biotechnology Holding, Inc.
Ex. 1002
Page 365 of 3077

<400> 350
Met Thr Asn Phe Leu Ile Val Val Ala Thr Val Leu Val Met Glu Leu
1 5 10 15

Thr Ala Tyr Ser Val His Arg Trp Ile Met His Gly Pro Leu Gly Trp
20 25 30

Gly Trp His Lys Ser His His Glu Glu His Asp His Ala Leu Glu Lys
35 40 45

Asn Asp Leu Tyr Gly Leu Val Phe Ala Val Ile Ala Thr Val Leu Phe
50 55 60

Thr Val Gly Trp Ile Trp Ala Pro Val Leu Trp Trp Ile Ala Leu Gly
65 70 75 80

Met Thr Val Tyr Gly Leu Ile Tyr Phe Val Leu His Asp Gly Leu Val
85 90 95

His Gln Arg Trp Pro Phe Arg Tyr Ile Pro Arg Lys Gly Tyr Ala Arg
100 105 110

Arg Leu Tyr Gln Ala His Arg Leu His His Ala Val Glu Gly Arg Asp
115 120 125

His Cys Val Ser Phe Gly Phe Ile Tyr Ala Pro Pro Val Asp Lys Leu
130 135 140

Lys Gln Asp Leu Lys Met Ser Gly Val Leu Arg Ala Glu Ala Gln Glu
145 150 155 160

Arg Thr

<210> 351
<211> 242
<212> PRT
<213> Agrobacterium aurantiacum

<400> 351
Met Ser Ala His Ala Leu Pro Lys Ala Asp Leu Thr Ala Thr Ser Leu
1 5 10 15

Ile Val Ser Gly Gly Ile Ile Ala Ala Trp Leu Ala Leu His Val His
20 25 30

Ala Leu Trp Phe Leu Asp Ala Ala Ala His Pro Ile Leu Ala Ile Ala
35 40 45

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 366 of 3077

```
Asn Phe Leu Gly Leu Thr Trp Leu Ser Val Gly Leu Phe Ile Ile Ala
    50                  55                  60

His Asp Ala Met His Gly Ser Val Val Pro Gly Arg Pro Arg Ala Asn
65                  70                  75                  80

Ala Ala Met Gly Gln Leu Val Leu Trp Leu Tyr Ala Gly Phe Ser Trp
                85                  90                  95

Arg Lys Met Ile Val Lys His Met Ala His His Arg His Ala Gly Thr
            100                 105                 110

Asp Asp Asp Pro Asp Phe Asp His Gly Gly Pro Val Arg Trp Tyr Ala
        115                 120                 125

Arg Phe Ile Gly Thr Tyr Phe Gly Trp Arg Glu Gly Leu Leu Leu Pro
    130                 135                 140

Val Ile Val Thr Val Tyr Ala Leu Ile Leu Gly Asp Arg Trp Met Tyr
145                 150                 155                 160

Val Val Phe Trp Pro Leu Pro Ser Ile Leu Ala Ser Ile Gln Leu Phe
                165                 170                 175

Val Phe Gly Thr Trp Leu Pro His Arg Pro Gly His Asp Ala Phe Pro
            180                 185                 190

Asp Arg His Asn Ala Arg Ser Ser Arg Ile Ser Asp Pro Val Ser Leu
        195                 200                 205

Leu Thr Cys Phe His Phe Gly Gly Tyr His His Glu His His Leu His
    210                 215                 220

Pro Thr Val Pro Trp Trp Arg Leu Pro Ser Thr Arg Thr Lys Gly Asp
225                 230                 235                 240

Thr Ala
```

<210> 352
<211> 589
<212> PRT
<213> Callitropsis nootkatensis

<400> 352

```
Met Ala Glu Met Phe Asn Gly Asn Ser Ser Asn Asp Gly Ser Ser Cys
1               5                   10                  15

Met Pro Val Lys Asp Ala Leu Arg Arg Thr Gly Asn His His Pro Asn
```

```
            20                  25                  30

Leu Trp Thr Asp Asp Phe Ile Gln Ser Leu Asn Ser Pro Tyr Ser Asp
        35                  40                  45

Ser Ser Tyr His Lys His Arg Glu Ile Leu Ile Asp Glu Ile Arg Asp
    50                  55                  60

Met Phe Ser Asn Gly Glu Gly Asp Glu Phe Gly Val Leu Glu Asn Ile
65                  70                  75                  80

Trp Phe Val Asp Val Val Gln Arg Leu Gly Ile Asp Arg His Phe Gln
                85                  90                  95

Glu Glu Ile Lys Thr Ala Leu Asp Tyr Ile Tyr Lys Phe Trp Asn His
            100                 105                 110

Asp Ser Ile Phe Gly Asp Leu Asn Met Val Ala Leu Gly Phe Arg Ile
        115                 120                 125

Leu Arg Leu Asn Arg Tyr Val Ala Ser Ser Asp Val Phe Lys Lys Phe
    130                 135                 140

Lys Gly Glu Glu Gly Gln Phe Ser Gly Phe Glu Ser Ser Asp Gln Asp
145                 150                 155                 160

Ala Lys Leu Glu Met Met Leu Asn Leu Tyr Lys Ala Ser Glu Leu Asp
                165                 170                 175

Phe Pro Asp Glu Asp Ile Leu Lys Glu Ala Arg Ala Phe Ala Ser Met
            180                 185                 190

Tyr Leu Lys His Val Ile Lys Glu Tyr Gly Asp Ile Gln Glu Ser Lys
        195                 200                 205

Asn Pro Leu Leu Met Glu Ile Glu Tyr Thr Phe Lys Tyr Pro Trp Arg
    210                 215                 220

Cys Arg Leu Pro Arg Leu Glu Ala Trp Asn Phe Ile His Ile Met Arg
225                 230                 235                 240

Gln Gln Asp Cys Asn Ile Ser Leu Ala Asn Asn Leu Tyr Lys Ile Pro
                245                 250                 255

Lys Ile Tyr Met Lys Lys Ile Leu Glu Leu Ala Ile Leu Asp Phe Asn
            260                 265                 270
```


LCY Biotechnology Holding, Inc.
Ex. 1002
Page 368 of 3077

```
Ile Leu Gln Ser Gln His Gln His Glu Met Lys Leu Ile Ser Thr Trp
        275                 280                 285

Trp Lys Asn Ser Ser Ala Ile Gln Leu Asp Phe Phe Arg His Arg His
    290                 295                 300

Ile Glu Ser Tyr Phe Trp Trp Ala Ser Pro Leu Phe Glu Pro Glu Phe
305                 310                 315                 320

Ser Thr Cys Arg Ile Asn Cys Thr Lys Leu Ser Thr Lys Met Phe Leu
                325                 330                 335

Leu Asp Asp Ile Tyr Asp Thr Tyr Gly Thr Val Glu Glu Leu Lys Pro
            340                 345                 350

Phe Thr Thr Thr Leu Thr Arg Trp Asp Val Ser Thr Val Asp Asn His
        355                 360                 365

Pro Asp Tyr Met Lys Ile Ala Phe Asn Phe Ser Tyr Glu Ile Tyr Lys
    370                 375                 380

Glu Ile Ala Ser Glu Ala Glu Arg Lys His Gly Pro Phe Val Tyr Lys
385                 390                 395                 400

Tyr Leu Gln Ser Cys Trp Lys Ser Tyr Ile Glu Ala Tyr Met Gln Glu
                405                 410                 415

Ala Glu Trp Ile Ala Ser Asn His Ile Pro Gly Phe Asp Glu Tyr Leu
            420                 425                 430

Met Asn Gly Val Lys Ser Ser Gly Met Arg Ile Leu Met Ile His Ala
        435                 440                 445

Leu Ile Leu Met Asp Thr Pro Leu Ser Asp Glu Ile Leu Glu Gln Leu
    450                 455                 460

Asp Ile Pro Ser Ser Lys Ser Gln Ala Leu Leu Ser Leu Ile Thr Arg
465                 470                 475                 480

Leu Val Asp Asp Val Lys Asp Phe Glu Asp Glu Gln Ala His Gly Glu
                485                 490                 495

Met Ala Ser Ser Ile Glu Cys Tyr Met Lys Asp Asn His Gly Ser Thr
            500                 505                 510

Arg Glu Asp Ala Leu Asn Tyr Leu Lys Ile Arg Ile Glu Ser Cys Val
        515                 520                 525
```

```
Gln Glu Leu Asn Lys Glu Leu Leu Glu Pro Ser Asn Met His Gly Ser
    530                 535                 540

Phe Arg Asn Leu Tyr Leu Asn Val Gly Met Arg Val Ile Phe Phe Met
545                 550                 555                 560

Leu Asn Asp Gly Asp Leu Phe Thr His Ser Asn Arg Lys Glu Ile Gln
                565                 570                 575

Asp Ala Ile Thr Lys Phe Phe Val Glu Pro Ile Ile Pro
            580                 585

<210> 353
<211> 522
<212> PRT
<213> Candida viswanathii

<400> 353
Met Asp Gln Glu Met Ser Arg Pro Lys Gln Ala Ala Ala Ala Ala Val
1               5                   10                  15

Ala Pro Ser Ala Ala Lys Ser Val Ala Pro Glu Glu Asp Glu Asp Glu
            20                  25                  30

Thr Ser Ser Glu Glu Leu Glu Ile Lys Ala Pro Val Lys Pro Leu Pro
        35                  40                  45

Leu Glu Glu Cys Thr Arg Ile Leu Lys Glu Gly Lys Val Lys Thr Leu
    50                  55                  60

Ser Asn Ala Glu Val Ser Ser Leu Val Val Gly Gly Lys Leu Pro Leu
65                  70                  75                  80

Tyr Ala Leu Glu Lys Gln Leu Gly Asp His Lys Arg Ala Val Ala Val
                85                  90                  95

Arg Arg Lys Ala Ile Ala Lys Leu Ala Asn Ala Pro Val Leu Glu Thr
            100                 105                 110

Asn Arg Leu Pro Tyr Ser His Tyr Asp Tyr Asp Arg Val Phe Gly Ala
        115                 120                 125

Cys Cys Glu Asn Val Ile Gly Tyr Met Pro Ile Pro Val Gly Val Ala
    130                 135                 140

Gly Pro Leu Val Ile Asp Gly Lys Pro Tyr His Ile Pro Met Ala Thr
145                 150                 155                 160
```

```
Thr Glu Gly Cys Leu Val Ala Ser Thr Met Arg Gly Cys Lys Ala Ile
                165                 170                 175

Asn Ala Gly Gly Gly Val Glu Thr Val Leu Thr Lys Asp Gly Met Thr
            180                 185                 190

Arg Gly Pro Cys Val Arg Phe Pro Thr Leu Lys Arg Ala Gly Ala Ala
        195                 200                 205

Lys Leu Trp Ile Asp Ser Glu Glu Gly Gln Ile Thr Ile Lys Lys Ala
    210                 215                 220

Phe Asn Ser Thr Ser Arg Phe Ala Arg Leu Gln His Ile Gln Thr Ala
225                 230                 235                 240

Leu Ala Gly Thr Ser Leu Phe Ile Arg Phe Arg Thr Thr Thr Gly Asp
                245                 250                 255

Ala Met Gly Met Asn Met Ile Ser Lys Gly Val Glu Tyr Ser Leu Lys
            260                 265                 270

Tyr Met Val Glu Glu Cys Gly Trp Asp Asp Met Glu Val Ile Ser Val
        275                 280                 285

Ser Gly Asn Tyr Cys Thr Asp Lys Lys Pro Ala Ala Ile Asn Trp Ile
    290                 295                 300

Glu Gly Arg Gly Lys Ser Ile Val Ala Ala Ala Thr Ile Pro Ala Asp
305                 310                 315                 320

Val Val Thr Lys Val Leu Lys Ser Asp Val Asp Ala Leu Val Glu Leu
                325                 330                 335

Asn Val Ser Lys Asn Leu Val Gly Ser Ala Met Ala Gly Ser Val Gly
            340                 345                 350

Gly Phe Asn Ala His Ala Ala Asn Leu Val Thr Ala Val Tyr Leu Ala
        355                 360                 365

Cys Gly Gln Asp Pro Ala Gln Asn Val Glu Ser Ser Asn Cys Ile Thr
    370                 375                 380

Leu Met Glu Lys Asp Lys Gln Thr Gly Asp Leu Val Ile Ser Val Ser
385                 390                 395                 400

Met Pro Ser Ile Glu Val Gly Thr Ile Gly Gly Gly Thr Ile Leu Glu
                405                 410                 415
```

```
Pro Gln Ala Ala Met Leu Asp Leu Leu Gly Val Arg Gly Pro His Pro
            420                 425                 430

Thr Asn Pro Gly Asp Asn Ala Arg Gln Leu Ala Lys Ile Val Ala Ser
        435                 440                 445

Ala Val Leu Ala Ala Glu Leu Ser Leu Cys Ser Ala Leu Ala Ala Gly
    450                 455                 460

His Leu Val Gln Ser His Met Gln His Asn Arg Lys Gly Ala Thr Ala
465                 470                 475                 480

Ala Ala Ala Ala Ala Ala Ala Pro Ala Ile Ala Asn Gly Arg Ala Asn
                485                 490                 495

Gly Thr Lys Thr Asn Gly Ser Ile Asn Gly Lys Asp Leu Lys Arg Leu
            500                 505                 510

Lys Asp Gly Ser Val Thr Cys Ile Lys Ser
        515                 520


<210> 354
<211> 709
<212> PRT
<213> Candida viswanathii

<400> 354
Met Thr Phe Thr Lys Lys Asn Val Ser Val Ser Gln Gly Pro Asp Pro
1               5                   10                  15

Arg Ser Ser Ile Gln Lys Glu Arg Asp Ser Ser Lys Trp Asn Pro Gln
            20                  25                  30

Gln Met Asn Tyr Phe Leu Glu Gly Ser Val Glu Arg Ser Glu Leu Met
        35                  40                  45

Lys Ala Leu Ala Gln Gln Met Glu Arg Asp Pro Ile Leu Phe Thr Asp
    50                  55                  60

Gly Ser Tyr Tyr Asp Leu Thr Lys Asp Gln Gln Arg Glu Leu Thr Ala
65                  70                  75                  80

Val Lys Ile Asn Arg Ile Ala Arg Tyr Arg Glu Gln Glu Ser Ile Asp
                85                  90                  95

Thr Phe Asn Lys Arg Leu Ser Leu Ile Gly Ile Phe Asp Pro Gln Val
            100                 105                 110
```

```
Gly Thr Arg Ile Gly Val Asn Leu Gly Leu Phe Leu Ser Cys Ile Arg
        115                 120                 125

Gly Asn Gly Thr Thr Ser Gln Leu Asn Tyr Trp Ala Asn Glu Lys Glu
    130                 135                 140

Thr Ala Asp Val Lys Gly Ile Tyr Gly Cys Phe Gly Met Thr Glu Leu
145                 150                 155                 160

Ala His Gly Ser Asn Val Ala Gly Leu Glu Thr Thr Ala Thr Phe Asp
                165                 170                 175

Lys Glu Ser Asp Glu Phe Val Ile Asn Thr Pro His Ile Gly Ala Thr
            180                 185                 190

Lys Trp Trp Ile Gly Gly Ala Ala His Ser Ala Thr His Cys Ser Val
        195                 200                 205

Tyr Ala Arg Leu Ile Val Asp Gly Gln Asp Tyr Gly Val Lys Thr Phe
    210                 215                 220

Val Val Pro Leu Arg Asp Ser Asn His Asp Leu Met Pro Gly Val Thr
225                 230                 235                 240

Val Gly Asp Ile Gly Ala Lys Met Gly Arg Asp Gly Ile Asp Asn Gly
                245                 250                 255

Trp Ile Gln Phe Ser Asn Val Arg Ile Pro Arg Phe Phe Met Leu Gln
            260                 265                 270

Lys Phe Cys Lys Val Ser Ala Glu Gly Glu Val Thr Leu Pro Pro Leu
        275                 280                 285

Glu Gln Leu Ser Tyr Ser Ala Leu Leu Gly Gly Arg Val Met Met Val
    290                 295                 300

Leu Asp Ser Tyr Arg Met Leu Ala Arg Met Ser Thr Ile Ala Leu Arg
305                 310                 315                 320

Tyr Ala Ile Gly Arg Arg Gln Phe Lys Gly Asp Asn Val Asp Pro Lys
                325                 330                 335

Asp Pro Asn Ala Leu Glu Thr Gln Leu Ile Asp Tyr Pro Leu His Gln
            340                 345                 350

Lys Arg Leu Phe Pro Tyr Leu Ala Ala Ala Tyr Val Ile Ser Ala Gly
        355                 360                 365
```


LCY Biotechnology Holding, Inc.
Ex. 1002
Page 373 of 3077

```
Ala Leu Lys Val Glu Asp Thr Ile His Asn Thr Leu Ala Glu Leu Asp
    370                 375                 380

Ala Ala Val Glu Lys Asn Asp Thr Lys Ala Ile Phe Lys Ser Ile Asp
385                 390                 395                 400

Asp Met Lys Ser Leu Phe Val Asp Ser Gly Ser Leu Lys Ser Thr Ala
                405                 410                 415

Thr Trp Leu Gly Ala Glu Ala Ile Asp Gln Cys Arg Gln Ala Cys Gly
            420                 425                 430

Gly His Gly Tyr Ser Ser Tyr Asn Gly Phe Gly Lys Ala Tyr Asn Asp
        435                 440                 445

Trp Val Val Gln Cys Thr Trp Glu Gly Asp Asn Asn Val Leu Ala Met
    450                 455                 460

Ser Val Gly Lys Pro Ile Val Lys Gln Val Ile Ser Ile Glu Asp Ala
465                 470                 475                 480

Gly Lys Thr Val Arg Gly Ser Thr Ala Phe Leu Asn Gln Leu Lys Asp
                485                 490                 495

Tyr Thr Gly Ser Asn Ser Ser Lys Val Val Leu Asn Thr Val Ala Asp
            500                 505                 510

Leu Asp Asp Ile Lys Thr Val Ile Lys Ala Ile Glu Val Ala Ile Ile
        515                 520                 525

Arg Leu Ser Gln Glu Ala Ala Ser Ile Val Lys Lys Glu Ser Phe Asp
    530                 535                 540

Tyr Val Gly Ala Glu Leu Val Gln Leu Ser Lys Leu Lys Ala His His
545                 550                 555                 560

Tyr Leu Leu Thr Glu Tyr Ile Arg Arg Ile Asp Thr Phe Asp Gln Lys
                565                 570                 575

Asp Leu Val Pro Tyr Leu Ile Thr Leu Gly Lys Leu Tyr Ala Ala Thr
            580                 585                 590

Ile Val Leu Asp Arg Phe Ala Gly Val Phe Leu Thr Phe Asn Val Ala
        595                 600                 605

Ser Thr Glu Ala Ile Thr Ala Leu Ala Ser Val Gln Ile Pro Lys Leu
    610                 615                 620
```

```
Cys Ala Glu Val Arg Pro Asn Val Val Ala Tyr Thr Asp Ser Phe Gln
625                 630                 635                 640

Gln Ser Asp Met Ile Val Asn Ser Ala Ile Gly Arg Tyr Asp Gly Asp
                645                 650                 655

Ile Tyr Glu Asn Tyr Phe Asp Leu Val Lys Leu Gln Asn Pro Pro Ser
            660                 665                 670

Lys Thr Lys Ala Pro Tyr Ser Asp Ala Leu Glu Ala Met Leu Asn Arg
        675                 680                 685

Pro Thr Leu Asp Glu Arg Glu Arg Phe Glu Lys Ser Asp Glu Thr Ala
    690                 695                 700

Ala Ile Leu Ser Lys
705


<210> 355
<211> 662
<212> PRT
<213> Candida viswanathii

<400> 355
Met Pro Thr Glu Leu Gln Lys Glu Arg Glu Leu Thr Lys Phe Asn Pro
1               5                   10                  15

Lys Glu Leu Asn Tyr Phe Leu Glu Gly Ser Gln Glu Arg Ser Glu Ile
            20                  25                  30

Ile Ser Asn Met Val Glu Gln Met Gln Lys Asp Pro Ile Leu Lys Val
        35                  40                  45

Asp Ala Ser Tyr Tyr Asn Leu Thr Lys Asp Gln Gln Arg Glu Val Thr
    50                  55                  60

Ala Lys Lys Ile Ala Arg Leu Ser Arg Tyr Phe Glu His Glu Tyr Pro
65                  70                  75                  80

Asp Gln Gln Ala Gln Arg Leu Ser Ile Leu Gly Val Phe Asp Pro Gln
                85                  90                  95

Val Phe Thr Arg Ile Gly Val Asn Leu Gly Leu Phe Val Ser Cys Val
            100                 105                 110

Arg Gly Asn Gly Thr Asn Ser Gln Phe Phe Tyr Trp Thr Ile Asn Lys
        115                 120                 125
```


LCY Biotechnology Holding, Inc.
Ex. 1002
Page 375 of 3077

```
Gly Ile Asp Lys Leu Arg Gly Ile Tyr Gly Cys Phe Gly Met Thr Glu
    130                 135                 140

Leu Ala His Gly Ser Asn Val Gln Gly Ile Glu Thr Thr Ala Thr Phe
145                 150                 155                 160

Asp Glu Asp Thr Asp Glu Phe Val Ile Asn Thr Pro His Ile Gly Ala
                165                 170                 175

Thr Lys Trp Trp Ile Gly Gly Ala Ala His Ser Ala Thr His Cys Ser
            180                 185                 190

Val Tyr Ala Arg Leu Lys Val Lys Gly Lys Asp Tyr Gly Val Lys Thr
        195                 200                 205

Phe Val Val Pro Leu Arg Asp Ser Asn His Asp Leu Glu Pro Gly Val
    210                 215                 220

Thr Val Gly Asp Ile Gly Ala Lys Met Gly Arg Asp Gly Ile Asp Asn
225                 230                 235                 240

Gly Trp Ile Gln Phe Ser Asn Val Arg Ile Pro Arg Phe Phe Met Leu
                245                 250                 255

Gln Lys Tyr Cys Lys Val Ser Arg Leu Gly Glu Val Thr Met Pro Pro
            260                 265                 270

Ser Glu Gln Leu Ser Tyr Ser Ala Leu Ile Gly Gly Arg Val Thr Met
        275                 280                 285

Met Met Asp Ser Tyr Arg Met Thr Ser Arg Phe Ile Thr Ile Ala Leu
    290                 295                 300

Arg Tyr Ala Ile His Arg Arg Gln Phe Lys Lys Lys Asp Thr Asp Thr
305                 310                 315                 320

Ile Glu Thr Lys Leu Ile Asp Tyr Pro Leu His Gln Lys Arg Leu Phe
                325                 330                 335

Pro Phe Leu Ala Ala Ala Tyr Leu Phe Ser Gln Gly Ala Leu Tyr Leu
            340                 345                 350

Glu Gln Thr Met Asn Ala Thr Asn Asp Lys Leu Asp Glu Ala Val Ser
        355                 360                 365

Ala Gly Glu Lys Glu Ala Ile Asp Ala Ala Ile Val Glu Ser Lys Lys
    370                 375                 380
```

```
Leu Phe Val Ala Ser Gly Cys Leu Lys Ser Thr Cys Thr Trp Leu Thr
385                 390                 395                 400

Ala Glu Ala Ile Asp Glu Ala Arg Gln Ala Cys Gly Gly His Gly Tyr
                405                 410                 415

Ser Ser Tyr Asn Gly Phe Gly Lys Ala Tyr Ser Asp Trp Val Val Gln
            420                 425                 430

Cys Thr Trp Glu Gly Asp Asn Asn Ile Leu Ala Met Asn Val Ala Lys
        435                 440                 445

Pro Met Val Arg Asp Leu Leu Lys Glu Pro Glu Gln Lys Gly Leu Val
    450                 455                 460

Leu Ser Ser Val Ala Asp Leu Asp Asp Pro Ala Lys Leu Val Lys Ala
465                 470                 475                 480

Phe Asp His Ala Leu Ser Gly Leu Ala Arg Asp Ile Gly Ala Val Ala
                485                 490                 495

Glu Asp Lys Gly Phe Asp Ile Thr Gly Pro Ser Leu Val Leu Val Ser
            500                 505                 510

Lys Leu Asn Ala His Arg Phe Leu Ile Asp Gly Phe Phe Lys Arg Ile
        515                 520                 525

Thr Pro Glu Trp Ser Glu Val Leu Arg Pro Leu Gly Phe Leu Tyr Ala
    530                 535                 540

Asp Trp Ile Leu Thr Asn Phe Gly Ala Thr Phe Leu Gln Tyr Gly Ile
545                 550                 555                 560

Ile Thr Pro Asp Val Ser Arg Lys Ile Ser Ser Glu His Phe Pro Ala
                565                 570                 575

Leu Cys Ala Lys Val Arg Pro Asn Val Val Gly Leu Thr Asp Gly Phe
            580                 585                 590

Asn Leu Thr Asp Met Met Thr Asn Ala Ala Ile Gly Arg Tyr Asp Gly
        595                 600                 605

Asn Val Tyr Glu His Tyr Phe Glu Thr Val Lys Ala Leu Asn Pro Pro
    610                 615                 620

Glu Asn Thr Lys Ala Pro Tyr Ser Lys Ala Leu Glu Asp Met Leu Asn
```


LCY Biotechnology Holding, Inc.
Ex. 1002
Page 377 of 3077

625 630 635 640

Arg Pro Asp Leu Glu Val Arg Glu Arg Gly Glu Lys Ser Glu Glu Ala
645 650 655

Ala Glu Ile Leu Ser Ser
660

<210> 356
<211> 267
<212> PRT
<213> Candida viswanathii

<400> 356
Met Val Ser Thr Lys Thr Tyr Thr Glu Arg Ala Ser Ala His Pro Ser
1 5 10 15

Lys Val Ala Gln Arg Leu Phe Arg Leu Met Glu Ser Lys Lys Thr Asn
20 25 30

Leu Cys Ala Ser Ile Asp Val Thr Thr Thr Ala Glu Phe Leu Ser Leu
35 40 45

Ile Asp Lys Leu Gly Pro His Ile Cys Leu Val Lys Thr His Ile Asp
50 55 60

Ile Ile Ser Asp Phe Ser Tyr Glu Gly Thr Ile Glu Pro Leu Leu Val
65 70 75 80

Leu Ala Glu Arg His Gly Phe Leu Ile Phe Glu Asp Arg Lys Phe Ala
85 90 95

Asp Ile Gly Asn Thr Val Met Leu Gln Tyr Thr Ser Gly Val Tyr Arg
100 105 110

Ile Ala Ala Trp Ser Asp Ile Thr Asn Ala His Gly Val Thr Gly Lys
115 120 125

Gly Val Val Glu Gly Leu Lys Arg Gly Ala Glu Gly Val Glu Lys Glu
130 135 140

Arg Gly Val Leu Met Leu Ala Glu Leu Ser Ser Lys Gly Ser Leu Ala
145 150 155 160

His Gly Glu Tyr Thr Arg Glu Thr Ile Glu Ile Ala Lys Ser Asp Arg
165 170 175

Glu Phe Val Ile Gly Phe Ile Ala Gln Arg Asp Met Gly Gly Arg Glu
180 185 190

Glu Gly Phe Asp Trp Ile Ile Met Thr Pro Gly Val Gly Leu Asp Asp
195 200 205

Lys Gly Asp Ala Leu Gly Gln Gln Tyr Arg Thr Val Asp Glu Val Val
210 215 220

Leu Thr Gly Thr Asp Val Ile Ile Val Gly Arg Gly Leu Phe Gly Lys
225 230 235 240

Gly Arg Asp Pro Glu Val Glu Gly Lys Arg Tyr Arg Asp Ala Gly Trp
245 250 255

Lys Ala Tyr Leu Lys Arg Thr Gly Gln Leu Glu
260 265

<210> 357
<211> 906
<212> DNA
<213> Chronobacter sakazakii

<400> 357
atgaatgcta atgccgttaa gtcctccggt caggaaatcg aattgcaagc tttgagggat 60

gccttgcaaa cacgcttgga cgaattgctc ccacccgggc aggagagaga cttggtttgc 120

gcagccatgc gcgaaggggc attgacgcca gggaagagag ttcgtccatt acttttgatc 180

ttggcagcga gggacttggg ttgcgacgca tcccagccag ccttgatgga tttagcatgt 240

gcagtagaga tggttcatgc ggcctcactc atgttggatg atattccctg tatggacaac 300

gccctttga gacgaggtaa acctacaatc catcgacagt acggtgagtc ggtggcaata 360

ttggcggctg ttgcgttgct tagcagagca ttcggcgttg tggcacaggc gaaccctcta 420

tctgactcat gtaagaccca agcagtgtcg gagctttcct cggccgttgg tttgcagggg 480

ttggtgcagg gtcaatttcg tgatttgtcc gaagggaacc aggcaaggag tgcagaagca 540

attttggcca ccaacgattt gaaaacctca gtgttgttcg acgctacctt acagattgcg 600

gccattgcgg caggcgcaag tgcgtcagtt agacacaagt tgagagagtt ctctcgacac 660

ttggggcaag ccttccagtt gttggacgat ttagccgatg gattaaacca cacgggaaag 720

gatatcaaca aagacgctgg caagagcacc cttgtcgcaa tgttgggtcc ggaagctgtg 780

caccaacgct tgcgagacca tcttcttcgc gcagatgaac acttgactgg tgcatgttca 840

agaggtgcct ccacccgtcg cttcatgtat gcatggtttg acaaacagct tgcaatgttt 900

gggtaa 906

<210> 358
<211> 1479

<212> DNA
<213> Chronobacter sakazakii

<400> 358
atgaccaaaa cggttgtaat aggatcagga tttgggggtc tcgcactagc cattagactt 60

caagccgctg gtgtaccaac cttgttgctt gagcagcgag acaagccggg aggccgtgca 120

tatgtgtacg aggacaaagg atttaccttt gacgccgggc ccactgtcat tacggaccca 180

tccgctatcg aagagttatt tacccttgcc ggtaaaaata tagctgatta cgtggatttg 240

ttgccagtaa ccccccttcta ccgtttgtgt tgggaaaatg gtcaagtgtt taattacgat 300

aacgaccaag ccagtttgga agcacaaatc gcacgtttta acccaaggga tgtggagggt 360

taccgacaat tcctcgcgta ttcacaagca gtatttaaag aggggtactt gaaactaggc 420

gcggttccat tcttgagctt cagagacatg ttgcgtgcgg gaccccaatt ggccagactc 480

caagcctgga ggagtgtcta tggtatggtt agcaagttta tcgagaacga tcatttgcga 540

caagctttca gcttccacag tttgttggtg ggtggcaacc cattcgcaac aagctccatc 600

tatacactca tccatgcctt ggaaaggcag tggggcgtct ggtttgccag aggaggcacc 660

ggagctttgg ttcagggctt agtaaagttg tttactgatt tgggggggaga aattgaactc 720

aacgcaaagg tcaccaggtt ggatacccaa ggagacaaaa taagtggcgt gactttggca 780

gatggtaggc ggattccagc tagggccgtc gcatccaacg cagatgtggt ccatacctac 840

aacaatttgt tgggtcatca tccgagaggc gtttcgcaag ctgcctctct tagaagaaag 900

cgcatgagca attccctctt tgtgttgtat ttcgggctca accatcacca ctcccaatta 960

gcccaccata cggtctgttt tggaccacgt tataaggggt tgatagaaga cattttcaag 1020

agagacagct tggcagacga tttttccttg tatcttcacg ctccgtgtgt gacagaccct 1080

tcattggctc caccaggatg tggatcatac tacgtattgg ctcctgtgcc acacttgggt 1140

actgcaaact tgaactggga tgtagagggg ccacgattaa gagacaggat ctttgaatac 1200

ttagaacaac actatatgcc gggtcttaga gatcaacttg ttacacaccg tatgttcact 1260

ccgtttgact ttagagacca attaggtgcc taccacggtt ccgctttctc cgttgagcca 1320

atcctaactc aatctgcttg gtttcgccca cacaacagag actctcgaat tgataacctt 1380

tatttggttg gagctggaac gcatcccggt gcaggaattc cgggtgtcat cgggagcgcg 1440

aaagcaacgg ccggcttaat gttggagggc cacgcgtaa 1479

<210> 359
<211> 972
<212> DNA
<213> Chronobacter sakazakii

<400> 359
atgagtgaca agcctttgtt aacacatgct acagagacca ttgaggccgg atcaaagagt 60

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 380 of 3077

tttgcgaccg ccagtaagct ctttgacgcc aagacacgta gaagtgctct tatgttgtac 120

gcttggtgtc gtcattgcga tgacgtcact gacggtcaag ccttgggatt tagggctgca 180

gacgctccta cagatacacc acaggctcgt atcgcattgt tgcgtgctct aaccttggaa 240

gcgtatgctg gtaaaccaat gcgtgaacca aactttgcag ctttccagga ggttgctcta 300

gcacatcaga tcccacccgc attggcactc gatcacttgg aagggtttgc gatggatgta 360

cgtgaagaac gctaccatac ctttgacgat accctaagat actgttatca tgttgcaggc 420

gttgttgggt tgatgatggc tagagttatg ggtgtccgag acgaggctgt cttagataga 480

gcctgcgatt taggtttggc ttttcagtta acaaatattg cccgagacat tgtcgaagat 540

gcggcaatcg gacgatgcta tctcccagag gcctggttgc aagaagaagg actttgtgcg 600

gatacattga ctgacagagc acacagacca gctttggcta ggctagccgc acgattggtg 660

gatgaggcag agccatacta tgcctcagcg agagctggat tggccggact cccattaagg 720

tcagcctggg ccatcgctac tgcccacggg gtttacagag aaatcggcgt aaaggtgaag 780

agggctggtg tgaacgcgtg ggaaacgcga cagggaacgt cgaaggccga gaaactcgcc 840

ttgttggcca agggggcggt gatggcagtc tcctccagag gagccagttc aagtccacga 900

ccttccgcgt tgtggcaaag accacgcgct caggatgatc ggtacgcgca tgccgcccct 960

cctgccgctt aa 972

<210> 360
<211> 1131
<212> DNA
<213> Xanthophyllomyces dendrorhous

<400> 360
atggactacg ccaacatcct cactgccatt ccattggaat ttacgccaca agatgatata 60

gttttgttgg aaccctacca ctacctcggt aaaaaccccg ggaaggaaat ccgatcacaa 120

ttgattgagg cattcaacta ctggttagat gttaaaaagg aggatctcga ggttattcag 180

aatgtagtcg gaatgttgca cacagcatcc cttttgatgg atgatgtcga agactcttct 240

gtactacgac gtggatcacc ggtcgcgcat ttgatttacg gaatcccaca aaccattaac 300

accgccaact acgtttattt cctagcttat caggaaatat tcaaattgag gcctacccca 360

atccctatgc cagtcatccc cccatcgtcg gcctcattgc agtccagcgt aagctcagct 420

agctcctcgt catctgcgtc atcggaaaac gggggaacat caacacccaa cagtcaaatt 480

ccattcagca aggacaccta ccttgacaaa gtaatcaccg atgaaatgct ttccttgcac 540

aggggccagg gcttagaact tttctggaga gactcgctaa catgtcccag tgaggaagag 600

tacgttaaga tggtccttgg caaaacaggg gggctattcc gaattgctgt caggttgatg 660

atggctaaaa gtgaatgcga tattgacttc gttcaattgg tcaatttaat ttcgatatat 720

ttccagatta gagatgacta catgaacttg caatcgtccg agtatgctca caacaagaac 780

ttcgcggagg atctaacgga gggcaagttt tcattcccta cgattcactc gattcacgca 840

aatccatcaa gtcgtttggt catcaacact ttgcaaaaga agtcaacaag ccctgaaatc 900

cttcatcact gtgtgaatta catgagaact gaaactcact cctttgagta cacccaagag 960

gtattgaaca cactcagtgg agccttagag agagaacttg gtcggttgca aggagagttt 1020

gccgaagcca acagcaaaat cgacttgggg gatgtggaaa gtgaaggtcg cacaggtaaa 1080

aacgtaaaac ttgaggcaat cttgaaaaag cttgccgaca tcccattata a 1131

<210> 361
<211> 1749
<212> DNA
<213> Xanthophyllomyces dendrorhous

<400> 361
atgggcaagg aacaagatca agacaaacct actgccatta tcgtcggttg tggtataggc 60

ggcatcgcta ccgccgctcg tcttgctaag gagggctttc aagtcacggt gttcgagaag 120

aacgattata gcggaggcag gtgttcgcta atcgaacgcg atggttaccg atttgaccag 180

ggtccgagtt tgttattgct cccggacttg tttaaacaaa cctttgaaga ccttggagaa 240

aagatggagg attgggtaga cttaataaaa tgcgaaccga actacgtttg tcattttcat 300

gacgaagaga ctttcacctt tagtacggac atggcattgt tgaagagaga agtcgagaga 360

tttgaaggaa aggatggttt cgacagattt ctatccttta ttcaagaagc ccaccgtcat 420

tatgagttgg ctgtcgtgca cgtcttacaa aaaaatttcc caggtttcgc tgcattcctt 480

cgccttcaat ttataggcca gatcttagca ttgcacccat ttgaatcaat atggacccgc 540

gtgtgcagat acttcaagac tgaccgcttg agaagagttt tttcgttcgc agtcatgtac 600

atgggtcaaa gcccttactc agcgccggga acatatagtt tgttgcagta caccgagtta 660

acggagggaa tatggtaccc aaggggcggt ttctggcagg ttccaaatac cttgttgcaa 720

attgttaaac gaaacaaccc ttcagctaaa ttcaacttca atgcaccagt gagtcaggtc 780

ctcttgagtc ctgccaaaga cagagcaacc ggagtgaggt tggagagtgg cgaagagcat 840

cacgcggacg ttgtcatcgt caacgcagac cttgtttatg ctagcgagca tcttattcct 900

gatgacgcta ggaacaagat tgggcaatta ggagaggtga aaaggtcttg gtgggcggat 960

ttggtgggag gtaagaaact caaaggttcg tgttcttcat tgtcattcta ctggtccatg 1020

gaccgaatag tggacggttt ggggggtcac aatatctttt tggccgagga cttcaagggt 1080

agttttgaca ccattttcga ggaactagga ttgccggctg atccttcttt ttacgtaaat 1140

gttccatctc gcatcgatcc gtcagctgcc ccggaaggta aggatgccat tgtgatacta 1200

gtgccatgtg gtcatattga tgccagcaac ccacaagatt acaataagtt ggttgctagg 1260

gctaggaagt tcgtgattca gacattgtcg gcaaagttgg gacttccaga cttcgaaaaa 1320

atgattgttg cggaaaaggt gcacgatgct ccgtcgtggg aaaaggagtt caacctcaag 1380

gatggttcaa tcttgggttt ggcgcacaac ttcatgcaag tgctcgggtt caggccgtcg 1440

actagacacc ctaagtatga caaactcttt tttgttggag ctagcacaca tcctggtaca 1500

ggggttccaa tcgtccttgc tggtgccaaa ttgacagcta atcaagtgtt ggagtccttc 1560

gacagatctc cagcccccga tccaaatatg agcctatctg ttccatacgg caagccttta 1620

aagtcgaacg ggacgggcat cgatagtcaa gtccagttga agttcatgga tttggaaaga 1680

tgggtttact tgttggtttt gttaatcgga gccgttattg ccagaagcgt tggagttctc 1740

gcattttaa 1749

<210> 362
<211> 2022
<212> DNA
<213> Xanthophyllomyces dendrorhous

<400> 362
atgacggcat tggcgtacta ccaaatccac ttgatttata ccttgcccat actagggttg 60

ttgggtttgt taacttctcc aatccttacc aaatttgata tataaaaat ctctatcttg 120

gtttttattg cattttcagc aacaactcca tgggacagtt ggataattag aaatggggca 180

tggacttacc cttcggccga gtcgggtcaa ggcgtctttg ggacttttct agatgttcca 240

tatgaagaat acgctttctt cgttatccaa acggttataa caggtttggt ctacgtcctc 300

gccacaaggc atttgttacc atcgctcgcc ttgccaaaga cacgatcctc ggcattaagc 360

ctcgcgttga aggcattgat tccgctccca attatctatt tgttcacagc ccacccaagt 420

ccttcaccag atcccttggt taccgatcat tacttctata tgagggcgtt gtcactcttg 480

atcacgcctc ccaccatgtt acttgctgcc ttgagcgggg aatacgcgtt cgactggaag 540

tccggccgag caaaaagcac gattgcagct atcatgattc caacggtata cctaatttgg 600

gtcgattacg tcgcagtagg acaggacagc tggagtatca acgacgaaaa aatcgtgggc 660

tggcgccttg ggggggtcct tcctatagaa gaggcaatgt tctttttgtt gacgaatctc 720

atgatcgtgt tagggctttc agcatgcgac catactcaag ccttgtatct tctccatggt 780

agaactattt atggtaataa aaaaatgccg tcgtcgttcc cattgattac cccgccagtg 840

ctttctttgt tcttctcaag tagaccgtat tcatcgcaac ccaagcgtga tttggaattg 900

gcggttaagt tgttggaaga gaaatcccgt agcttcttcg tggcaagtgc cggttttcca 960

agtgaagtga gagagaggtt ggtgggctta tacgcgtttt gcagagtcac ggacgacctt 1020

atagacagtc cagaagtgtc ctctaaccca cacgcgacga tcgatatggt tagcgatttt 1080

ttgaccttgc tttttggacc acccttgcac ccatcccaac cagataaaat tttgagctca 1140

cctctactcc cgcccagtca cccaagcagg cccacgggga tgtacccact ccccccgcca 1200

ccgagtttgt ctccagccga attggttcag ttcttgacgg aacgcgttcc agtgcagtat 1260

cactttgcgt tcaggttgtt ggcgaaatta caagggttga tcccaagata tccgttagac 1320

gagctattga gaggttacac caccgatttg atattccctt tgtccaccga ggctgtgcaa 1380

gccagaaaaa ccccaatcga gactacggca gatcttttgg attacggttt gtgcgtagcg 1440

gggagcgtcg ccgagttatt agtctacgtt tcgtgggcat ccgctccatc acaggttcca 1500

gccactattg aggaaagaga ggccgtgttg gtcgctagca gagagatggg tacagcgttg 1560

caattagtga atattgctcg agatattaaa ggtgatgcca ctgagggccg tttctacttg 1620

ccattgagtt ttttcggttt gcgtgacgaa tcaaaactcg caataccaac agattggact 1680

gaacctaggc cacaagactt cgataaattg ttgtcgctct ccccaagctc cactcttcca 1740

tcatcaaatg cgtccgaatc cttcagattt gagtggaaga cgtacagtct tcccctagtg 1800

gcctacgccg aagacttggc gaaacacagt tacaagggaa ttgaccgtct acccaccgag 1860

gtgcaggcag gtatgagggc ggcgtgcgcc agttaccttt tgatagggcg cgaaatcaag 1920

gttgtgtgga agggtgacgt aggtgagcga cggacggtgg ccggttggcg aagagtcaga 1980

aaagttcttt ctgtggttat gtcgggttgg gagggacaat aa 2022

<210> 363
<211> 909
<212> DNA
<213> Pantoea ananatis

<400> 363
atgacggtgt gtgctaagaa gcacgtccac ttgactagag acgccgccga acaattattg 60

gccgacattg acaggagact tgaccaactt ttgccggttg aaggtgagag ggatgttgtg 120

ggagctgcaa tgcgagaggg ggctttagcc ccaggtaaga ggataagacc tatgttactt 180

ctattgacag cgagggatct aggttgtgct gtctcacatg atggtttgct agatttggcc 240

tgtgcggtcg agatggtcca tgccgccagc ctcatcctcg acgatatgcc ctgtatggat 300

gacgctaaat tgagacgtgg tcggccaacg attcactcgc actacgggga acatgttgca 360

atcctagccg cagtcgcctt gttgtcaaag gccttcggcg ttatcgccga cgccgacggc 420

ttgacgccat tggcaaaaaa cagagcggtg agtgagttga gtaacgctat aggtatgcaa 480

gggttggttc aaggacaatt caaagatttg tccgaggggg acaagccaag aagtgctgaa 540

gcaattttga tgaccaatca cttcaaaaca tccacattgt tctgcgcatc tatgcaaatg 600

gcatccattg ttgcaaacgc ctcatcggag gccagagatt gtctacatcg attttctctt 660

gatttgggcc aagcgttcca gttgttggac gaccttacag acggtatgac cgatactggc 720

aaagattcga accaagacgc aggtaaaagt acgttggtaa atctcttggg tcctagagct 780

gtcgaggaga ggcttagaca acatctccag ttagcatccg aacatctatc agctgcctgc 840

caacacggtc acgccactca acactttatt caagcgtggt ttgacaagaa gttggcggct 900

gtaagctaa 909

<210> 364
<211> 1479
<212> DNA
<213> Pantoea ananatis

<400> 364
atgaaaccca caaccgttat tggtgctggc tttggcggtt tggcattggc cataagattg 60

caagctgccg gcatacctgt gttattgctt gaacagcggg acaagccagg aggacgggcg 120

tacgtctatg aagatcaggg tttcactttt gatgctgggc ccacggttat cacagacccg 180

tcggcgatcg aggaactttt tgccttggct ggaaagcaat tgaaagaata tgtcgagtta 240

cttccagtta caccatttta tcgattgtgc tgggaatccg gtaaggtgtt taattacgac 300

aacgaccaaa ctaggcttga ggcccaaatc caacaattca acccccggga tgttgagggt 360

tatcgccaat tcctcgacta tagcagagct gtcttcaagg aagggtactt gaaattgggc 420

actgtgccat tcttgtcttt ccgagatatg ttacgagcag ccccacagtt ggcgaaattg 480

caagcatgga gaagcgttta cagtaaagtt gcttcttata ttgaggacga acatttgaga 540

caggcctttt catttcacag cttgttggtg ggaggtaacc cattcgctac tagctctatt 600

tacaccttga ttcacgccct agaaagggaa tggggtgttt ggttcccaag aggtgggaca 660

ggcgctttag tgcaaggaat gatcaaattg ttccaagact taggagggga agtggtgttg 720

aacgcgcgtg tgtctcacat ggaaactaca gggaacaaga tagaagcagt gcatctcgaa 780

gacggccgtc gattcttgac acaagcagtt gcatcgaatg ccgacgtcgt tcacacctac 840

cgcgacttgc tatctcaaca tccagccgct gtgaagcaat ccaataaatt acaaacaaaa 900

agaatgtcaa actccttatt tgttttgtac ttcggtctaa atcaccacca tgatcaactc 960

gctcatcaca cagtctgctt cggtccaaga tatagagaat tgatcgacga gatattcaat 1020

catgacgggt tggccgaaga tttttctttg tatctacacg ctccatgcgt cactgatagc 1080

tctctcgctc ctgaggggtg cggttcttac tacgttttgg cccctgtacc tcacttgggt 1140

acggcgaact tggattggac cgtagagggt cccaagctac gtgacagaat ctttgcatac 1200

ttagaacaac actacatgcc ggggttacgc agccaattgg ttacacatag aatgttcacg 1260

ccgttcgact tccgagacca actaaacgcg tatcacggct ctgctttttc agttgagcca 1320

gtgctcacgc aaagcgcatg gtttaggcct cacaacaggg acaaaacgat cactaactta 1380

tacctcgtgg gggccgggac ccacccagga gcaggaatac caggggtcat tggttcggct 1440

aaggcgactg ccggcttgat gcttgaggac ttgatttaa 1479

<210> 365
<211> 930
<212> DNA
<213> Pantoea ananatis

<400> 365
atgaacaacc cctcattgtt gaaccatgcc gtggagacaa tggcagtcgg ttctaagagc 60

tttgctaccg catcgaaatt gtttgatgcc aaaactcgga ggtcagtttt gatgctctac 120

gcgtggtgtc gtcattgtga cgatgttatc gatgaccaaa ccctcggctt ccaagctcgg 180

cagccggcat tgcaaacccc agaacaaagg ttgatgcaat tggagatgaa gaccaggcaa 240

gcttatgccg gcagccagat gcacgaaccc gcctttgcgg cttttcaaga ggttgcaatg 300

gctcatgata tagcacctgc ttatgccttt gaccacttgg agggtttcgc tatggacgtc 360

agggaagccc agtatagcca attggatgat accttaaggt actgttatca tgtcgcagga 420

gttgtgggat tgatgatggc gcagatcatg ggggttcgag acaacgctac cttggataga 480

gcatgtgatt tggggttggc gtttcaattg actaatatcg cgcgagatat tgtcgacgat 540

gcacacgcag gtagatgcta cttgccagcc tcttggctag aacacgaagg cctcaacaag 600

gagaattacg ccgcccctga aaatcgacaa gctttgtcaa gaatagcgag aagattggtg 660

caagaggccg agccatatta tttgagtgcg acagctggtt tagccggatt gccattgagg 720

agtgcatggg ccatagccac ggcgaaacaa gtataccgta aaataggggt caaagtcgag 780

caagctggtc agcaagcatg ggatcagcgg caatcaacca caacgccaga gaagttgact 840

ttgctcttgg ccgccagtgg acaggctcta acctcccgta tgagagccca tcccccaagg 900

ccagctcact tgtggcagag gccattgtaa 930

<210> 366
<211> 2241
<212> DNA
<213> Xanthophyllomyces dendrorhous

<400> 366
atggcaaccc tatctgactt ggttatcttg ttgttgggtg ccttgttagc attggggttc 60

tataacaagg acagattgct cggttcaagt tccagcagtg cctctaccac atcaggttca 120

agcgctgcaa ctgctaacgg tagtaagccc acatattcta atggcaacgg gaacgccttt 180

aaaggcgacc cacgagattt cgtcgcgaga atgaaggatc aaaagaaaag gttggcagtt 240

ttttatgggt cccaaactgg aactgctgag gagtacgcca cccgaatcgc caaggaagcg 300

aaatctcgtt tcggcgtttc gagtttggtg tgtgacattg aggaatatga ttttgaaaaa 360

ttggatcaag tccccgagga ttgcgccatt gttttttgca tggctaccta tggtgaaggt 420

gaaccaacag acaatgctgt ccaattcatc gaaatgatca gccaggatga cccagagttt 480

agtgaaggta gtactttgga tggtttgaag tacgttgtct ttggattggg gaataagaca 540

tacgagcaat acaatgttgt cggccggcag ttggacgcta gattaaccgc acttggtgcg 600

accagagtag gtgagcgagg tgaaggagat gatgacaaaa gtatggagga agattatctc 660

gcatggaagg atgatatgtt cgccgctttg gccaccacat taagttttga agagggagct 720

tcgggggaga ctccagattt cgtggtcact gaagtcccga accaccccat cgaaaaggtc 780

ttccagggcg aattgtccag cagagccttg ttgggctcta aaggcgttca cgacgccaag 840

aatccatacg cttcgccagt tttggcctgc cgggaattat tcacaggcgg agatcgaaac 900

tgtatccacc ttgagtttga catcaccggg tccggaatca cctaccaaac aggtgatcac 960

gtcgcagttt ggccatccaa tcctgacgtt gaagtggaaa ggcttttggc agtgttgggt 1020

ctcacatctc cagagaagag acggatgatt atccaagtgg tgagccttga tccaactttg 1080

gccaaggttc cattccctac accaaccacc tatgatgccg ttttcagaca ttaccttgac 1140

attagtgcag tggcttcacg ccagaccttg gctgtccttg ccaaatacgc cccttccgaa 1200

caggctgccg aattcttaac caggttggga actgacaagc aagcgtacca tactgaagtg 1260

gttggaggtc acctaagatt ggccgaggtg cttcaattgg ctgcagggaa cgatattact 1320

gttatgccta cagctgaaaa tacaaccgtc tggaacatac cattcgacca cgttgtgtct 1380

gacgtttccc gcttgcaacc ccgattttac tcgataagca gttctcctaa gcttcacccg 1440

aacagcattc acgtaacggc tgtgatactt aaatacgaaa gtcaagccac tgaccgtcac 1500

ccagcgagat gggtgtttgg tttggggaca aactatttgc ttaatgttaa gcaagcagca 1560

aacaatgaaa ccactccaat gatctctgac ggacaagacg atgtgccaga acacgtgtca 1620

gcgccaaaat acacgttgga gggaccccgt ggctcctata agcacgacga tcaattgttt 1680

aaagtgccaa tccatgtgcg taggtccact ttcagacttc caaccagccc taagatccct 1740

gtgatcatga ttgggccagg taccggtgtc gctcctttca gaggattcat ccaagaaaga 1800

attgcccttg ccagaaggtc cattgctaag aatgggcctg acgccttggc ggattgggcc 1860

ccaatctatc ttttctacgg tagtcgcgat gagcaagact tcttgtacgc ggaagagtgg 1920

cctgcttatg aagcggagtt gcaagggaaa ttcaagattc atgttgcttt cagtagatcc 1980

gggccaagaa agccagatgg gagtaagatt tacgtgcaag atcttttgtg ggatcaaaag 2040

gaagtcatca aatctgccat agtggaaaaa cgtgcatctg tttatatttg tggggacggt 2100

cgaaatatgt cgaaagatgt cgagcagaag ttggccgcga tgttggccga aagtaaaaac 2160

ggtagtgcgg ccgtggaagg agccgcggaa gtcaagtcat tgaaagagag atctcgatta 2220

ttgatggatg tgtggagcta a 2241

<210> 367
<211> 1674
<212> DNA

<213> Xanthophyllomyces dendrorhous

<400> 367
atgttcattt tggtgttgtt gaccggtgct ttgggtttgg cagcttttag ctgggcttcg 60

attgccttct tttctcttta tttggcgccc cgtcgcagta gcttgtataa tttgcaaggt 120

ccaaaccaca ctaactattt caccggcaat tttttagata tactctcggc taggacgggt 180

gaggagcatg ctaagtatag ggaaaaatac ggatccacgt tgagatttgc cggcatcgcc 240

ggtgccccag ttttgaactc aactgatcca aaagtcttca accatgtgat gaaggaggca 300

tatgattacc caaaacctgg tatggctgcg agagtgttga ggattgcaac gggtgacgga 360

gtggttacgg ctgagggtga ggcacataaa cgacatcgaa gaatcatgat cccttccttg 420

tcagcacaag ccgttaagtc catggttccc atcttccttg agaagggcat ggagctcgtt 480

gacaagatga tggaggacgc agccgagaaa gacatggcag tgggcgaatc cgcaggtgag 540

aagaaggcta ctcgacttga aacggaaggc gttgacgtaa aggactgggt tgggagagca 600

actttggatg tcatggcttt ggccggattt gattacaaat cagattcgct acaaaataag 660

acgaacgaat tgtatgtcgc tttcgtgggg cttacagacg gatttgcacc caccttggac 720

agttttaagg ccatcatgtg ggatttcgtg ccgtactttc gcacaatgaa gagaagacat 780

gagattccct tgacccaggg tttggctgtt agtcgccgtg ttggcattga actaatggag 840

caaaagaagc aagcagtgtt gggttcggca tcggatcaag ccgtggacaa aaaggacgtg 900

cagggacgcg acattcttag tctcttggtc agagcgaaca ttgccgcaaa tcttcccgaa 960

tctcaaaagt tgtcggacga agaggttctt gctcaaattt cgaatttgct attcgcgggt 1020

tatgagactt caagtactgt tctcacatgg atgttccatc gattgtcgga ggacaaagca 1080

gtgcaagata aattaaggga ggaaatctgt caaatagata cagacatgcc aacactagac 1140

gagttgaatg cattgccgta cttggaggcg tttgtgaaag agtcactaag actcgatccc 1200

ccgtctccat atgctaatag ggagtgtttg aaggatgaag atttcattcc attggccgaa 1260

cctgtcattg gacgcgatgg gtccgtgatc aatgaagtca gaatcacaaa ggggacaatg 1320

gtgatgcttc cattgttcaa catcaataga tcgaaattca tatacggtga ggacgccgaa 1380

gaatttagac cagaaaggtg gttggaggat gttaccgact cccttaacag catagaagct 1440

ccatatggcc atcaagcgtc cttcatatca ggaccaagag cctgctttgg atggcggttc 1500

gcggtcgcag aaatgaaagc attcttgttc gtaacactaa ggagagtgca atttgagccg 1560

atcataagcc atccagagta cgaacatatt actctcatta tcagcagacc ccggatcgtt 1620

ggcagagaga aggaaggata tcagatgcgt ttgcaggtaa agccagtgga gtaa 1674

<210> 368
<211> 489
<212> DNA

<213> Agrobacterium aurantiacum

<400> 368
atgacaaact tcttgatcgt tgtggccacc gtgcttgtta tggaattgac cgcctactcg 60

gtgcatagat ggataatgca cggaccattg ggttggggat ggcacaaatc acaccatgag 120

gaacacgatc acgccttaga aaaaaatgac ctctatggct tggtgttcgc tgtcattgcg 180

accgttttgt tcaccgtcgg ttggatctgg gcaccagtgt tgtggtggat tgcacttggt 240

atgacggttt atggcctcat atattttgtg ctccacgatg gtctagttca ccaaaggtgg 300

cctttcagat acattcccag gaagggttat gcacggcgat tatatcaagc ccatcgtctt 360

catcacgctg tggagggaag ggatcactgt gtcagctttg ggttcatcta cgcaccaccc 420

gttgacaagt tgaaacaaga tttgaaaatg tctggtgtct tgcgggctga ggcccaagag 480

aggacgtaa 489

<210> 369
<211> 729
<212> DNA
<213> Agrobacterium aurantiacum

<400> 369
atgagtgccc atgccttgcc aaaggcagat ctaactgcca ctagtttaat tgtgtcgggc 60

ggtattatcg cggcttggtt ggcattgcat gtacatgcat tgtggttctt ggacgcggca 120

gcgcacccta tcctcgcaat cgccaatttt cttggactaa cgtggttgtc agtaggtcta 180

tttatcatcg cccacgatgc gatgcacgga tccgtagtcc ctggtagacc tagagcaaat 240

gctgcaatgg gccagttggt gctttggttg tacgctggct tcagttggag aaagatgatt 300

gttaagcata tggcacatca ccgtcacgca ggtaccgatg acgacccaga ttttgatcat 360

ggaggtccag tgcgatggta cgcacgcttc ataggaacct actttggctg gagagaaggt 420

ttgcttttgc cagtcatagt aaccgtctac gctttaatac ttggagatcg atggatgtac 480

gtcgtgttct ggccattgcc atcaatcttg gcctctatcc aattgttcgt ttttgggact 540

tggttaccac atagacccgg tcacgatgca ttcccagata gacacaacgc ccgatcatcc 600

agaatatccg atccagtttc tttgctaact tgcttccact tcggtggtta tcaccacgaa 660

caccatttgc atccgacggt cccatggtgg aggttgccta gcactagaac taaaggtgat 720

accgcttaa 729

<210> 370
<211> 1770
<212> DNA
<213> Callitropsis nootkatensis

<400> 370
atggcagaaa tgttcaatgg aaacagcagt aatgacggga gctcatgtat gccggtaaag 60

gatgcattga gacgtacggg caaccaccac ccaaatttgt ggactgacga tttcattcaa 120

tcgcttaata gtccatactc ggactcctcc taccataaac atcgcgaaat attgatcgat 180

gagattcgcg atatgttttc aaatggggaa ggggacgaat ttggcgtgtt ggaaaatatt 240

tggtttgtcg atgttgttca acggttagga atcgaccgtc actttcagga agagattaaa 300

acagcgttgg actacatcta caagttttgg aaccacgact cgatatttgg tgaccttaac 360

atggtcgctt tgggtttcag aatcctaaga ttgaaccgtt acgtggcaag cagcgacgtt 420

ttcaagaagt ttaagggtga agagggccaa ttttcaggct tcgaaagttc agaccaagat 480

gcgaagttgg aaatgatgct aaacttgtat aaagcaagtg agctagattt tccagacgag 540

gatatcctta aggaagctag ggcctttgcg agtatgtact taaagcacgt aatcaaggaa 600

tacggtgata tccaagagag caagaaccca ttgcttatgg agatagaata cacattcaag 660

tacccatggc gttgtcgttt gcctcgactt gaggcttgga atttcattca cattatgaga 720

caacaagact gtaatatttc attggctaac aacctttaca agattccgaa gatatacatg 780

aaaaagatct tggagttggc catcttggac tttaatattc tccagtcaca acaccaacat 840

gaaatgaaac tcatctccac gtggtggaag aactctagcg ctatccaatt ggacttttc 900

agacaccgcc acatcgaaag ctacttctgg tgggcttcac cgttgttcga gcccgagttt 960

tctacatgtc ggatcaactg tacgaagttg tcaactaaaa tgtttctact tgacgatata 1020

tacgacacat acggtactgt ggaagaacta aagccgttta cgaccacatt aacgaggtgg 1080

gacgtgtcta cagtcgacaa ccatccagac tacatgaaga tcgcttttaa cttcagctat 1140

gaaatctaca aggagatcgc ctccgaggcg gaacgtaaac acggtccatt cgtgtataaa 1200

tatttgcaat catgctggaa gtcctatata gaagcttaca tgcaagaagc agagtggatt 1260

gcttcaaatc atattccagg cttcgatgag tacctcatga atggtgtaaa gtcatccggc 1320

atgagaatat tgatgatcca cgcattaatc ctaatggata caccattgag cgacgagatt 1380

ttggaacaat tagatatacc gagttcgaag tcacaggcct tgttgtcgtt gattaccaga 1440

ctcgtggatg atgtcaagga ctttgaggac gaacaagccc atggggaaat ggcatcctcc 1500

attgaatgtt acatgaaaga taaccacggt tcaacaagag aagatgcatt aaactacttg 1560

aagatccgaa tcgagtcgtg tgtccaggag ttgaataagg aattgttaga accttcgaat 1620

atgcacggta gttttaggaa tctttatttg aatgttggca tgcgtgtcat atttttcatg 1680

ctcaacgatg gggacttgtt cactcattcg aacaggaagg aaattcagga tgccatcacc 1740

aagttcttcg tcgaaccgat tataccttaa 1770

<210> 371
<211> 1212
<212> DNA
<213> Candida viswanathii

<400> 371
atgaccctcc caccagtcta catcgtctcc acagcaagaa ctccaattgg ttccttccaa 60

ggttcgttat cttccttgac ctactccgat ttgggtgccc acgccgttaa ggcagcttta 120

gccaaggtcc cacagatcaa gccacaggac gtcgacgaga tcgtctttgg tggtgtcttg 180

caagccaacg ttggtcaggc accagccaga caagtcgcct tgaaggcagg cttgccagac 240

tctatcattg cctccaccat taacaaagtc tgtgcctccg gtatgaaggc cgtcattatc 300

ggtgcccaaa acatcatctg tggaaccagt gacatcgtcg ttgttggtgg tgccgaatcc 360

atgtctaaca ctccatacta cttgccaagc gccagaagcg gggccagata cggtgacgcc 420

gtcatggttg atggtgtcca gaaggatggt ttgttggatg tctacgaaga aaaattgatg 480

ggtgttgctg ctgaaaagtg tgccaaggac cacggattca gcagagaaga ccaggacaac 540

tttgccatca actcttacaa gaaggctggc aaggctttga gtgaaggtaa gttcaaatca 600

gaaatcgccc cagtcaccat taaaggattc agaggcaagc cagatactgt tattgaaaat 660

gatgaagaga ttggcaaatt caacgaagac agactcaagt ctgccagaac tgtcttccaa 720

aaggaaaacg gtactgttac tgctccaaac gcttctaaat tgaacgatgg tggtgctgcc 780

ttggtcttgg tttctgaagc caagttgaag caattgggct tgaagccatt ggccaagatc 840

tctggttggg gtgaagctgc cagaactcca ttcgatttca ccattgcccc agctttggct 900

gttccaaagg ctgtcaagca cgctggtttg accgttgaca gagtcgactt ctttgaattg 960

aacgaagcct tctccgttgt tggtttggcc aacgctgaat tggtcaagat cccattggaa 1020

aaattgaacg tctacggtgg tgctgtcgcc atgggtcacc cattgggttg cagtggtgct 1080

agaattattg tcaccttgtt gtccgtcttg acccaagaag gcggtagatt tggtgctgct 1140

ggtgtctgta acggtggtgg tggtgcctct gccatcgtca ttgaaaagat tgactccgat 1200

gccaagttgt aa 1212

<210> 372
<211> 1350
<212> DNA
<213> Candida viswanathii

<400> 372
atgactaacg caccacaaaa cattggtatc aaaggaatcg aagtctacat tccaggccaa 60

gctgtcaatc aatcagactt ggaaaaattc gacggcatcc cagccggcaa gtacaccatc 120

ggcttgggcc agacaaacat ggcttttgtc aacgacagag aggacatcta ctccatcgcc 180

ttgactgttg tctccagatt gatcaagcat tacaatgttg acaccaacaa cgtcggtcgt 240

ttagaggtcg gtaccgagac cttgttggac aagtccaagt ctgtcaagtc cgtgttgatg 300

caattgttcc ccgacaacaa cgacattgaa ggtatcgaca ctgtcaacgc ctgctacggt 360

ggtacctccg ctgtcatcaa tgccatcaac tggatcgagt cgtcttcctg ggacggcaga 420

gacgccattg ttgttgctgg tgacattgct atctacgaca agggtgctgc tagaccaact 480

ggtggtgctg gttccattgc tatgttgatt ggtcctgacg ctcctattgt ctttgactct 540

gtgcgtggct cattcatgga acacgcctat gatttctaca agccagactt caccagtgaa 600

tacccagttg tcgatggtca cttctccttg tcctgttacg ttaaggctgt tgacaactgt 660

tacaagaact actccaagaa ggtcaccggc aacgtcgaca agaccgtcgg cgtttacgac 720

cactttgact acaacgcttt ccacgtgcca acctgtaagt tggtcaccaa gagttacgcc 780

agattgttgt acaacgacta caagtcggat ccatccaagt ttgctgactt gattgacgaa 840

agcaccagac aacacattga cagtttgtcc tacgaagcat ccttgaccga caaggtcttg 900

gaaaaaactt tcgtcacatt ggccaaagaa gagaccaaaa agagagtgca accagccttg 960

caagtgccaa ccaacaccgg taacatgtac actgcttcgt catgggtctc cttggcctcg 1020

ttgttgtact atgttggcgc tgaaaacttg aaagaaaaga gaatcggtct cttctcctac 1080

ggttcaggtt tggcctccac cttgttgtcc gtcactgttg tcggcgacgt ttctccaatc 1140

accaaggtct tggacttcga ctacaaatta ggcgaaggaa gaaagatcca gagcccagaa 1200

gaatacttgg ctgccattga attgagagaa aaggcacact tgcaaaagag cttcaagcca 1260

caaggttccc tcgataactt gagccaaggc acctactact tgactgaggt tgacgacaag 1320

tttagaagaa gttatgctat aaaggaataa 1350

<210> 373
<211> 3120
<212> DNA
<213> Candida viswanathii

<400> 373
atgctaagtt tcatcaccga agcaagtggc agaatagccc agactgcagc ccacagacca 60

attcacttca tggtgatacc tgcgttacta gcatccatag catacttgtc catcatcgat 120

gattacgtcc ccgaacacat cagagcccag tacatgtcag gcgtcagcta tttccaccct 180

caagggtccg cctcggattt ggacaagtgg atcgaaatcc acgacactac ccaatacgcc 240

aacgccaacc aaatcagtgt catccctttg agattcagac gattccatga ctcaatccca 300

cagatcgcta acgccatcaa gatttcaaac aatgaacaga tcttgattgt tccttcggat 360

aaagcagagt cgacagtaga aggtttgagc gagatcactg agaatggcat cacttggagg 420

gctagaaata acgacaagtt gtccaagtat tatgattacg ccaggtacgg gttgcttaga 480

atccaggacg ctatacacaa tgctgataat ttcgacatcc tcttggtgtt tgttgcctac 540

cttggaatgt ggtactcact tatcaaagtg tttattgaca tgagacgcat tggctccaag 600

ttttggttag cctttggcac gttgacttcc tcgacttttt cattcttgtt tgccttggtc 660

atttcaaaca aatttttgga cgctaaagtt tccttgagaa gtctttctga atgcatccca 720

tttttggttg ccattattgg gttcaaacat aaagtggcca tcacgactag tgttgcccag 780

tcatccactt cttcacccga ggatgttcct catgtcgtgg gcaaagctgt ctcagaccag 840

tgtttgttta tcttgagaga ccatttggtt gtcattattg gatttctcgc ctgtgcggct 900

tatggcaatg aattgaaagg tttaagaaac ttctgtatct tgggggcatt gattttgtcc 960

tttgacattg tttccgttta tactttcttt tctgcgattt tggccttgaa agtcgagatt 1020

aaccgtgccc gccgtactca agatttacaa cacgtattgg aagaggacgg catttcatcc 1080

ttagttgccg ctagagttgc cgaacgaaat gcaactattg aacaccccaa cgaaaccaac 1140

ttcttctcat ctaataattc ttccattgtt tatttcaaag ttattatgag tcttgggttt 1200

ttcgccttcc atgcattttg gttgggcagc tcgtggttat acaacacttc tgatggcggt 1260

agccatggaa gcttttcgtt cctcagcaac atcccattat tgacacaaga catctccaac 1320

tccatcccaa ttggcagaag aggaaccgtg gttaccattt tgccaaccac cttctacatg 1380

ccatccggta tcatcatcca atttgaagat atggtctact tagcattgag caagatcagc 1440

agtgctatca gagacagttt gatttccaag tgtattgtgt ttgcattgac catttccatt 1500

gtcaccaaca tctacttctt gaatgctgcc cgtttccaag tttcagctac ccgtaagttg 1560

attgaccaag aaatgtctcg tccaaaacaa gccgccgccg ccgctgctgc tgctccatct 1620

gctgcaaaat cagtggcgcc agaagaagac gaagacgaaa cctccagtga agaacttgaa 1680

atcaaagctc ctgtaaagcc gttatcactt gaagaatgta ccagaatcct caaagagggc 1740

aaagtcaaga ccttgtccaa cgctgaagtg tctagcttgg ttgttggcgg taaattgcca 1800

ttgtacgctt tggagaaaca attgggcgac aacaagagag ccgtggctgt tcgtcgtaag 1860

gctattgcaa aattagctaa cgcgccagta ttagaaacca accgtttgcc ttactcccac 1920

tatgattatg atagagtctt tggtgcttgt tgtgagaatg ttattgggta catgccaatc 1980

cctgttggtg ttgctggacc attagtgatt gatggcaagc cataccacat ccccatggcc 2040

actaccgaag gttgtcttgt tgcctctact atgcgtggct gtaaagctat taacgctgga 2100

ggtggtgttg aaaccgtctt aaccaaagat ggtatgacta gaggtccttg tgttcgtttc 2160

ccaactttga aaagagccgg tgctgcaaag ttgtggattg actccgaaga aggtcaaatc 2220

accatcaaaa aggcattcaa ctcgacttct agatttgctc gtttgcaaca tattcaaact 2280

gccctcgctg gtacctcgtt gtttattaga tttagaacta ctactggtga tgctatgggc 2340

atgaacatga tttccaaagg tgttgaatat tccttgaagc acatggttga agaatgtggc 2400

tgggacgata tggaagttat ctctgtttct ggtaactact gtaccgataa gaaaccagct 2460

gctattaact ggattgaagg tagaggtaag tctattgttg ctgctgctac catccctgct 2520

gatgttgtca ctaaagtctt gaaatccgat gtcgatgcct tagttgaatt gaatgtttcc 2580

aaaaacttgg ttggctcagc catggccggt tcagttggtg gtttcaatgc tcatgcagcc 2640

aacttggtca ctgctgttta cttggcctgt ggtcaagacc ccgcacagaa tgtcgaatcc 2700

tcaaactgta ttaccttgat ggaaaaggac aaacagactg gcgacttggt catttctgtt 2760

tctatgccat ctattgaagt tggtacgatt ggtggtggta ccattttgga accacaagca 2820

gccatgttgg acttgttggg cgtgcgtggc cctcacccaa ccaatccagg tgacaatgca 2880

agacaattgg ccaagattgt tgcttcggca gtgttagcag ctgaattgtc actttgttct 2940

gctttggccg ctggtcactt ggtgcaatcc cacatgcaac acaaccgtaa aggagccacc 3000

ccagctgcgg cccctgcaat cagcaacgga agtgccaagg gaaccaaaac caatgggtct 3060

ataaatggga aagacttgaa gcgtctcaaa gacggatccg ttacatgtat aaagtcatag 3120

<210> 374
<211> 1299
<212> DNA
<213> Candida viswanathii

<400> 374
atgtcagtct cgcctttcgt tgtcagtgct ccagggaaag tcatcatctt tggagaacat 60

tcagcagttt atgggaaacc cgccattgca gccgccttga gtttgcgatg ttacctactc 120

gtttccccgt cagttgatga tgcaaacaca atcagattgc agttcccaga catccaattg 180

gaccattcct ggaacatcaa cgatatcccc tgggatgaaa tcaaaccgtt cgttaaatat 240

gacgccaata acaaaccact caccccatcg gagttggtac cagagatcct cgataagttg 300

tctcccttgt tgactgattt cgataacaaa atgcattact acgcttgttt ttgtttcctt 360

tacttgtacg tcaatttatg ttccactgaa actccaggaa ccaccttcat cgtcagatcc 420

accttgccaa tcggtgccgg gttagggtcg tctgcatcca cgtctgtgtg cttgtcgtct 480

gcattagcat tgttaggcgg ctgtatcagt gagccttcga tatcagccac cgacaagatc 540

ctcaacgaag acatacccga tttggagttc gtcgataagt ggtccttgat aggtgagaag 600

tgcttccacg ggaacccatc cggtatcgac aacgcagtag ccacgtttgg cggtgccgtc 660

atgtttcaaa gaacttcggc tcctgagcaa ccatccatca gaaccaacat gagaaacttc 720

ccagctatca aactcttact tacaaacacc aaagtcccaa aaagtacagc tgatttagtc 780

gccggtgttg gcaaactcaa cgcggagttc aacccgataa ccacgtcgat tttgaccgct 840

atggagcatt tgtcacagga agcttacaaa gtcatgatca ctccggggtt cggcaaggac 900

gaaacaaaca cgttgcgtaa gttggtcaat atcaaccatg gtctattggt agcgttgggc 960

gtgtcacatc cgtcgctaga aaccgtcaag atcattggtg acagacacaa gattggcgct 1020

actaagttga ctggcgctgg tggtggcggg tgtgcgatca ctttggttaa tgataatgtc 1080

gaggagtctg ttatccaaga agctatcaaa gagtttgcaa ctgagggata tgagagtttt 1140

gaaacgtcgt tgggtggtaa aggtgtcggt gtcctattta gtggtgatgt ttctgataat 1200

ggcaagtttt ctcaatctgt gttctgcaac tatcccgacc gtgctagcat tgaagatgct 1260

ttaggaatga taaacgttaa agaatggaag ttctggtga 1299

<210> 375
<211> 1263
<212> DNA
<213> Candida viswanathii

<400> 375
atgtcaaaag cttttagtgc ccctggtaaa gcattgttag ctggcggata cttggttctt 60

gagccaacat atgatgcata tgtgacggca ttgtcctcga gaatgcacgc tataataaca 120

ccccagaagc cagcatccat ttcaaaaatc aaaatcagtt cgcctcagtt tgccaatggt 180

gagtgggagt atcacgtcac gtctaatgag aaacccaagg acatcaagtc cagaagcaat 240

ccgtttttgg aggctaccat tttcattgtt ttgtcttata tccaaccaac agaaccattc 300

gatctcgatt tggtcatata ttcagaccca ggataccatt cacaagaaca caccacgcag 360

aaagtgtcca gcaacggcaa gaaaaagttc ctttatcatt cgcgtgcaat taacgatgtt 420

gagaagacag gattaggttc ttcggccggt ctagtttctg tcgtcactac cagtttgtta 480

tcctacttta ttccgggaat tgaagagtca aacaaggaca tgttgcacaa tgtagcacag 540

attgcccatt gttttgccca gaagaagatt gggtccgggt tcgatgtggc cactgccatc 600

tatggtctga ttgtctaccg aagattccag ccaagcttga tcaacgacgt tttcgagatc 660

ttagaggaaa ccccgggtag attccccggt gctttgaagt cgttagttga gtctaattgg 720

gaattcaaac atgaaaggtg cgtactacca ccaaagatca agctcttgat gggagatatc 780

aaaggaggat cagagactcc gaaattggtt tccaaaatac tccaatggaa gaaggacaaa 840

ccagaagaaa gtgggttggt gtatgatcaa cttaatagcg caaatgtggc gttcatgaag 900

aagatcagta cattgaatga atcgagtcag gtccaagaga ttgacgaatt gtctgattac 960

ataagcgcgg tacgcaaagg gttgcaagaa ttgacagaaa aatccaaagt cccagttgaa 1020

ccacctgtgc aaactgaatt gttggaccgc attgcaaaac tcccaggttg tctaggtggg 1080

gttgtgcccg gtgctggtgg gtacgatgct atagctgtgc ttgtgttgga aaaggaggtt 1140

gagaacttta gaaacaagac acttgaaaac ccagagtact accacaacgt gtactgggtt 1200

gacttggagg aagaaaccga gggtgtagtg gtggaaaact gggaggacta tatcggttta 1260

tag 1263

<210> 376
<211> 1166
<212> DNA
<213> Candida viswanathii

<400> 376
atgtattcag cttcagtcac agcacctgtc aacatagcgg tatgtatgtt cccacctagt 60

tgctcctccc cgactcatca ctaacaccct tatagaccct caagtactgg gggaaacgcg 120

acaagaccct aaacttgcca accaactcct ccatatcagt caccctctcc caggatgatt 180

tgcgaacctt aaccacagca gcagcgtcca ccactttcga caaggaccaa ttgtggctca 240

atggcaagtt ggagtcgttg gacactccga gaacgcaagc gtgcttggtc gatttaagga 300

aattgagagc cgatgttgag caagccaacg ctgacctccc aaagctttcc acgatgaagc 360

tccacattgt ttcggaaaac aacttcccca ccgccgctgg tttggcttcc tcggctgctg 420

ggtttgctgc gttgatcaca gccatcgcca agttgtacga attgccccaa gacatgtcgg 480

agttgtccaa gattgctcgt aaaggttctg gttctgcgtg cagatccctc tttggcgggt 540

ttgttgcgtg ggagatgggc gacgccgccg acggacaaga ctccaaggca gttgaagttg 600

cgcctttgga ccactggcct agcatgaggg cggttatttt ggtggtcagc gacgacaaaa 660

aggacacccc aagcaccacc ggtatgcaag ccaccgtgca aacctccgac ttgtttgccc 720

accgcgtcac caaagtcgtc ccccagtgtt tcgaagagat gaagaaggcc atcgtcgcca 780

aggactttcc taaatttacc gagttaacca tgaaggattc caactcgttc cacgctgttt 840

gtttggactc gtaccctcca atcttctacc tcaacgacac ctccaagaag atcatcaagt 900

tggtggaggc catcaacaag cacgccacca tcgctgccta cacgttcgac gcgggcccca 960

acgccgtcat ctactacgac gccgctaacg aagacgaggt cttgtcccag ctctacaagt 1020

cgtttggcca cgttcaagga tggaagaaag cctacaccgc ggaaacagcc gttgccggtg 1080

tttcacgtat tattcaaaca tctattggtc aaggaccgca agttacaaac gaatcattaa 1140

tcaacgaaag tgggttacca aactaa 1166

<210> 377
<211> 861
<212> DNA
<213> Candida viswanathii

<400> 377
atgtcatcag aatacgctaa actagtggca agtttctccc caaacgacat attggctaaa 60

tggccagagg tcactccatt gaagaagata tcaggcatcc caagatcagc cgagtcagac 120

agctccaacg gctcccacaa caacacagaa cttttcaatg gccacgacga ggagcaaatc 180

agattgatgg aggagctttg cattgttttg gactacgacg acaagcctgt tggggcagga 240

acaaagaagt tgtgtcatat catggacaat atcaacgaag gattgttgca ccgtgccttc 300

tcggtgtttt tattcaatga ggacgggaaa ttattattac aacaacgtgc cgatgaaaaa 360

atcaccttcc ctgccatgtg gacaaatact tgctgttccc atcctttgtg tgttcctagt 420

gagttggggg ttgatgccga tgccaaagat gtcaacaact tggataacgc cgtgcatgga 480

gccaaggtgg ccgcacaaag aaaattggac catgagttgg gcatcccatt cagtgacact 540

cccttggacc agttcacgta cttgactaga attcactata aatcagccag tggtgcagaa 600

gactccaaat ggggtgagca tgaaatagac tatattttga tattgaagac caagaatgac 660

atcaccatca atgccaacta taatgaagtc aaggattata agtatgtgga cgccaaggag 720

ttgcaggaga tgtttgaaga taaggacttg gtgttcaccc cgtggtttaa attgatctgt 780

caatctttct tattcaaatg gtggaataat ttaagtgatt tagaaaagta ccaagatact 840

gagatacata gattacttta a 861

<210> 378
<211> 1056
<212> DNA
<213> Candida viswanathii

<400> 378
atgtctgata aagcagccgc tagagagaga ttcctctctg tttttgagtg tgccgtcgag 60

gaattgaaag aagtcttggt ttctcacaag atgccgcaag aagcaattga ctggtttgtc 120

aagaacttga actacaacac ccccggcggt aagttgaaca gaggtttgtc tgttgtcgac 180

acctacgcta tcttgaacaa caccaccgct gacaagttga acgatgaaca atacaagaag 240

gtcgccttgt tgggctggtc aattgaattg ttgcaagctt actttttggt tgccgatgat 300

atgatggacc aatccaagac cagaagagga cagaaatgtt ggtacttggt cgaaggtgtt 360

ggaaacattg caattaatga ctccttcatg ttggaaggtg ccatttacgt cttgttgaag 420

aagcacttcc gtcaagatcc atactatgtc gacttgttgg acttgttcca cgaagtcacc 480

ttccagaccg aattgggtca attattggac ttggtgactg ctgatgaaga agtcgtcgac 540

ttggacaagt tctccttgga caagcactcg ttcattgtca ttttcaaaac cgcatactac 600

tccttctact tgcctgttgc tttggccatg tacatgagcg gtatcagcag cgaagaagac 660

ttgaagcaag tcagagatat cttgatccca ttgggtgagt acttccaaat ccaggacgat 720

ttcttggact gtttcggaac cccagaacaa attggcaaga tcggtactga tatcaaagac 780

aacaagtgtt cctgggtggt caaccaagct ttgttgcatg ctactccaga acaacgtaag 840

ttgttggacg acaactacgg taagaaagac gacgagtctg aacagagatg caaggacttg 900

ttcaagtcca tgggcattga aaagatctac cacgactacg aagagtcaat tgttgctaaa 960

ttaagagaac aaatcgataa agttgatgaa tcaagaggtt tgaaaaaaga tgtcttgacc 1020

gctttcttgg gcaaggttta caagagatcc aaatag 1056

<210> 379
<211> 996
<212> DNA

<213> Candida viswanathii

<400> 379
atgtcattca acatcgatag cttaatccaa ccaggcgccg cctacgaccc atctatgacc 60

gacgccatca tgaagccata caggtacatc tccgaggtcc ccagcaacaa ccacaacgtg 120

cgtacgcgct tcttgcttgc attcaacgag ttgttctacg gcatgaagaa cgaggacttg 180

ctccaccgca tcagccacat catctccgtg ttccacaact cgtccttgtt gatcgacgac 240

atcgaggacg actcgcagtt gcgccgcggc atgccggtgg cacacgtcaa gtacggcgtc 300

ccgttgacca tcaactgcgg caacatgatg tacttcgtcg cggtccagaa agcgattgac 360

ttggcgggcg aggctgggtc cgcggagttg aagttcgaga cgtcgcagat cttggttgac 420

gaaatgatga atgttcacca tggacaaggc ttggatattt actggagaga ctacttgaag 480

gacctcgagc acttgcctga gatcgaggat tacttgggca tggtcaagga caagactggg 540

agcttgttca gattggccat caaattgttg ctgttgcatt cggacgtggg cgaggacaat 600

ggtttggtcg cgattgccaa cttgttgggc ataatctacc aaatcagaga cgactacttg 660

aacttggtcg atatcaagta ctccgccatg aagggcgtta cttgtgaaga cttgattgaa 720

ggtaagttat ccttgcctat tttgcattgc ttgagaacca cgactaattc cccggttcac 780

gagatcctat acaactacaa caccagcgct gagcgcgcca aacagaacgc cttgattgaa 840

gagtgcttga cttacatgaa gaacaagtcg agatccttgc aatatacttt ggacttgatc 900

aagaccttgg agcgcaagat caaggctatg atgaccaagt atccaaacct ggagaactcc 960

ggcttgataa aaatcattga caggttgtgt gatctt 996

<210> 380
<211> 855
<212> DNA
<213> Candida viswanathii

<400> 380
atggacgacg ttgattctgc tttagccgat aatgttaaat ccttcgctgc cggtggtttt 60

ggtggtattt gtgccgtgtt gaccggtcat ccattcgact tggtcaaggt cagattgcaa 120

actgggttgt acaaatcgtc agtgcaatgt gtcaaagaaa caatagccaa agacgggttg 180

tttgggttat acagaggtgt cttgcctcca ttgttgggtg tcaccccaat gtttgctgtt 240

tccttctggg gttacgacgt tggtaagaaa ttggtttctt ccttcactgg caaatcagtc 300

gacaagtttg aaatcaaaga catttcgact gccggtttca tcagtgccat cccaaccacc 360

ttggttgccg ctccattcga aagagtcaag gtcatgatgc agatccaaga aggtgccaag 420

tccaagtcca tgggtgccgt tgttgctgaa atgtacagaa ccggtggtat cagatccatc 480

ttcaagggta ctgtggctac tttggccaga gatggtccag gttctgcctt gtatttcgct 540

acttatgaat gggtcaagaa ggaattgact gcccctggtg aagacttgtc cttgtttgct 600

attaccactg ctggtggttt cgctggtatt gccatgtggt tgggtgtttt cccaattgac 660

accatcaagt ctacccagca gtcttccaac gtcaaggtct ccattgtgca ggccactaag 720

aacatctacg ctaaaggtgg tatcaaggct ttcttccctg gtgttggtcc tgccttggcc 780

agagctttcc cagccaacgc tgctaccttc ctaggggttg aattggccag aaagttcttg 840

gacaaagtta tataa 855

<210> 381
<211> 2091
<212> DNA
<213> Candida viswanathii

<400> 381
atgggtgccc ctttaacagt cgccgttggc gaagcaaaac caggcgaaac cgctccaaga 60

agaaaagccg ctcaaaaaat ggcctctgtc gaacgcccaa cagactcaaa ggcaaccact 120

ttgccagact tcattgaaga gtgttttgcc agaaacggca ccagagatgc catggcctgg 180

agagacttgg tcgaaatcca cgtcgaaacc aaacaggtta ccaaaatcat tgacggcgaa 240

cagaaaaagg tcgataagga ctggatctac tacgaaatgg gtccttacaa ctacatatcc 300

taccccaagt tgttgacgtt ggtcaagaac tactccaagg gtttgttgga gttgggcttg 360

gccccagatc aagaatccaa gttgatgatc tttgccagta cctcccacaa gtggatgcag 420

accttcttag cctccagttt ccaaggtatc cccgttgtca ccgcctacga caccttgggt 480

gagtcgggct tgacccactc cttggtgcaa accgaatccg atgccgtgtt caccgacaac 540

caattgttgt cctccttgat tcgtcctttg gagaaggcca cctccgtcaa gtatgtcatc 600

cacggggaaa agattgaccc taacgacaag agacagggcg gcaaaatcta ccaggatgcg 660

gaaaaggcca aggagaagat tttacaaatt agaccagata ttaaatttat ttctttcgac 720

gaggttgttg cattgggtga acaatcgtcc aaagaattgc atttcccaaa accagaagac 780

ccaatctgta tcatgtacac ctcgggttcc accggtgctc caaagggtgt ggttatcacc 840

aatgccaaca ttgttgccgc cgtgggtggt atctccacca atgctactag agacttggtt 900

agaactgtcg acagagtgat tgcatttttg ccattggccc acattttcga gttggccttt 960

gagttggtta ccttctggtg gggggctcca ttgggttacg ccaatgtcaa gactttgacc 1020

gaagcctcct gcagaaactg tcagccagac ttgattgaat tcaaaccaac catcatggtt 1080

ggtgttgctg ccgtttggga atcggtcaga aagggtgtct tgtctaaatt gaaacaggct 1140

tctccaatcc aacaaaagat cttctgggct gcattcaatg ccaagtctac tttgaaccgt 1200

tatggcttgc caggcggtgg gttgtttgac gctgtcttca agaaggttaa agccgccact 1260

ggtggccaat tgcgttatgt gttgaatggt gggtccccaa tctctgttga tgcccaagtg 1320

tttatctcca ccttgcttgc gccaatgttg ttgggttacg gtttgactga aacctgtgcc 1380

aataccacca ttgtcgaaca cacgcgcttc cagattggta ctttgggtac cttggttgga 1440

tctgtcactg ccaagttggt tgatgttgct gatgctggat actacgccaa gaacaaccag 1500

ggtgaaatct ggttgaaagg cggtccagtt gtcaaggaat actacaagaa cgaagaagaa 1560

accaaggctg cattcaccga agatggctgg ttcaagactg gtgatattgg tgaatggacc 1620

gccgacggtg gtttgaacat cattgaccgt aagaagaact tggtcaagac tttgaatggt 1680

gaatacattg ctttggagaa attggaaagt atttacagat ccaaccactt gattttgaac 1740

ttgtgtgttt acgctgacca aaccaaggtc aagccaattg ctattgtctt gccaattgaa 1800

gccaacttga agtctatgtt gaaggacgaa aagattatcc cagatgctga ttcacaagaa 1860

ttgagcagct tggttcacaa caagaaggtt gcccaagctg tcttgagaca cttgctccaa 1920

accggtaaac aacaaggttt gaaaggtatt gaattgttgc agaatgttgt cttgttggat 1980

gacgagtgga ccccacagaa tggttttgtt acttctgccc aaaagttgca gagaaagaag 2040

attttagaaa gttgtaaaaa agaagttgaa gaggcataca agtcgtctta g 2091

<210> 382
<211> 1953
<212> DNA
<213> Candida viswanathii

<400> 382
atgtcaggat tagaaatagc cgctgctgcc atccttggta gtcagttatt ggaagccaaa 60

tatttaattg ccgacgacgt gctgttagcc aagacagtcg ctgtcaatgc cctcccatac 120

ttgtggaaag ccagcagagg taaggcatca tactggtact ttttcgagca gtccgtgttc 180

aagaacccaa acaacaaagc gttggcgttc ccaagaccaa gaaagaatgc ccccaccccc 240

aagaccgacg ccgagggatt ccagatctac gacgatcagt ttgacctaga agaatacacc 300

tacaaggaat tgtacgacat ggttttgaag tactcataca tcttgaagaa cgagtacggc 360

gtcactgcca acgacaccat cggtgtttct tgtatgaaca agccgctttt cattgtcttg 420

tggttggcat tgtggaacat tggtgccttg cctgcgttct tgaacttcaa caccaaggac 480

aagccattga tccactgtct taagattgtc aacgcttcgc aagttttcgt tgacccggac 540

tgtgattccc caatcagaga taccgaggct cagatcagag aggaattgcc acatgtgcaa 600

ataaactaca ttgacgagtt tgccttgttt gacagattga gactcaagtc gactccaaaa 660

cacagagccg aggacaagac cagaagacca accgatactg actcctccgc ttgtgcattg 720

atttacacct cgggtaccac cggtttgcca aaagccggta tcatgtcctg gagaaaagcc 780

ttcatggcct cggttttctt tggccacatc atgaagattg actcgaaatc gaacgtcttg 840

accgccatgc ccttgtacca ctccaccgcg gccatgttgg ggttgtgtcc tactttgatt 900

gtcggtggct gtgtctccgt gtcccagaaa ttctccgcta cttcgttctg gacccaggcc 960

agattatgtg gtgccaccca cgtgcaatac gtcggtgagg tctgtcgtta cttgttgaac 1020

tccaagcctc atccagacca agacagacac aatgtcagaa ttgcctacgg taacgggttg 1080

cgtccagata tatggtctga gttcaagcgc agattccaca ttgaaggtat cggtgagttc 1140

tacgccgcca ccgagtcccc tatcgccacc accaacttgc agtacggtga gtacggtgtc 1200

ggcgcctgtc gtaagtacgg gtccctcatc agcttgttat tgtctaccca gcagaaattg 1260

gccaagatgg acccagaaga cgagagtgaa atctacaagg accccaagac cgggttctgt 1320

accgaggccg cttacaacga gccaggtgag ttgttgatga gaatcttgaa ccctaacgac 1380

gtgcagaaat ccttccaggg ttattatggt aacaagtccg ccaccaacag caaaatcctc 1440

accaatgttt tcaaaaaagg tgacgcgtgg tacagatccg gtgacttgtt gaagatggac 1500

gaggacaaat tgttgtactt tgtcgacaga ttaggtgaca ctttccgttg gaagtccgaa 1560

aacgtctccg ccaccgaggt cgagaacgaa ttgatgggct ccaaggcctt gaagcagtcc 1620

gtcgttgtcg gtgtcaaggt gccaaaccac gaaggtagag cctgttttgc cgtctgtgaa 1680

gccaaggacg agttgagcca tgaagaaatc ttgaaattga ttcactctca cgtgaccaag 1740

tctttgcctg tgtatgctca acctgcgttc atcaagattg gcaccattga ggcttcgcac 1800

aaccacaagg ttcctaagaa ccaattcaag aaccaaaagt tgccaaaggg tgaagacggc 1860

aaggatttga tctactggtt gaatggcgac aagtaccagg agttgactga agacgattgg 1920

tctttgattt gtaccggtaa agccaaattg tag 1953

<210> 383
<211> 1566
<212> DNA
<213> Candida viswanathii

<400> 383
atggaccaag aaatgtctcg tccaaaacaa gccgctgctg ctgctgttgc tccatctgcc 60

gcaaaatcag tggcgccaga agaagacgaa gacgaaacct ccagcgagga acttgaaatc 120

aaagctcctg tcaagccatt accacttgaa gaatgtacca gaatcctcaa agagggcaaa 180

gtcaagacct tgtccaatgc tgaagtgtct agtttggttg ttggtggtaa attgccattg 240

tacgctttgg agaagcaatt gggcgaccac aagagagccg tggctgttcg tcgtaaggct 300

attgcaaaat tagctaatgc gccagtatta gaaaccaacc gtttacctta ctcccactat 360

gattatgata gagtttttgg tgcttgttgt gagaatgtta ttggatacat gccaatccct 420

gttggtgttg ctggaccatt agttattgat ggcaagccat accacatccc catggccact 480

actgaaggtt gtcttgttgc ctctactatg cgtggttgta aagctattaa cgctggaggt 540

ggtgttgaaa ctgtcttgac caaagatggt atgactagag gtccttgtgt tcgtttccca 600

actttgaaaa gagccggcgc tgcaaagttg tggattgact ccgaagaagg tcaaatcacc 660

atcaaaaagg cattcaactc gacttctaga tttgctcgtt tgcaacatat tcaaactgcc 720

ctcgctggta cctcgttgtt tattagattt agaactacta ctggtgatgc catgggtatg 780

aacatgattt ccaaaggtgt tgagtattcc ttgaagtaca tggttgaaga atgtggctgg 840

gacgatatgg aagttatctc tgtttctggt aactactgta ccgataagaa accagctgct 900

attaactgga tcgaaggtag aggtaagtct attgttgccg ctgctaccat tcctgctgat 960

gttgtcacta aagtcttgaa atccgatgtc gatgcattag ttgaattgaa tgtttctaag 1020

aacttggttg gttcagccat ggccggttca gttggtggtt tcaacgccca tgcagccaac 1080

ttggtcacag ctgtttattt ggcctgtggt caggaccctg cacaaaacgt cgaatcttca 1140

aactgtatta ctttgatgga aaaggacaaa cagactggtg acttggtgat ttctgtttct 1200

atgccatcta tcgaagttgg tacgattggt ggtggtacca ttttggaacc acaagcagcc 1260

atgttggact tgttgggcgt gcgtggtcct cacccaacca accctggtga caatgcaaga 1320

caattggcca agattgttgc ttcggcagtg ttagcagctg aattgtcact ttgttccgct 1380

ttggctgctg gtcacttggt gcaatcccac atgcaacaca accgtaaagg agctaccgcc 1440

gccgccgccg ccgcagcagc ccctgcaatc gccaacggcc gtgccaatgg aaccaaaacc 1500

aatgggtcta tcaatgggaa agacttgaag cgtctcaaag acggatcagt tacatgtata 1560

aagtca 1566

<210> 384
<211> 2130
<212> DNA
<213> Candida viswanathii

<400> 384
atgactttta caaagaaaaa cgttagtgta tcacaaggtc ctgaccctag atcatccatc 60

caaaaggaaa gagacagctc caaatggaac cctcaacaaa tgaactactt cttggaaggc 120

tccgtcgaaa gaagtgagtt gatgaaggct ttggcccaac aaatggaaag agacccaatc 180

ttgttcacag acggctccta ctacgacttg accaaggacc aacaaagaga attgaccgcc 240

gtcaagatca acagaatcgc cagatacaga gaacaagaat ccatcgacac tttcaacaag 300

agattgtcct tgattggtat ctttgaccca caggtcggta ccagaattgg tgtcaacctc 360

ggtttgttcc tttcttgtat cagaggtaac ggtaccactt cccaattgaa ctactgggct 420

aacgaaaagg aaaccgctga cgttaaaggt atctacggtt gtttcggtat gaccgaattg 480

gcccacggtt ccaacgttgc tggtttggaa accaccgcca catttgacaa ggaatctgac 540

gagtttgtca tcaacacccc acacattggt gccaccaagt ggtggattgg tggtgctgct 600

cactccgcca cccactgttc tgtctacgcc agattgattg ttgacggtca agattacggt 660

gtcaagactt ttgttgtccc attgagagac tccaaccacg acctcatgcc aggtgtcact 720

gttggtgaca ttggtgccaa gatgggtaga gatggtatcg ataacggttg gatccaattc 780

tccaacgtca gaatcccaag attctttatg ttgcaaaagt tctgtaaggt ttctgctgaa 840

ggtgaagtca ccttgccacc tttggaacaa ttgtcttact ccgccttgtt gggtggtaga 900

gtcatgatgg ttttggactc ctacagaatg ttggctagaa tgtccaccat tgccttgaga 960

tacgccattg gtagaagaca attcaagggt gacaatgtcg atccaaaaga tccaaacgct 1020

ttggaaaccc aattgataga ttacccattg caccaaaaga gattgttccc atacttggct 1080

gctgcctacg tcatctccgc tggtgccctc aaggttgaag acaccatcca taacaccttg 1140

gctgaattgg acgctgccgt tgaaaagaac gacaccaagg ctatctttaa gtctattgac 1200

gacatgaagt cattgtttgt tgactctggt tccttgaagt ccactgccac ttggttgggt 1260

gctgaagcca ttgaccaatg tagacaagcc tgtggtggtc acggttactc gtcctacaac 1320

ggcttcggta aagcctacaa cgattgggtt gtccaatgta cttgggaagg tgacaacaat 1380

gtcttggcca tgagtgttgg taagccaatt gtcaagcaag ttatcagcat tgaagatgcc 1440

ggcaagaccg tcagaggttc caccgctttc ttgaaccaat tgaaggacta cactggttcc 1500

aacagctcca aggttgtttt gaacactgtt gctgacttgg acgacatcaa gactgtcatc 1560

aaggctattg aagttgccat catcagattg tcccaagaag ctgcttctat tgtcaagaag 1620

gaatctttcg actatgtcgg cgctgaattg gttcaactct ccaagttgaa ggctcaccac 1680

tacttgttga ctgaatacat cagaagaatt gacacctttg accaaaagga cttggttcca 1740

tacttgatca ccctcggtaa gttgtacgct gccactattg tcttggacag atttgccggt 1800

gtcttcttga ctttcaacgt tgcctccacc gaagccatca ctgctttggc ctctgtgcaa 1860

attccaaagt tgtgtgctga agtcagacca aacgttgttg cttacaccga ctccttccaa 1920

caatccgaca tgattgtcaa ttctgctatt ggtagatacg atggtgacat ctatgagaac 1980

tactttgact tggtcaagtt gcagaaccca ccatccaaga ccaaggctcc ttactctgat 2040

gctttggaag ccatgttgaa cagaccaacc ttggacgaaa gagaaagatt tgaaaagtct 2100

gatgaaaccg ctgctatctt gtccaagtaa 2130

<210> 385
<211> 1989
<212> DNA
<213> Candida viswanathii

<400> 385
atgcctaccg aacttcaaaa agaaagagaa ctcaccaagt tcaacccaaa ggagttgaac 60

tacttcttgg aaggttccca agaaagatcc gagatcatca gcaacatggt cgaacaaatg 120

caaaaagacc ctatcttgaa ggtcgacgct tcatactaca acttgaccaa agaccaacaa 180

agagaagtca ccgccaagaa gattgccaga ctctccagat actttgagca cgagtaccca 240

gaccaacagg cccagagatt gtcgatcctc ggtgtctttg acccacaagt cttcaccaga 300

atcggtgtca acttgggttt gtttgtttcc tgtgtccgtg gtaacggtac caactcccag 360

ttcttctact ggaccataaa taagggtatc gacaagttga gaggtatcta tggttgtttt 420

ggtatgactg agttggccca cggttccaac gtccaaggta ttgaaaccac cgccactttt 480

gacgaagaca ctgacgagtt tgtcatcaac accccacaca ttggtgccac caagtggtgg 540

atcggtggtg ctgcgcactc cgccacccac tgctccgtct acgccagatt gaaggtcaaa 600

ggaaaggact acggtgtcaa gacctttgtt gtcccattga gagactccaa ccacgacctc 660

gagccaggtg tgactgttgg tgacattggt gccaagatgg gtagagacgg tatcgataac 720

ggttggatcc agttctccaa cgtcagaatc ccaagattct ttatgttgca aaagtactgt 780

aaggtttccc gtctgggtga agtcaccatg ccaccatctg aacaattgtc ttactcggct 840

ttgattggtg gtagagtcac catgatgatg gactcctaca gaatgaccag tagattcatc 900

accattgcct tgagatacgc catccacaga agacaattca agaagaagga caccgatacc 960

attgaaacca agttgattga ctacccattg catcaaaaga gattgttccc attcttggct 1020

gccgcttact tgttctccca aggtgccttg tacttagaac aaaccatgaa cgcaaccaac 1080

gacaagttgg acgaagctgt cagtgctggt gaaaaggaag ccattgacgc tgccattgtc 1140

gaatccaaga aattgttcgt cgcttccggt tgtttgaagt ccacctgtac ctggttgact 1200

gctgaagcca ttgacgaagc tcgtcaagct tgtggtggtc acggttactc gtcttacaac 1260

ggtttcggta aagcctactc cgactgggtt gtccaatgta cctgggaagg tgacaacaac 1320

atcttggcca tgaacgttgc caagccaatg gttagagact tgttgaagga gccagaacaa 1380

aagggattgg ttctctccag cgttgccgac ttggacgacc cagccaagtt ggttaaggct 1440

ttcgaccacg ccctttccgg cttggccaga gacattggtg ctgttgctga agacaagggt 1500

ttcgacatta ccggtccaag tttggttttg gtttccaagt tgaacgctca cagattcttg 1560

attgacggtt tcttcaagcg tatcacccca gaatggtctg aagtcttgag acctttgggt 1620

ttcttgtatg ccgactggat cttgaccaac tttggtgcca ccttcttgca gtacggtatc 1680

attaccccag atgtcagcag aaagatttcc tccgagcact tcccagcctt gtgtgccaag 1740

gttagaccaa acgttgttgg tttgactgat ggtttcaact tgactgacat gatgaccaat 1800

gctgctattg gtagatatga tggtaacgtc tacgaacact acttcgaaac tgtcaaggct 1860

ttgaacccac cagaaaacac caaggctcca tactccaagg ctttggaaga catgttgaac 1920

cgtccagacc ttgaagtcag agaaagaggt gaaaagtccg aagaagctgc tgaaatcttg 1980

tccagttaa 1989

<210> 386

<211> 804
<212> DNA
<213> Candida viswanathii

<400> 386
atggttagca caaaaacata cacagaaagg gcatcagcac acccctccaa ggttgcccaa 60

cgtttattcc gcttaatgga gtccaaaaag accaacctct gcgcctcgat cgacgtgacc 120

acaaccgccg agttcctttc gctcatcgac aagctcggtc cccacatctg tctcgtgaag 180

acgcacatcg atatcatctc agacttcagc tacgagggca cgattgagcc gttgcttgtg 240

cttgcagagc gccacgggtt cttgatattc gaggacagga agtttgctga tatcggaaac 300

accgtgatgt tgcagtacac ctcgggggta taccggatcg cggcgtggag tgacatcacg 360

aacgcgcacg gagtgactgg gaagggcgtc gttgaagggt tgaaacgcgg tgcggagggg 420

gtagaaaagg aaaggggcgt gttgatgttg gcggagttgt cgagtaaagg ctcgttggcg 480

catggtgaat atacccgtga gacgatcgag attgcgaaga gtgatcggga gttcgtgatt 540

gggttcatcg cgcagcggga catggggggt agagaagaag ggtttgattg gatcatcatg 600

acgcctggtg tggggttgga tgataaaggc gatgcgttgg gccagcagta taggactgtt 660

gatgaggtgg ttctgactgg taccgatgtg attattgtcg ggagagggtt gtttggaaaa 720

ggaagagacc ctgaggtgga gggaaagaga tacagggatg ctggatggaa ggcatacttg 780

aagagaactg gtcagttaga ataa 804

<210> 387
<211> 1125
<212> DNA
<213> Candida viswanathii

<400> 387
atgtccgtta aaaccaaaac catcactatt ctcccaggtg accacgtcgg taccgagatc 60

gttgccgagg caatcaagtc cttacatgcc attgagtccc tgaccccata ccaaaaagtt 120

cattttgagt tcaagcacca cttgataggc ggtgccgcca ttgacgccac tggtgtgccg 180

ctcccagacg acgccttagc cgccgccaag tcctcagacg ccgtgttgct cggggccgtc 240

ggtggaccca aatggggaac cggcgctgtg cgtccagaac agggattgtt gaagatccgt 300

aaggagttga acctttatgc caacatcaga ccatgcaatt tcgctagcga ctccttgctt 360

gagctctctc cgttacgtcc tgaggtcgtc aagggcacca acttgatcat tgtgcgtgag 420

ctagttggtg ggatttactt tggtgagaga caggagcagg aagaaagcga ggatgggaaa 480

tcggcttggg atactgaaaa gtacaccgtt gatgaagtcg cgagaattac acgcatggct 540

gcgttcatgg cattgcaaca caccccacca ttgccaatct ggtccttgga caaagccaac 600

gtgcttgcct cgtcaagatt gtggagaaag acagtggata aaatcatcag tgaagagttc 660

cctaccttgg ctgtccagca ccagttgatt gactccgcgg ctatgatctt gatccagaac 720

ccaacgaagt tgaacgggat catcatcacg tcgaacatgt ttggtgatat catatccgat 780

gaagcgtccg tgatcccagg gtccttggga ttgttgccat ctgcgtcatt ggcttcgtta 840

ccggacacaa acactgcgtt tggtctctat gagccatgcc atggttcggc acctgatttg 900

ccagagaaca aagtcaaccc tattgcgaca atcttgtccg ttgctagtat gttgagattg 960

agtttggatt gtgttaagga agccgaggct ttggaacagg ccgtgaagga agtcttggat 1020

aaggggatca gaacggcgga tttgagaggt agtagcacaa ccactgaagt cggtgatgcc 1080

gttgctgaaa ccgtctccag aatcttgaaa gaagctaagg cttga 1125

<210> 388
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 388
ttccgcttaa tggagtccaa a 21

<210> 389
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 389
gcatacaacg gaaatttgct tt 22

<210> 390
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 390
taaacgttgg gcaaccttgg 20

<210> 391
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 391
atggacgacg ttgattctgc tttagccgat aatgtt 36

<210> 392
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 392
ttatataact ttgtccaaga actttctggc c 31

<210> 393
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 393
gaatagaaga gagtgactct tttg 24

<210> 394
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 394
gattgattgt tatagtttct ttctttc 27

<210> 395
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 395
ctgcaggcat gatgatctg 19

<210> 396
<211> 25
<212> DNA
<213> Artificial Sequence

<220>

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 407 of 3077

<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 396
atcgattaaa ttctttaatt gaggg 25

<210> 397
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 397
gaatagaaga gagtgactct tttg 24

<210> 398
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 398
agatcatcat gcctgcagaa aatcagaggc tactcc 36

<210> 399
<211> 43
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 399
gtcactctct tctattctta tataactttg tccaagaact ttc 43

<210> 400
<211> 48
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 400
ccctcaatta aagaatttaa tcgatatgaa tgctaatgcc gttaagtc 48

<210> 401
<211> 47
<212> DNA
<213> Artificial Sequence

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 408 of 3077

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 401
tcaaaagagt cactctcttc tattcttacc caaacattgc aagctgt 47

<210> 402
<211> 54
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 402
ccctcaatta aagaatttaa tcgatatgac caaaacggtt gtaataggat cagg 54

<210> 403
<211> 47
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 403
tcaaaagagt cactctcttc tattcttacg cgtggccctc caacatt 47

<210> 404
<211> 60
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 404
ccctcaatta aagaatttaa tcgatatgag tgacaagcct ttgttaacac atgctacaga 60

<210> 405
<211> 47
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 405
tcaaaagagt cactctcttc tattcttaag cggcaggagg ggcggca 47

<210> 406
<211> 49

<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 406
ccctcaatta aagaatttaa tcgatatggc agaaatgttc aatggaaac 49

<210> 407
<211> 47
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 407
tcaaaagagt cactctcttc tattcttaag gtataatcgg ttcgacg 47

<210> 408
<211> 48
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 408
aagaaagaaa ctataacaat caatcatgaa tgctaatgcc gttaagtc 48

<210> 409
<211> 47
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 409
tcaaaagagt cactctcttc tattcttacc caaacattgc aagctgt 47

<210> 410
<211> 54
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 410
aagaaagaaa ctataacaat caatcatgac caaaacggtt gtaataggat cagg 54

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 410 of 3077

<210> 411
<211> 47
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 411
tcaaaagagt cactctcttc tattcttacg cgtggccctc caacatt 47

<210> 412
<211> 60
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 412
aagaaagaaa ctataacaat caatcatgag tgacaagcct ttgttaacac atgctacaga 60

<210> 413
<211> 47
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 413
tcaaaagagt cactctcttc tattcttaag cggcaggagg ggcggca 47

<210> 414
<211> 49
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 414
aagaaagaaa ctataacaat caatcatggc agaaatgttc aatggaaac 49

<210> 415
<211> 47
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 415
tcaaaagagt cactctcttc tattcttaag gtataatcgg ttcgacg 47

<210> 416
<211> 47
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 416
ccctcaatta aagaatttaa tcgatatggc aaccctatct gacttgg 47

<210> 417
<211> 51
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 417
tcaaaagagt cactctcttc tattcttagc tccacacatc catcaataat c 51

<210> 418
<211> 49
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 418
ccctcaatta aagaatttaa tcgatatgtt cattttggtg ttgttgacc 49

<210> 419
<211> 47
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 419
tcaaaagagt cactctcttc tattcttact ccactggctt tacctgc 47

<210> 420
<211> 47
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 420
ccctcaatta aagaatttaa tcgatatgag tgcccatgcc ttgccaa 47

<210> 421
<211> 54
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 421
tcaaaagagt cactctcttc tattcttaag cggtatcacc tttagttcta gtgc 54

<210> 422
<211> 50
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 422
ccctcaatta aagaatttaa tcgatatgac aaacttcttg atcgttgtgg 50

<210> 423
<211> 47
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 423
tcaaaagagt cactctcttc tattcttacg tcctctcttg ggcctca 47

<210> 424
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 424
gatccacaag ttttcactag aattgg 26

<210> 425
<211> 42
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 425
cactgcttaa ttaagcgctg tcactcttat atcagtatga cc 42

<210> 426
<211> 41
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 426
cactgcggat ccggtacccc accacgacaa caggattccc c 41

<210> 427
<211> 52
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 427
cgagatttaa ttaattagcc aacaggacaa tagagttaaa aagaatgaat tg 52

<210> 428
<211> 56
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 428
cgagatctgc aggcggccgc ttctgtgagt atttgttatt gaaaatgaag ggatgg 56

<210> 429
<211> 42
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 429
ccacgtcggt accgagatcg ttgccgaggc aatcaagtcc tt 42

<210> 430
<211> 50
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 430
aacggcttcg tctaaacaac cacggatctt caacaatccc tgttctggac 50

<210> 431
<211> 50
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 431
gtccagaaca gggattgttg aagatccgtg gttgtttaga cgaagccgtt 50

<210> 432
<211> 45
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 432
agtgtttgtg tccggtaacg accgaaatat tacaattgga gctcc 45

<210> 433
<211> 45
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 433
ggagctccaa ttgtaatatt tcggtcgtta ccggacacaa acact 45

<210> 434
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 434
tttcagcaac ggcatcacc 19

<210> 435
<211> 22
<212> DNA

<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 435
acctttatgc caacatcaga cc 22

<210> 436
<211> 47
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 436
aacggcttcg tctaaacaac ccatcaacgg tgtacttttc agtatcc 47

<210> 437
<211> 47
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 437
ggatactgaa aagtacaccg ttgatgggtt gtttagacga agccgtt 47

<210> 438
<211> 42
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 438
ttgcaatgcc atgaacgccc gaaatattac aattggagct cc 42

<210> 439
<211> 42
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 439
ggagctccaa ttgtaatatt tcgggcgttc atggcattgc aa 42

<210> 440

<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 440
cagatggcaa caatcccaag 20

<210> 441
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 441
aattaaccta tggtgcac 18

<210> 442
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 442
ttaattaaaa gcttggcgta atc 23

<210> 443
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 443
ctcgtgctag tcagtcttgc acgctttggg tg 32

<210> 444
<211> 44
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 444
atgattacgc caagctttta attaacaaca cggcgtctga ggac 44

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 417 of 3077

<210> 445
<211> 45
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 445
actgagagtg caccataggt taattaactc ggggccgtcg gtgga 45

<210> 446
<211> 37
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 446
aagcgtgcaa gactgactag cacgagcgaa gatgggg 37

<210> 447
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 447
ctcggggccg tcggtgga 18

<210> 448
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 448
caacacggcg tctgaggact tgg 23

<210> 449
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 449

gctcaacaat tgtctgacaa gatctc 26

<210> 450
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 450
gagattgtta actttatggg gcaataac 28

<210> 451
<211> 44
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 451
tccccatctt cgctcgtgct agtcaaaggg aagaagagtc gttg 44

<210> 452
<211> 44
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 452
cgtcggcacc caaagcgtgc aagacgtcga cctaaattcg caac 44

<210> 453
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 453
gtgctagtca agataaaccg aggcatggaa g 31

<210> 454
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 454
cgtgcaagac gcatagactg caagtaagca gc 32

<210> 455
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 455
cagtctatgc gtcttgcacg ctttgggtgc 30

<210> 456
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 456
cggtttatct tgactagcac gagcgaagat gg 32

<210> 457
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 457
aaaggaataa gctgggtgtt tgctcatttt g 31

<210> 458
<211> 43
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 458
cgttagtcat ttttgttaat tggtgttttg tttaataatg ggg 43

<210> 459
<211> 36
<212> DNA
<213> Artificial Sequence

<220>

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 420 of 3077

<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 459
aattaacaaa aatgactaac gcaccacaaa acattg 36

<210> 460
<211> 47
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 460
aacacccagc ttattccttt atagcataac ttcttctaaa cttgtcg 47

<210> 461
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 461
aaagtcatag gctgggtgtt tgctcatttt g 31

<210> 462
<211> 43
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 462
aacttagcat ttttgttaat tggtgttttg tttaataatg ggg 43

<210> 463
<211> 34
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 463
aattaacaaa aatgctaagt ttcatcaccg aagc 34

<210> 464
<211> 40
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 464
aacacccagc ctatgacttt atacatgtaa cggatccgtc 40

<210> 465
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 465
gttctggtga gctgggtgtt tgctcatttt g 31

<210> 466
<211> 43
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 466
agactgacat ttttgttaat tggtgttttg tttaataatg ggg 43

<210> 467
<211> 33
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 467
aattaacaaa aatgtcagtc tcgcctttcg ttg 33

<210> 468
<211> 41
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 468
aacacccagc tcaccagaac ttccattctt taacgtttat c 41

<210> 469
<211> 31

<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 469
cggtttatag gctgggtgtt tgctcatttt g 31

<210> 470
<211> 40
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 470
aacacccagc ctataaaccg atatagtcct cccagttttc 40

<210> 471
<211> 43
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 471
ctgaatacat ttttgttaat tggtgttttg tttaataatg ggg 43

<210> 472
<211> 34
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 472
aattaacaaa aatgtattca gcttcagtca cagc 34

<210> 473
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 473
gattacttta agctgggtgt ttgctcattt tg 32

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 423 of 3077

<210> 474
<211> 43
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 474
ctgatgacat ttttgttaat tggtgttttg tttaataatg ggg 43

<210> 475
<211> 39
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 475
aattaacaaa aatgtcatca gaatacgcta aactagtgg 39

<210> 476
<211> 49
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 476
aacacccagc ttaaagtaat ctatgtatct cagtatcttg gtacttttc 49

<210> 477
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 477
atccaaatag gctgggtgtt tgctcatttt g 31

<210> 478
<211> 43
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 478
tatcagacat ttttgttaat tggtgttttg tttaataatg ggg 43

<210> 479
<211> 34
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 479
aattaacaaa aatgtctgat aaagcagccg ctag 34

<210> 480
<211> 37
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 480
aacacccagc ctatttggat ctcttgtaaa ccttgcc 37

<210> 481
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 481
taccaaacta agctgggtgt ttgctcattt tg 32

<210> 482
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 482
aacacccagc ttagtttggt aacccacttt cgttg 35

<210> 483
<211> 44
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 425 of 3077

<400> 483
cttttgcaca tttttgttaa ttggtgtttt gtttaataat gggg 44

<210> 484
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 484
ccaattaaca aaaatgtcaa aagcttttag tgcccc 36

<210> 485
<211> 751
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 485
aagggaagaa gagtcgttga gttgatgtaa ttaagctggc acgtagatac cagaaggttc 60

tagagtagag cttgggtggt gtttggccct gtttggacca cggatagaga tggagaatcc 120

cttggttaga gcggagagga aaaaattgaa actttgcata tcccacttca ttatccttga 180

tgtaaccgtt ttatggggta attaaagtgt ggaaaaataa tcagggagac atattcccga 240

tcaattgggt ggtggtcgct caatttctgt gagtagtagg ctcagtggtg tgtattggga 300

ttggtagtag tctgtataag cagtgttata taacccattg cttgttgatt cctattttgc 360

tggcaaaagt gacaactgta gttgtgagat aatcctcggt tattacgcct gggggggcag 420

acagccaaag ttgtgcccgt gcgacaatgg catcagaaga aacagaaaaa aaaaacacag 480

gcatttttat ccacatgcac actaccccca ctattcctgt ctgcagtgtg cttgtgtgtg 540

gccccccgca gaatcaacag ggcaaactct ggagcctgaa tctttatata aacttcaggc 600

attggccccc cttttcacaa ttcttcacat ccaccatttt ttttcttctt tcctaccata 660

ttagtttttt tttattcttt tcctacctat ctgattatta tcaaacatct ggtcatcctc 720

aaaagaaaga aagaaactat aacaatcaat c 751

<210> 486
<211> 531
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 426 of 3077

<400> 486
gagctccaat tgtaatattt cgggagaaat atcgttgggg taaaacaaca gagagagaga 60

gggagagatg gttctggtag aattataatc tggttgttgc aaatgctact gatcgactct 120

ggcaatgtct gtagctcgct agttgtatgc aacttaggtg ttatgcatac acacggttat 180

tcggttgaat tgtggagtaa aaattgtctg agttgtgtct tagctactgg ctggcccccc 240

gcgaaagata atcaaaatta cacttgtgaa tttttgcaca cacaccgatt aacatttccc 300

ttttttgtcc accgatacac gcttgcctct tctttttttt ctctgtgctt ccccctcctg 360

tgacttttttc caccattgat ataaaatcaa ctccatttcc ctaaaatctc cccagattct 420

aaaaacaact tcttctcttc tgcttttcct tttttttgt tatatttatt taccatccct 480

tttttttgaa tagttattcc ccactaacat tgttcaaatc ttcacgacat a 531

<210> 487
<211> 415
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 487
cgacgggtac aacgagaatt gtattgaatt gatcaagaac atgatcttgg tgttacagaa 60

catcaagttc ttggaccaga ctgagaatgc acagatatac aaggcgtcat gtgataaaat 120

ggatgagatt tatccacaat tgaagaaaga gtttatggaa agtggtcaac cagaagctaa 180

acaggaagaa gcaaacgaag aggtgaaaca agaagaagaa ggtaaataag tattttgtat 240

tatataacaa acaaagtaag gaatacagat ttatacaata aattgccata ctagtcacgt 300

gagatatctc atccattccc caactcccaa gaaaaaaaaa aagtgaaaaa aaaaatcaaa 360

cccaaagatc aacctcccca tcatcatcgt catcaaaccc ccagctcaat tcgca 415

<210> 488
<211> 745
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 488
aaaatcagag gctactccgg agatggcaca ttatcacgtg ggcagtccca tctctctgac 60

aatggcactg acaatgccat gtcattctag aagtcgctcg ttccattgct agtcgtgtag 120

gtgtcttact cagaccatat gtttggctgt gtgtgtggta agggggtaag atatctctag 180

gaggaggcaa ctctactcga aagtacgaaa tggccgacgc caatccgacc tgccattgtg 240

caggacgcgg gtcgagatcc ggaagtgccc tccttctgag gttgttggtt tgttccagtt 300

ggcataagaa attaataatc gtttagtcaa tggtacagta ttaataattc ctgacaaaac 360

cgtgccctac aagcagtgtg agttccgggg gtagcactag tggtcgtggt ggtagggccc 420

acagagagta tttccggaac cgactgatcc tttgattcca atattccaag atttgggttt 480

gtggttgtgt gtaatgcatg aggttatttt gtcaacctcc ttccttctct acggttgttg 540

ctcctcctcc tcttccttcc ctttgttcat ttaaatacac acacaccact ccccttccca 600

cccacaagaa ctttttttt ttttactttc ttcttctctt actttcttct tcttctttca 660

ttacctgtta gattatattt atcaacctca attggtttat tatctaccac ccatacaacc 720

cccacccctg ccacatccac caaca 745

<210> 489
<211> 614
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 489
cggaagttgt ttaccgacct gaccgtaaat ttgctgctga aagaaacgtg tcaaacaaga 60

ccaattggct caattgaccc tgtggaaatg ctttgttgac caccaatgct tccaccaaac 120

gttacttttt ttttgcaatc ggatggtatg ggtctggggt tcacctgttt tgtaaagcta 180

cagaaggtgg catatttctc tgatcaggtg ttttttttt cggctgctgc tgctcgtggt 240

ggtgtagtgg tagtggtgtg tgtgtgtgtg tgtgcgtgcg tgtggaagga cgcttttgc 300

tctctgactc ctcccaatca gaagttgcta tagtggtgaa acaacaatgg atgataatgc 360

cccgggcggt gcgtgtccga cacaaaccac tacatttttt agctgggagc ctactgccac 420

tacgacccac ccacccatgg tcaacaaaaa aattctgaca aattataaaa taacccttga 480

attccccctt ggaaaaattt ttggtatttc tctctctctt ttcctttccc tcttcttttt 540

ctctccatca atcaattgac gttcagtaac tcaattaatt acatcacatc cctcaattaa 600

agaatttaat cgat 614

<210> 490
<211> 360
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 490
agataaaccg aggcatggaa gaagtcagag acttcctaat tgggaaatat cacatgaaaa 60

tgtatggatg tgtgcaggac acgtctactg agtgccatac tttgtgccca ccgcgaaatt 120

aagggcagcc atcagattac gtaagcagtg gtatgcagtt gttacagcgt caagtttttc 180

tgtattaccc caaagttttc ttctctcccc tgcaaaagtt ttcatatgca aaactgatcc 240

cacagcagta taaataccca ataactttcc ccaaccaaca gcaacaacct ctcattcttt 300

tttttcctat tcttttcttt aatcaacccc cattattaaa caaaacacca attaacaaaa 360

<210> 491
<211> 500
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 491
gtcttgcacg ctttgggtgc cgacgacgcc ttggcccact caagtattag attcggtatt 60

ggtagattca ccactgaggc agaggtcgac tatgtcatca aggctattaa cgaaagagtt 120

gatttcttga gaaagatgtc tccattgtgg gagatggtgc aagaaggcat tgacttggac 180

tcaattgaat ggagtggtca ttagtgtgac cccctatcgc tacggtcttc ctcttcgttt 240

catttaagta tgacttgttt tgtttattta tctcgtgtat agaatagtaa ttatttttgg 300

tataagccat gttggtattc gtgagtcaga tgtgagctgt atgttctcct gaaccggcgt 360

aacttccctt ccgttgccgc gggaatgcga caacaacgac gggaaaaaaa tcccagtata 420

aaaccacccc aaatgacgtg tgtgaaattt ttcactccag gtttctctcc ctctttttg 480

tccccccaaa ccaatcacca 500

<210> 492
<211> 184
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 492
gaatagaaga gagtgactct tttgataaga gtcgcaaatt tgatttcata agtatatatt 60

cattatgtaa agtagtaaat ggaaaattca ttaaaaaaaa agcaaatttc cgttgtatgc 120

atactccgaa cacaaaacta gccccggaaa aacccttagt tgatagttgc gaatttaggt 180

cgac 184

<210> 493
<211> 267
<212> DNA

<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 493
taaatattgt aataaatagg tctatataca tacactaagc ttctaggacg tcattgtagt 60

cttcgaagtt gtctgctagt ttagttctca tgatttcgaa aaccaataac gcaatggatg 120

tagcagggat ggtggttagt gcgttcctga caaacccaga gtacgccgcc tcaaaccacg 180

tcacattcgc cctttgcttc atccgcatca cttgcttgaa ggtatccacg tacgagttgt 240

aatacacctt gaagaacggc ttcgtct 267

<210> 494
<211> 374
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 494
gctgggtgtt tgctcatttt gtttcaatgc atggatgcac actggtcttt ttatgatttg 60

aaatgaatag atttgatcat catcattttt tttttgcaaa tatttatcat ataatgtgag 120

ttttctttgg tgtatatttg ttttctacag tatttagata atattatttg aaaagaatat 180

atgcttaaaa gaaactacaa gttagagttc gacgatgaag tggttgtcat gtcggcgact 240

ttaccccaga gtaagaagga attgtcgtgg tcgatattgg ttgaaacata tcgggactag 300

tcttgcgacc gagcgcaggg caattgtgga gtcacacata caatgaaaat tggctgctta 360

cttgcagtct atgc 374

<210> 495
<211> 300
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 495
ggatcctcta catatgaaca ttctatattg aaaaatgcat gtatattatt gtcgttaacg 60

ttttttaaaa attcatggtc ttttctaaaa atgcagatag ttgagcaaca gaatcctctt 120

tatcaatcac cacaaccacc gactgtttgc tactctcatt cgtaacaatg aatctattcg 180

acgtgaccgt agcctcacta gactcagatt ggctggtgtt cctcgcaaga ttgggttctt 240

cattaggcga gtaaaactcc tcttcctcgt cctcctcccc atcttcgctc gtgctagtca 300

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 430 of 3077

<210> 496
<211> 50
<212> PRT
<213> Candida viswanathii

<400> 496
Met Phe Asn Phe Lys Leu Ser Gln Gln Val Leu Lys Asn Ser Thr Lys
1 5 10 15

Ser Ile Met Pro Ile Leu Lys Lys Pro Phe Ser Thr Ser His Ala Lys
20 25 30

Gly Asp Leu Phe Lys Tyr Gln Ser Gln Leu Pro Lys Leu Pro Val Pro
35 40 45

Thr Leu
50

<210> 497
<211> 32
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 497
Met Pro Ile Leu Lys Lys Pro Phe Ser Thr Ser His Ala Lys Gly Asp
1 5 10 15

Leu Phe Lys Tyr Gln Ser Gln Leu Pro Lys Leu Pro Val Pro Thr Leu
20 25 30

<210> 498
<211> 32
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 498
Met Pro Ile Leu Lys Lys Pro Phe Ser Thr Ser His Ala Lys Gly Asp
1 5 10 15

Leu Phe Lys Tyr Gln Ser Gln Leu Pro Lys Leu Pro Val Pro Thr Leu
20 25 30

<210> 499
<211> 50
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
polypeptide

<400> 499
Met Phe Asn Phe Lys Leu Ser Gln Gln Val Leu Lys Asn Ser Thr Lys
1 5 10 15

Ser Ile Met Pro Ile Leu Lys Lys Pro Phe Ser Thr Ser His Ala Lys
20 25 30

Gly Asp Leu Phe Lys Tyr Gln Ser Gln Leu Pro Lys Leu Pro Val Pro
35 40 45

Thr Leu
50

<210> 500
<211> 27
<212> PRT
<213> Candida viswanathii

<400> 500
Lys Trp Tyr Leu Val Asp Ser Ala Asn Glu Met Lys Asp Val Leu Thr
1 5 10 15

Lys Gly Leu Leu Thr Asp Ala Lys Pro Lys Leu
20 25

<210> 501
<211> 27
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
peptide

<400> 501
Lys Trp Tyr Leu Val Asp Ser Ala Asn Glu Met Lys Asp Val Leu Thr
1 5 10 15

Lys Gly Leu Leu Thr Asp Ala Lys Pro Lys Leu
20 25

<210> 502
<211> 24
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
peptide

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 432 of 3077

<400> 502
Lys Trp Tyr Leu Val Asp Ser Ala Asn Glu Met Lys Asp Val Leu Thr
1   5   10   15

Lys Gly Leu Leu Thr Asp Ala Lys
            20

<210> 503
<211> 24
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
      peptide

<400> 503
Lys Trp Tyr Leu Val Asp Ser Ala Asn Glu Met Lys Asp Val Leu Thr
1   5   10   15

Lys Gly Leu Leu Thr Asp Ala Lys
            20

<210> 504
<211> 17
<212> PRT
<213> Candida viswanathii

<400> 504
Met Ser Thr Tyr Gln Phe Gln Glu Thr Leu Glu Lys Leu Pro Ile Pro
1   5   10   15

Asp

<210> 505
<211> 49
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
      polypeptide

<400> 505
Met Phe Asn Phe Lys Leu Ser Gln Gln Val Leu Lys Asn Ser Thr Lys
1   5   10   15

Ser Ile Met Pro Ile Leu Lys Lys Pro Phe Ser Thr Ser His Ala Lys
            20   25   30

Gly Ser Thr Tyr Gln Phe Gln Glu Thr Leu Glu Lys Leu Pro Ile Pro
        35   40   45

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 433 of 3077

Asp

<210> 506
<211> 41
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 506
Met Ser Ala Leu Arg Ser Phe Gln Arg Ser Ser Asn Val Ala Lys Ser
1 5 10 15

Thr Leu Lys Asn Ser Val Arg Thr Tyr Ser Thr Tyr Gln Phe Gln Glu
20 25 30

Thr Leu Glu Lys Leu Pro Ile Pro Asp
35 40

<210> 507
<211> 46
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 507
Met Leu Ser Arg Thr Thr Leu Arg Val Ala Arg Gln Gln Thr Arg Leu
1 5 10 15

Leu Ser Thr Ser Arg Ile Leu Phe Asn Ser Lys Thr Asp Gln Ser Thr
20 25 30

Tyr Gln Phe Gln Glu Thr Leu Glu Lys Leu Pro Ile Pro Asp
35 40 45

<210> 508
<211> 636
<212> PRT
<213> Candida viswanathii

<400> 508
Met Pro Thr Glu Leu Gln Lys Glu Arg Glu Leu Thr Lys Phe Asn Pro
1 5 10 15

Lys Glu Leu Asn Tyr Phe Leu Glu Gly Ser Gln Glu Arg Ser Glu Ile
20 25 30

Ile Ser Asn Met Val Glu Gln Met Gln Lys Asp Pro Ile Leu Lys Val

35 40 45

Asp Ala Ser Tyr Tyr Asn Leu Thr Lys Asp Gln Gln Arg Glu Val Thr
50 55 60

Ala Lys Lys Ile Ala Arg Leu Ser Arg Tyr Phe Glu His Glu Tyr Pro
65 70 75 80

Asp Gln Gln Ala Gln Arg Leu Ser Ile Leu Gly Val Phe Asp Pro Gln
85 90 95

Val Phe Thr Arg Ile Gly Val Asn Leu Gly Leu Phe Val Ser Cys Val
100 105 110

Arg Gly Asn Gly Thr Asn Ser Gln Phe Phe Tyr Trp Thr Ile Asn Lys
115 120 125

Gly Ile Asp Lys Leu Arg Gly Ile Tyr Gly Cys Phe Gly Met Thr Glu
130 135 140

Leu Ala His Gly Ser Asn Val Gln Gly Ile Glu Thr Thr Ala Thr Phe
145 150 155 160

Asp Glu Asp Thr Asp Glu Phe Val Ile Asn Thr Pro His Ile Gly Ala
165 170 175

Thr Lys Trp Trp Ile Gly Gly Ala Ala His Ser Ala Thr His Cys Ser
180 185 190

Val Tyr Ala Arg Leu Lys Val Lys Gly Lys Asp Tyr Gly Val Lys Thr
195 200 205

Phe Val Val Pro Leu Arg Asp Ser Asn His Asp Leu Glu Pro Gly Val
210 215 220

Thr Val Gly Asp Ile Gly Ala Lys Met Gly Arg Asp Gly Ile Asp Asn
225 230 235 240

Gly Trp Ile Gln Phe Ser Asn Val Arg Ile Pro Arg Phe Phe Met Leu
245 250 255

Gln Lys Tyr Cys Lys Val Ser Arg Ser Gly Glu Val Thr Met Pro Pro
260 265 270

Ser Glu Gln Leu Ser Tyr Ser Ala Leu Ile Gly Gly Arg Val Thr Met
275 280 285

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 435 of 3077

```
Met Met Asp Ser Tyr Arg Met Thr Ser Arg Phe Ile Thr Ile Ala Leu
    290                 295                 300

Arg Tyr Ala Ile His Arg Arg Gln Phe Lys Lys Lys Asp Thr Asp Thr
305                 310                 315                 320

Ile Glu Thr Lys Leu Ile Asp Tyr Pro Leu His Gln Lys Arg Leu Phe
                325                 330                 335

Pro Phe Leu Ala Ala Ala Tyr Leu Phe Ser Gln Gly Ala Leu Tyr Leu
            340                 345                 350

Glu Gln Thr Met Asn Ala Thr Asn Asp Lys Leu Asp Glu Ala Val Ser
        355                 360                 365

Ala Gly Glu Lys Glu Ala Ile Asp Ala Ala Ile Val Glu Ser Lys Lys
    370                 375                 380

Leu Phe Val Ala Ser Gly Cys Leu Lys Ser Thr Cys Thr Trp Leu Thr
385                 390                 395                 400

Ala Glu Ala Ile Asp Glu Ala Arg Gln Ala Cys Gly Gly His Gly Tyr
                405                 410                 415

Ser Ser Tyr Asn Gly Phe Gly Lys Ala Tyr Ser Asp Trp Val Val Gln
            420                 425                 430

Cys Thr Trp Glu Gly Asp Asn Asn Ile Leu Ala Met Asn Val Ala Lys
        435                 440                 445

Pro Met Val Arg Asp Leu Leu Lys Glu Pro Glu Gln Lys Gly Leu Val
    450                 455                 460

Leu Ser Ser Val Ala Asp Leu Asp Asp Pro Ala Lys Leu Val Lys Ala
465                 470                 475                 480

Phe Asp His Ala Leu Ser Gly Leu Ala Arg Asp Ile Gly Ala Val Ala
                485                 490                 495

Glu Asp Lys Gly Phe Asp Ile Thr Gly Pro Ser Leu Val Leu Val Ser
            500                 505                 510

Lys Leu Asn Ala His Arg Phe Leu Ile Asp Gly Phe Phe Lys Arg Ile
        515                 520                 525

Thr Pro Glu Trp Ser Glu Val Leu Arg Pro Leu Gly Phe Leu Tyr Ala
    530                 535                 540
```


LCY Biotechnology Holding, Inc.
Ex. 1002
Page 436 of 3077

```
Asp Trp Ile Leu Thr Asn Phe Gly Ala Thr Phe Leu Gln Tyr Gly Ile
545                 550                 555                 560


Ile Thr Pro Asp Val Ser Arg Lys Ile Ser Ser Glu His Phe Pro Ala
                565                 570                 575


Leu Cys Ala Lys Val Arg Pro Asn Val Val Gly Leu Thr Asp Gly Phe
            580                 585                 590


Asn Leu Thr Asp Met Met Thr Asn Ala Ala Ile Gly Arg Tyr Asp Gly
        595                 600                 605


Asn Val Tyr Glu His Tyr Phe Glu Thr Val Lys Ala Leu Asn Pro Pro
    610                 615                 620


Glu Asn Thr Lys Ala Pro Tyr Ser Lys Ala Leu Glu
625                 630                 635


<210> 509
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
      oligonucleotide


<220>
<221> modified_base
<222> (4)..(6)
<223> a, c, t, g, unknown or other

<220>
<221> modified_base
<222> (8)..(8)
<223> a, c, t, g, unknown or other

<220>
<221> modified_base
<222> (10)..(21)
<223> a, c, t, g, unknown or other

<220>
<221> misc_feature
<222> (10)..(21)
<223> This region may encompass 9-12 nucleotides

<400> 509
cggnnntnan nnnnnnnnnn nccg                                          24


<210> 510
<211> 24
<212> DNA
<213> Artificial Sequence
```


LCY Biotechnology Holding, Inc.
Ex. 1002
Page 437 of 3077

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<220>
<221> modified_base
<222> (4)..(6)
<223> a, c, t, g, unknown or other

<220>
<221> modified_base
<222> (8)..(8)
<223> a, c, t, g, unknown or other

<220>
<221> modified_base
<222> (10)..(21)
<223> a, c, t, g, unknown or other

<220>
<221> misc_feature
<222> (10)..(21)
<223> This region may encompass 8-12 nucleotides

<400> 510
cggnnntnrn nnnnnnnnn nccg 24

<210> 511
<211> 48
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<220>
<221> modified_base
<222> (7)..(42)
<223> a, c, t, g, unknown or other

<220>
<221> misc_feature
<222> (7)..(42)
<223> This region may encompass 4-36 nucleotides

<400> 511
cyccrdnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnhyggrg 48

<210> 512
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<220>
<221> modified_base
<222> (4)..(6)
<223> a, c, t, g, unknown or other

<220>
<221> modified_base
<222> (10)..(21)
<223> a, c, t, g, unknown or other

<400> 512
cggnnnttan nnnnnnnnnn ncag 24

<210> 513
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 513
cggttattac gcctgggggg gcag 24

<210> 514
<211> 13
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<220>
<221> modified_base
<222> (7)..(7)
<223> a, c, t, g, unknown or other

<400> 514
cttgtgncat gtg 13

<210> 515
<211> 9
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
peptide

<400> 515
Asp Tyr Lys Asp Asp Asp Asp Lys Gly
1 5

<210> 516
<211> 14

<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic peptide

<400> 516
Gly Lys Pro Ile Pro Asn Pro Leu Leu Gly Leu Asp Ser Thr
1 5 10

<210> 517
<211> 10
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic peptide

<400> 517
Glu Gln Lys Leu Ile Ser Glu Glu Asp Leu
1 5 10

<210> 518
<211> 11
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic peptide

<400> 518
Gln Pro Glu Leu Ala Pro Glu Asp Pro Glu Asp
1 5 10

<210> 519
<211> 9
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic peptide

<400> 519
Tyr Pro Tyr Asp Val Pro Asp Tyr Ala
1 5

<210> 520
<211> 11
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic peptide

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 440 of 3077

<400> 520
Tyr Thr Asp Ile Glu Met Asn Arg Leu Gly Lys
1 5 10

<210> 521
<211> 6
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic 6xHis tag

<400> 521
His His His His His His
1 5

<210> 522
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic peptide

<220>
<221> MOD_RES
<222> (3)..(5)
<223> Any amino acid

<220>
<221> MISC_FEATURE
<222> (3)..(5)
<223> This region may encompass 1-3 residues

<400> 522
Cys Cys Xaa Xaa Xaa Cys Cys
1 5

<210> 523
<211> 6
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic peptide

<400> 523
Cys Cys Pro Gly Cys Cys
1 5

<210> 524
<211> 6
<212> PRT
<213> Unknown

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 441 of 3077

<220>
<223> Description of Unknown:
Thrombin recognition site sequence

<400> 524
Leu Val Pro Arg Gly Ser
1 5

<210> 525
<211> 5
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic peptide

<400> 525
Asp Asp Asp Asp Lys
1 5

<210> 526
<211> 7
<212> PRT
<213> Tobacco etch virus

<400> 526
Glu Asn Leu Tyr Phe Gln Gly
1 5

<210> 527
<211> 8
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic peptide

<400> 527
Leu Glu Val Leu Phe Gln Gly Pro
1 5

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 442 of 3077

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application	:	Aeling K, et al.
International Application No.	:	PCT/US2018/041579
International Filing Date	:	11 July 2018
For	:	**BIOLOGICAL METHODS FOR PREPARING TERPENES**

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

DECLARATION OF SEQUENCE LISTING

Sir:

As required under 37 C.F.R. §§ 1.821 and 1.825, Applicants' undersigned attorney hereby states that the information recorded in the enclosed computer readable form (CRF) is identical to the enclosed paper or PDF image copy of the Sequence Listing.

Additionally, as required under 37 C.F.R. § 1.825, Applicants' undersigned attorney hereby states, upon information and belief, that the enclosed Sequence Listing does not include any new matter, and is believed to be fully supported by the above-identified patent application as filed.

Respectfully submitted

SILVIA SALVADORI, P.C.

January 13, 2020 By: /Silvia Salvadori/

Silvia Salvadori, Reg. No. 48,265
270 Madison Avenue, 8th Floor
New York, NY 10016
Phone: (212) 897-1938
Fax: (646) 219-6529

Doc code: IDS
Doc description: Information Disclosure Statement (IDS) Filed

PTO/SB/08a (02-18)
Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (**Not for submission under 37 CFR 1.99**)		
	Application Number	
	Filing Date	
	First Named Inventor	Kimberly Ann Aeling
	Art Unit	
	Examiner Name	
	Attorney Docket Number	SS118-1026US

U.S.PATENTS						Remove
Examiner Initial*	Cite No	Patent Number	Kind Code[1]	Issue Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1					
If you wish to add additional U.S. Patent citation information please click the Add button.						Add

U.S.PATENT APPLICATION PUBLICATIONS						Remove
Examiner Initial*	Cite No	Publication Number	Kind Code[1]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1	20030044946	A1	2003-03-06	Longo	Entire Document
	2	20130066035	A1	2013-03-14	Burgard et al.	Entire Document
	3	20130302861	A1	2013-11-14	Vainstein et al.	Entire Document
	4	20140113343	A1	2014-04-24	Daviet et al.	Entire Document
	5	20150010978	A1	2015-01-08	Heaps et al.	Entire Document
If you wish to add additional U.S. Published Application citation information please click the Add button.						Add
FOREIGN PATENT DOCUMENTS						Remove

EFS Web 2.1.18

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	
	Filing Date	
	First Named Inventor	Kimberly Ann Aeling
	Art Unit	
	Examiner Name	
	Attorney Docket Number	SS118-1026US

Examiner Initial*	Cite No	Foreign Document Number[3]	Country Code[2] i	Kind Code[4]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear	T[5]
	1	2012254044	JP	A	2012-12-27	Tamagawa et al.	Entire Document	
	2	2014100504	WO	A2	2014-06-26	Beardslee et al.	Entire Document	
	3	2018094110	WO	A2	2018-05-24	Naughton et al.	Entire Document	

If you wish to add additional Foreign Patent Document citation information please click the Add button Add

NON-PATENT LITERATURE DOCUMENTS Remove

Examiner Initials*	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc), date, pages(s), volume-issue number(s), publisher, city and/or country where published.	T[5]
	1	International Preliminary Report on Patentability of PCT/US2018/041579 of 2 December 2019.	
	2	Search Report and Written Opinion of PCT/US18/41579 of 13 September 2018.	
	3	Written Opinion of the International Preliminary Examining Authority of PCT/US2018/041579 of 13 September 2019.	

If you wish to add additional non-patent literature document citation information please click the Add button Add

EXAMINER SIGNATURE

Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through a citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

EFS Web 2.1.18

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	
	Filing Date	
	First Named Inventor	Kimberly Ann Aeling
	Art Unit	
	Examiner Name	
	Attorney Docket Number	SS118-1026US

[1] See Kind Codes of USPTO Patent Documents at www.USPTO.GOV or MPEP 901.04. [2] Enter office that issued the document, by the two-letter code (WIPO Standard ST.3). [3] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [4] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [5] Applicant is to place a check mark here if English language translation is attached.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	
	Filing Date	
	First Named Inventor	Kimberly Ann Aeling
	Art Unit	
	Examiner Name	
	Attorney Docket Number	SS118-1026US

CERTIFICATION STATEMENT

Please see 37 CFR 1.97 and 1.98 to make the appropriate selection(s):

That each item of information contained in the information disclosure statement was first cited in any communication from a foreign patent office in a counterpart foreign application not more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(1).

OR

☐ That no item of information contained in the information disclosure statement was cited in a communication from a foreign patent office in a counterpart foreign application, and, to the knowledge of the person signing the certification after making reasonable inquiry, no item of information contained in the information disclosure statement was known to any individual designated in 37 CFR 1.56(c) more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(2).

See attached certification statement.

The fee set forth in 37 CFR 1.17 (p) has been submitted herewith.

☒ A certification statement is not submitted herewith.

SIGNATURE

A signature of the applicant or representative is required in accordance with CFR 1.33, 10.18. Please see CFR 1.4(d) for the form of the signature.

Signature	/Silvia Salvadori/	Date (YYYY-MM-DD)	2020-01-13
Name/Print	Silvia Salvadori	Registration Number	48265

This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 1 hour to complete, including gathering, preparing and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

Privacy Act Statement

The Privacy Act of 1974 (P.L. 93-579) requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C. 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether the Freedom of Information Act requires disclosure of these record s.

2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.

3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.

4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).

5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.

6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).

7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (i.e., GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.

8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspections or an issued patent.

9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 448 of 3077

(19)日本国特許庁(JP)　(12)公開特許公報(A)　(11)特許出願公開番号
特開2012-254044
(P2012-254044A)
(43)公開日　平成24年12月27日(2012.12.27)

(51)Int.Cl.			F I			テーマコード（参考）
C12N	***1/19***	***(2006.01)***	C12N	1/19	ZNA	4B024
C12P	***7/04***	***(2006.01)***	C12P	7/04		4B050
C12N	*15/09*	*(2006.01)*	C12N	15/00	A	4B064
C12N	*9/10*	*(2006.01)*	C12N	9/10		4B065

審査請求 未請求　請求項の数 15　ＯＬ　（全 27 頁）

(21)出願番号　特願2011-129527 (P2011-129527)
(22)出願日　平成23年6月9日 (2011.6.9)

（出願人による申告）平成２２年度、独立行政法人新エネルギー・産業技術総合開発機構・グリーン・サステイナブルケミカルプロセス基盤技術開発委託研究、産業技術力強化法第１９条の適用を受ける特許出願

(71)出願人　000253503
キリンホールディングス株式会社
東京都中央区新川二丁目１０番１号
(74)代理人　100117787
弁理士　勝沼　宏仁
(74)代理人　100091487
弁理士　中村　行孝
(74)代理人　100107342
弁理士　横田　修孝
(74)代理人　100111730
弁理士　伊藤　武泰
(72)発明者　玉　川　英　幸
東京都中央区新川二丁目１０番１号　キリンホールディングス株式会社内

最終頁に続く

(54)【発明の名称】酵母におけるアセチルＣｏＡを経由する代謝経路を利用した物質の製造法

(57)【要約】
【課題】アセチルＣｏＡを経由する代謝産物を高生産し得る酵母菌株の提供。
【解決手段】以下の酵素：（ａ）アセチルＣｏＡ合成酵素（ＡＣＳ）；および（ｂ）アセチルＣｏＡアセチル基転移酵素（ＥＲＧ１０）の活性が増強されている、酵母菌株。
【選択図】なし

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 449 of 3077

【特許請求の範囲】
【請求項１】
以下の酵素：
（ａ）アセチルＣｏＡ合成酵素（ＡＣＳ）；および
（ｂ）アセチルＣｏＡアセチル基転移酵素（ＥＲＧ１０）
の活性が増強されている、酵母菌株。
【請求項２】
以下の遺伝子：
（ａ）アセチルＣｏＡ合成酵素（ＡＣＳ）の活性を有するポリペプチドをコードする遺伝子；および
（ｂ）アセチルＣｏＡアセチル基転移酵素（ＥＲＧ１０）の活性を有するポリペプチドをコードする遺伝子
の発現が増強されている、請求項１に記載の酵母菌株。
【請求項３】
ＡＣＳ活性を有するポリペプチドが、
（ｉ）配列番号１２または配列番号１４で表されるアミノ酸配列を含むポリペプチド、または
（ｉｉ）配列番号１２または配列番号１４で表されるアミノ酸配列において、１もしくは数個のアミノ酸が欠失、置換、付加もしくは挿入されたアミノ酸配列を含み、かつＡＣＳ活性を有するポリペプチド
である、請求項２に記載の酵母菌株。
【請求項４】
ＥＲＧ１０活性を有するポリペプチドが、
（ｉ）配列番号６２で表されるアミノ酸配列を含むポリペプチド、または
（ｉｉ）配列番号６２で表されるアミノ酸配列において、１もしくは数個のアミノ酸が欠失、置換、付加もしくは挿入されたアミノ酸配列を含み、かつＥＲＧ１０活性を有するポリペプチド
である、請求項２または３に記載の酵母菌株。
【請求項５】
さらに、以下の遺伝子：
（ｃ）ＡＴＰ－クエン酸除去付加酵素（ＡＣＬ）の活性を有するポリペプチドをコードする遺伝子；
（ｄ）ミトコンドリア内膜カルニチン輸送体（ＣＲＣ１）の活性を有するポリペプチドをコードする遺伝子；および
（ｅ）カルニチンアセチル転移酵素遺伝子（ＹＡＴ１）の活性を有するポリペプチドをコードする遺伝子
からなる群から選択される少なくとも一種の遺伝子の発現が増強されている、請求項１～４のいずれか一項に記載の酵母菌株。
【請求項６】
ＡＣＬ活性を有するポリペプチドが、ＡＣＬ１活性を有するポリペプチドおよびＡＣＬＹ活性を有するポリペプチドからなるものであり、
ＡＣＬ１活性を有するポリペプチドが、
（ｉ）配列番号３２で表されるアミノ酸配列を含むポリペプチド、または
（ｉｉ）配列番号３２で表されるアミノ酸配列において、１もしくは数個のアミノ酸が欠失、置換、付加もしくは挿入されたアミノ酸配列を含み、かつＡＣＬ１活性を有するポリペプチド
であり、
ＡＣＬＹ活性を有するポリペプチドが、
（ｉ）配列番号３４で表されるアミノ酸配列を含むポリペプチド、または
（ｉｉ）配列番号３４で表されるアミノ酸配列において、１もしくは数個のアミノ酸が欠失

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 450 of 3077

、置換、付加もしくは挿入されたアミノ酸配列を含み、かつＡＣＬＹ活性を有するポリペプチド
である、請求項５に記載の酵母菌株。
【請求項７】
ＣＲＣ１活性を有するポリペプチドが、
（ｉ）配列番号２４で表されるアミノ酸配列を含むポリペプチド、または
（ｉｉ）配列番号２４で表されるアミノ酸配列において、１もしくは数個のアミノ酸が欠失、置換、付加もしくは挿入されたアミノ酸配列を含み、かつＣＲＣ１活性を有するポリペプチド
である、請求項５または６に記載の酵母菌株。
【請求項８】
ＹＡＴ１活性を有するポリペプチドが、
（ｉ）配列番号２８で表されるアミノ酸配列を含むポリペプチド、または
（ｉｉ）配列番号２８で表されるアミノ酸配列において、１もしくは数個のアミノ酸が欠失、置換、付加もしくは挿入されたアミノ酸配列を含み、かつＹＡＴ１活性を有するポリペプチド
である、請求項５～７のいずれか一項に記載の酵母菌株。
【請求項９】
クラブツリー効果陰性酵母である、請求項１～８のいずれか一項に記載の酵母菌株。
【請求項１０】
キャンディダ・ユティリスである、請求項１～９のいずれか一項に記載の酵母菌株。
【請求項１１】
アセチルＣｏＡを経由する代謝産物を合成するための遺伝子により形質転換されているものである、請求項１～１０のいずれか一項に記載の酵母菌株。
【請求項１２】
ＣｏＡ転移酵素サブユニットＡ（ｃｔｆＡ）、ＣｏＡ転移酵素サブユニットＢ（ｃｔｆＢ）、アセト酢酸脱炭酸酵素（ａｄｃ）および二級アルコール脱水素酵素（ｓａｄｈ）の活性を有するポリペプチドをコードしている４種の遺伝子により形質転換されてなる、請求項１１に記載の酵母菌株。
【請求項１３】
請求項１～１２のいずれか一項に記載の酵母菌株を培養することを含んでなる、物質を製造する方法。
【請求項１４】
前記物質が、アセチルＣｏＡを経由する代謝産物である、請求項１３に記載の方法。
【請求項１５】
前記物質がイソプロパノールである、請求項１３または１４に記載の方法。
【発明の詳細な説明】
【発明の背景】
【０００１】
技術分野
本発明は、酵母において、アセチルＣｏＡを経由する代謝経路を利用した物質の製造方法に関する。
【０００２】
背景技術
深刻化する環境問題を背景に、バイオ燃料が石油の代替燃料として注目を集めている。エタノールはバイオ燃料としてよく知られているが、ガソリンより低いエネルギー密度しか有さないこと、輸送燃料として機能させるためには限られた濃度範囲でガソリンと混合する必要があることから、燃料としての欠点も多い。イソプロパノールは、１－ブタノールよりも高いオクタン価を有し、また、工業原料としても利用される汎用性の高い溶剤である。イソプロパノールは、ナフサ由来のプロピレンの水和によって製造されており、再

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 451 of 3077

生可能な供給源から商業的に生産されたことはない。

【０００３】

微生物によるアセトンやブタノールの発酵生産は、２０世紀初頭に産業的に行われていた。絶対嫌気性菌のある種のクロストリジウム属細菌はアセトンとブタノールを同時に生産することができる（アセトン・ブタノール発酵）。アセトンは細菌細胞内に存在するアセチルＣｏＡを原料として生産される。アセチルＣｏＡは、アセチルＣｏＡアセチル基転移酵素（チオラーゼ）によってアセトアセチルＣｏＡに変換され、さらにＣｏＡ転移酵素によってアセト酢酸に変換され、さらにアセト酢酸脱炭酸酵素によってアセトンへと変換される。ある種のクロストリジウム属細菌はさらに二級アルコールの脱水素酵素を有し、アセトンからイソプロパノールを生産する。Ｃｈｅｎらは、クロストリジウム・ベイジェリンキを含む５２株のクロストリジウム属細菌の中で、最大で３０ｍＭ（１．８ｇ／Ｌ）のイソプロパノールを生産する株が存在したことを報告している（非特許文献１：Chen, J.-S., and S. F. Hiu. Acetone-butanol-isopropanol production by Clostridium beijerinckii (synonym, Clostridium butylicum). Biotechnol. Lett. 1986. 8:371-376）。また、最近になってＨａｎａｉらおよびＪｏｊｉｍａらは、大腸菌にクロストリジウム属由来のＣｏＡ転移酵素、アセト酢酸脱炭酸酵素、二級アルコール脱水素酵素を導入し、内在性のアセチルＣｏＡアセチル基転移酵素を過剰発現させることによってイソプロパノールを効率的に生産できる大腸菌株を報告している（非特許文献２：T. Hanai, S. Atsumi, and J. C. Liao. Engineered Synthetic Pathway for Isopropanol Production in Escherichia coli. Appl Environ Microbiol, Dec. 2007, p. 7814-7818；非特許文献３：Toru Jojima & Masayuki Inui & Hideaki Yukawa. Production of isopropanol by metabolically engineered Escherichia coli. Appl Microbiol Biotechnol 2008 77:1219-1224）。しかしながら、大腸菌は、糖耐性、浸透圧耐性、塩耐性、有機溶剤耐性など各種ストレス耐性はあまり高くないことが多く、ファージ汚染によるリスクもある。また、イソプロパノール耐性が低いことは生産物による発酵の阻害が起こることを意味する。発酵阻害を回避するためにガスストリッピング法などを用いて生産物を培地から除去しながら発酵を続ける方法が知られているが、該方法の使用は発酵生産の工程を煩雑化させるだけでなく、コスト増加につながる。これらの問題を解決するための一つの手段として、高い有機溶剤耐性を有する微生物を用いてイソプロパノール生産をすることが考えられる。

【０００４】

単細胞真核微生物である酵母は高い有機溶剤耐性を示し、また、大腸菌などの微生物と同様に、培養設備、培養時間および培養コストの面で高等生物の細胞よりも安価に培養することができるという特徴を有する。特に、出芽酵母はビールやワインなどの酒類やパンの製造に用いられ、その食品としての安全性が確認されている。

【０００５】

酵母でイソプロパノールを生産するためには、その酵母菌株に、外因性のＣｏＡ転移酵素、アセト酢酸脱炭酸酵素、および二級アルコール脱水素酵素を導入する必要がある。これまで、酵母によるイソプロパノールの製造例としては、サッカロマイセス・セレビシエ（Ｓａｃｃｈａｒｏｍｙｃｅｓ　ｃｅｒｅｖｉｓｉａｅ）にイソプロピルアルコール生産酵素群の遺伝子を導入し、この形質転換体を培養することによりイソプロパノールを製造したことが報告されている（特許文献１：特開２０１１－９７９２９号公報）。しかし、この方法によるイソプロパノールの生産量は、培養液中の濃度として１００ｍｇ／Ｌ前後という少量に過ぎない。酵母は真核生物であり、オルガネラが存在することから、イソプロパノール合成の基質であるアセチルＣｏＡを細胞質に供給する経路が複数存在している。どの経路からアセチルＣｏＡを細胞質に供給することがイソプロパノール合成にとって有利であるのか、あるいはどのような酵母がそういった能力に優れているか、これまでに調査された例はない。

【０００６】

キャンディダ・ユティリス（Ｃａｎｄｉｄａ　ｕｔｉｌｉｓ）は、サッカロマイセス・セレビシエ（Ｓａｃｃｈａｒｏｍｙｃｅｓ　ｃｅｒｅｖｉｓｉａｅ）およびサッカロマイ

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 452 of 3077

セス・フラジリス（Ｓａｃｃｈａｒｏｍｙｃｅｓ　ｆｒａｇｉｌｉｓ）と同様に、ＦＤＡ（米国連邦食品医薬品局）によって食用酵母として認定されており、食品添加物として安全に使用できることが認められている食用酵母である。実際に、現在でも、キャンディダ・ユティリスは、日本をはじめとして、ドイツやアメリカなど世界各国で生産され、食飼料として利用されている。また、キャンディダ・ユティリスは、ペントースの資化による菌体生産、エチルアセテート、Ｌ－グルタミン酸、グルタチオン、インベルターゼなどの生産株として広く産業界で利用されてきた。近年、キャンディダ・ユティリスにおいても形質転換法が開発され、様々な化合物が代謝工学的に生産される技術が開発されている（非特許文献４：Hiroshi Shimada, Keiji Kondo, Paul D. Fraser, Yutaka Miura, Toshiko Saito, and Norihiko Misawa., Increased Carotenoid Production by the Food Yeast Candida utilis through Metabolic Engineering of the Isoprenoid Pathway., Appl Environ Microbiol, July 1998, p. 2676-2680, Vol. 64, No. 7；非特許文献５：Yutaka Miura, Keiji Kondo, Toshiko Saito, Hiroshi Shimada, Paul D. Fraser, and Norihiko MisawaProduction of the Carotenoids Lycopene, β-Carotene, and Astaxanthin in the Food Yeast Candida utilis., Appl Environ Microbiol, April 1998, p. 1226-1229, Vol. 64, No. 4；非特許文献６：Keiji Kondo', Yutaka Miura, Hidetaka Sone, Kazuo Kobayashi & Hiroshi lijima., High-level expression of a sweet protein, monellin, in the food yeast Candida utilis., Nature Biotechnology 15, 453 - 457 (1997)；非特許文献７：Yutaka Miura, Keiji Kondo, Hiroshi Shimada, Toshiko Saito, Katsumi Nakamura, Norihiko Misawa., Production of lycopene by the food yeast, Candida utilis that does not naturally synthesize carotenoid., Biotechnology and Bioengineering Volume 58 Issue 2-3, Pages 306-308；非特許文献８：Shigehito IKUSHIMA, Toshio FUJII, Osamu KOBAYASHI, Satoshi YOSHIDA and Aruto YOSHIDA., Genetic Engineering of Candida utilis Yeast for Efficient Production of L-Lactic Acid., Bioscience, Biotechnology, and Biochemistry Vol. 73 (2009) , No. 8 pp.1818-1824；非特許文献９：Yi-Ren Hong, Ya-Lei Chen, Lynn Farh, Wen-Jen Yang, Chen-Hua Liao and David Shiuan.,Recombinant Candida utilis for the production of biotin., Applied Microbiology and Biotechnology.,Volume 71(2006), Number 2, 211-221）。また、キャンディダ・ユティリスは、サッカロマイセス属とは異なり、クラブツリー（Ｃｒａｂｔｒｅｅ）効果陰性酵母であることから、グルコース存在下でも呼吸活性が低下しない。これにより、十分に通気を行う条件を採用することにより、エタノールの生成を抑えることができる。さらに、キャンディダ・ユティリスは呼吸活性も高く、一旦生成したエタノールも速やかに再同化されることから、効率的な菌体や代謝産物の生産を可能とする。

【０００７】

酵母でイソプロパノールのような物質を生産する場合、一般に、副産物としてのエタノールの生成が問題となる。このエタノール副生を抑制するために、エタノール生産の重要酵素であるピルビン酸脱炭酸酵素遺伝子（ＰＤＣ）を欠損させる取り組みがなされている（特許文献２：国際公開第２０１０／０９５７５１号パンフレット；特許文献３：国際公開第２０１０／０９５７５０号パンフレット；非特許文献１０：Flikweert MT, Van Der Zanden L, Janssen WM, et al. Pyruvate decarboxylase: an indispensable enzyme for growth of Saccharomyces cerevisiae on glucose. Yeast 12(1996): 247-257；非特許文献１１：van Maris AJA, Winkler AA, Porro D, et al. Homofermentative lactate production cannot sustain anaerobic growth of engineered Saccharomyces cerevisiae: Possible consequence of energy-dependent lactate export. Appl Environ Microbiol 70(2004): 2898-2905）。

【０００８】

【非特許文献１】Chen, J.-S., and S. F. Hiu. Acetone-butanol-isopropanol production by Clostridium beijerinckii (synonym, Clostridium butylicum). Biotechnol. Lett. 1986. 8:371-376

【非特許文献２】T. Hanai, S. Atsumi, and J. C. Liao. Engineered Synthetic Pathwa

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 453 of 3077

y for Isopropanol Production in Escherichia coli. Appl Environ Microbiol, Dec. 2007, p. 7814-7818
【非特許文献３】Toru Jojima & Masayuki Inui & Hideaki Yukawa. Production of isopropanol by metabolically engineered Escherichia coli. Appl Microbiol Biotechnol 2008 77:1219-1224
【非特許文献４】Hiroshi Shimada, Keiji Kondo, Paul D. Fraser, Yutaka Miura, Toshiko Saito, and Norihiko Misawa., Increased Carotenoid Production by the Food Yeast Candida utilis through Metabolic Engineering of the Isoprenoid Pathway., Appl Environ Microbiol, July 1998, p. 2676-2680, Vol. 64, No. 7
【非特許文献５】Yutaka Miura, Keiji Kondo, Toshiko Saito, Hiroshi Shimada, Paul D. Fraser, and Norihiko MisawaProduction of the Carotenoids Lycopene, β-Carotene, and Astaxanthin in the Food Yeast Candida utilis., Appl Environ Microbiol, April 1998, p. 1226-1229, Vol. 64, No. 4
【非特許文献６】Keiji Kondo,, Yutaka Miura, Hidetaka Sone, Kazuo Kobayashi & Hiroshi lijima., High-level expression of a sweet protein, monellin, in the food yeast Candida utilis., Nature Biotechnology 15, 453 - 457 (1997)
【非特許文献７】Yutaka Miura, Keiji Kondo, Hiroshi Shimada, Toshiko Saito, Katsumi Nakamura, Norihiko Misawa., Production of lycopene by the food yeast, Candida utilis that does not naturally synthesize carotenoid., Biotechnology and Bioengineering Volume 58 Issue 2-3, Pages 306-308
【非特許文献８】Shigehito IKUSHIMA, Toshio FUJII, Osamu KOBAYASHI, Satoshi YOSHIDA and Aruto YOSHIDA., Genetic Engineering of Candida utilis Yeast for Efficient Production of L-Lactic Acid., Bioscience, Biotechnology, and Biochemistry Vol. 73 (2009) , No. 8 pp.1818-1824
【非特許文献９】Yi-Ren Hong, Ya-Lei Chen, Lynn Farh, Wen-Jen Yang, Chen-Hua Liao and David Shiuan.,Recombinant Candida utilis for the production of biotin., Applied Microbiology and Biotechnology.,Volume 71(2006), Number 2, 211-221
【非特許文献１０】Flikweert MT, Van Der Zanden L, Janssen WM, et al. Pyruvate decarboxylase: an indispensable enzyme for growth of Saccharomyces cerevisiae on glucose. Yeast 12(1996): 247-257
【非特許文献１１】van Maris AJA, Winkler AA, Porro D, et al. Homofermentative lactate production cannot sustain anaerobic growth of engineered Saccharomyces cerevisiae: Possible consequence of energy-dependent lactate export. Appl Environ Microbiol 70(2004): 2898-2905
【特許文献１】特開２０１１－９７９２９号公報
【特許文献２】国際公開第２０１０／０９５７５１号パンフレット
【特許文献３】国際公開第２０１０／０９５７５０号パンフレット

【発明の概要】

【０００９】

本発明者らは、酵母におけるアセチルＣｏＡ合成酵素（ＡＣＳ）およびアセチルＣｏＡアセチル基転移酵素（ＥＲＧ１０）の活性を増強することにより、酵母内におけるアセチルＣｏＡを経由する物質の生産量が顕著に増加することを見出した。本発明はこの知見に基づくものである。

【００１０】

従って、本発明の目的は、様々な物質の発酵生産において中間物質として関与するアセチルＣｏＡならびにアセトアセチルＣｏＡの合成・代謝に係る酵素活性が増強された酵母菌株、ならびに該酵母菌株を用いた物質の製造法を提供することにある。

【００１１】

そして、本発明による酵母菌株は、以下の酵素：

（ａ）アセチルＣｏＡ合成酵素（ＡＣＳ）；および

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 454 of 3077

（ｂ）アセチルＣｏＡアセチル基転移酵素（ＥＲＧ１０）
の活性が増強されている酵母菌株である。
【００１２】
さらに、本発明による製造法は、本発明による酵母菌株を培養することを含んでなる、物質を製造する方法である。
【００１３】
本発明によれば、アセチルＣｏＡやアセトアセチルＣｏＡが関与する酵母の代謝経路が活性化される。アセチルＣｏＡやアセトアセチルＣｏＡが関与する代謝経路は多岐にわたるため、生成する代謝産物も多種多様である。従って、本発明によれば、酵母菌株の培養によって様々な物質を高効率で製造することが可能となる。また、製造しようとする物質に応じて、必要な酵素の遺伝子を発現可能な形で酵母に組み込み、得られた酵母菌株を培養することによりその物質を製造することも可能である。
【発明の具体的説明】
【００１４】
本発明による酵母菌株では、以下の遺伝子：（ａ）アセチルＣｏＡ合成酵素（ＡＣＳ）の活性を有するポリペプチドをコードする遺伝子；および（ｂ）アセチルＣｏＡアセチル基転移酵素（ＥＲＧ１０）の活性を有するポリペプチドをコードする遺伝子の発現が増強されている。
【００１５】
本明細書では、上記のアセチルＣｏＡ合成酵素（ＡＣＳ）遺伝子およびアセチルＣｏＡアセチル基転移酵素（ＥＲＧ１０）遺伝子とともに、イソプロパノールを合成するための４つの遺伝子を過剰発現する酵母菌株を作製し、これを培養することにより、極めて高濃度のイソプロパノールが生産されることが実証されている。この酵母菌株におけるイソプロパノール生産の経路は図１に示す通りである。図１に示される経路から、イソプロパノールの高い生産量が、アセチルＣｏＡおよびアセトアセチルＣｏＡの合成・代謝に係る酵素活性の増強に起因することは明らかである。従って、本発明によって、イソプロパノールのみならず、アセチルＣｏＡおよびアセトアセチルＣｏＡを経由する様々な物質を高生産することができる。また、図１に示されるように、ＡＣＳおよびＥＲＧ１０は細胞質において機能する酵素である。
【００１６】
細胞質に供給されるアセチルＣｏＡは、フラボノイド骨格を有する化合物、テルペノイド骨格を有する化合物、ステロール骨格を有する化合物、脂肪酸など、様々な有用な二次代謝物質の基幹物質である。よって、酵母菌株が天然の状態もしくは人為的に遺伝子を改変された状態でこれらの二次代謝産物の合成経路を有していれば、本発明の方法によりこれらの物質を高生産させることができる。
【００１７】
一般的に、微生物による物質生産を行う際には、副産物生産を抑制する取り組みがなされる場合が多い。例えば、酵母において物質を生産する場合、一般に、副産物としてのエタノールの生成を抑制するために、エタノール生産に関与する酵素遺伝子、例えばピルビン酸脱炭酸酵素遺伝子（ＰＤＣ）を欠損させる取り組みがなされている（国際公開第２０１０／０９５７５１号パンフレット；国際公開第２０１０／０９５７５０号パンフレット；Flikweert MT, Van Der Zanden L, Janssen WM, et al. Pyruvate decarboxylase: an indispensable enzyme for growth of Saccharomyces cerevisiae on glucose. Yeast 12 (1996): 247-257；van Maris AJA, Winkler AA, Porro D, et al. Homofermentative lactate production cannot sustain anaerobic growth of engineered Saccharomyces cerevisiae: Possible consequence of energy-dependent lactate export. Appl Environ Microbiol 70(2004): 2898-2905）。しかし、本明細書では、上記のＰＤＣを欠損させた酵母菌株よりも、ＰＤＣを欠損させなかった酵母菌株の方が顕著に高いイソプロパノール生産量を示すことが実証されている。よって、本発明による酵母菌株は、ＰＤＣなどのエタノール生産に関与する酵素遺伝子を欠失させずにそのまま維持していることも特徴の一つ

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 455 of 3077

であり、これによりイソプロパノール生産量、ひいてはアセチルＣｏＡやその代謝産物であるアセトアセチルＣｏＡの生産量が顕著に向上することは、驚くべきことである。

【００１８】

本明細書において酵素活性が「増強されている」とは、培養過程の少なくとも一時期において、活性増強のための手段を施されていない酵母と比較してより強い酵素活性が観察されることを意味する。酵素活性の増強は、当該酵素遺伝子の発現を増強すること、あるいは比活性のより高い酵素遺伝子への置き換えや、タンパク質工学的な方法による酵素遺伝子の改変によって実現することができる。本明細書において遺伝子の発現が「増強されている」とは、酵母の遺伝的改変により該酵母におけるその遺伝子の発現量が増加すること、つまり、その遺伝子を過剰発現させることを意味する。このような遺伝的改変の方法としては、例えば、その遺伝子により酵母を形質転換する方法（つまり、その遺伝子を発現し得る形で酵母にトランスフェクトする方法）、その酵母のゲノム上に存在する遺伝子のプロモーター活性を強化する方法、その酵母のゲノム上に存在する遺伝子のプロモーターを他のプロモーターと置換することにより該遺伝子の発現を増強する方法など、様々な方法が知られているが、好ましくはその遺伝子により酵母を形質転換する方法とされる。

【００１９】

酵母

本発明による酵母菌株は、特定の遺伝子が過剰発現するように遺伝的に改変された酵母である。遺伝的改変に用いる酵母は、アセチルＣｏＡまたはアセトアセチルＣｏＡを経由する代謝経路を有する酵母であればよく、特に制限されないが、好ましくは呼吸活性の高いクラブツリー効果陰性酵母とされ、より好ましくはクラブトゥリー陰性酵母であるキャンディダ・ユティリスとされる。キャンディダ・ユティリスの菌株は当技術分野において公知の様々な株、例えば、ＮＢＲＣ０３９６株、ＮＢＲＣ０６１９株、ＮＢＲＣ０９８８株、ＮＢＲＣ０６３９株、ＮＢＲＣ０６２６株、ＮＢＲＣ１０８６株、ＮＢＲＣ１０７０７株、ＡＴＣＣ９２５５株、ＡＴＣＣ２８９５５株、ＡＴＣＣ４４６３８株、ＡＴＣＣ９２０６株、ＡＴＣＣ９６６２１株、ＡＴＣＣ１８２０１株、およびＡＴＣＣ１５２３９株等であってよいが、好ましくはＮＢＲＣ１０８６株、ＡＴＣＣ９２５５株またはＡＴＣＣ１５２３９株とされる。

【００２０】

活性が増強される酵素およびその遺伝子

本発明による酵母菌株において活性が増強される酵素は、以下の酵素：（ａ）アセチルＣｏＡ合成酵素（ＡＣＳ）；および（ｂ）アセチルＣｏＡアセチル基転移酵素（ＥＲＧ１０）である。本発明の好ましい実施態様によれば、本発明による酵母菌株においてはこれら２種類の酵素の遺伝子が過剰発現され、すなわち、以下の遺伝子：（ａ）アセチルＣｏＡ合成酵素（ＡＣＳ）の活性を有するポリペプチドをコードする遺伝子；および（ｂ）アセチルＣｏＡアセチル基転移酵素（ＥＲＧ１０）の活性を有するポリペプチドをコードする遺伝子の発現が増強されている。

【００２１】

本発明による酵母菌株では、さらに、以下の遺伝子：（ｃ）ＡＴＰ－クエン酸除去付加酵素（ＡＣＬ）の活性を有するポリペプチドをコードする遺伝子；（ｄ）ミトコンドリア内膜カルニチン輸送体（ＣＲＣ１）の活性を有するポリペプチドをコードする遺伝子；および（ｅ）カルニチンアセチル転移酵素遺伝子（ＹＡＴ１）の活性を有するポリペプチドをコードする遺伝子からなる群から選択される少なくとも一種の遺伝子の発現が増強されていてもよく、好ましくはこれら（ｃ）、（ｄ）および（ｅ）の全ての遺伝子の発現が増強される。

【００２２】

これらの酵素および遺伝子は様々な酵母に由来するものが知られており、その起源は特に制限されない。これらの酵素および遺伝子は、育種に用いられる酵母の内在性遺伝子およびその発現産物であってよく、あるいは、育種に用いられる酵母以外の他の酵母に由来するものであってもよい。例えば、これらの酵素および遺伝子は、それぞれ独立に、キャ

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 456 of 3077

ンディダ・ユティリスまたはヤロウィア・リポリティカに由来するものとすることができる。それぞれの酵素の起源としては、例えば、アセチルＣｏＡ合成酵素（ＡＣＳ）はキャンディダ・ユティリス酵母、アセチルＣｏＡアセチル基転移酵素（ＥＲＧ１０）はキャンディダ・ユティリス酵母、ミトコンドリア内膜カルニチン輸送体（ＣＲＣ１）はキャンディダ・ユティリス酵母、カルニチンアセチル転移酵素遺伝子（ＹＡＴ１）はキャンディダ・ユティリス酵母、ＡＴＰ－クエン酸除去付加酵素（ＡＣＬ）はヤロウィア・リポリティカ酵母（ヤロウィア・リポリティカ由来のＡＣＬ１とＡＣＬＹがサブユニットを構成する遺伝子）が好ましい。

【００２３】

キャンディダ・ユティリス由来のアセチルＣｏＡ合成酵素（ＡＣＳ）遺伝子には、ＡＣＳ１遺伝子とＡＣＳ２遺伝子がある。ＡＣＳ１遺伝子のコード配列は配列番号１１で表され、コードされるアミノ酸配列は配列番号１２で表される。ＡＣＳ２遺伝子のコード配列は配列番号１３で表され、コードされるアミノ酸配列は配列番号１４で表される。アセチルＣｏＡ合成酵素（ＡＣＳ）は、酢酸をアセチルＣｏＡに変換する酵素である。本発明に用いられるＡＣＳ遺伝子はＡＣＳ１およびＡＣＳ２のいずれであってもよいが、好ましくはＡＣＳ２とされる。また、アセチルＣｏＡ合成酵素（ＡＣＳ）の活性を有するポリペプチドは、配列番号１２または配列番号１４で表されるアミノ酸配列において、１もしくは数個のアミノ酸が欠失、置換、付加もしくは挿入されたアミノ酸配列を含み、かつＡＣＳ活性を有するポリペプチドであってもよい。

【００２４】

キャンディダ・ユティリス由来のアセチルＣｏＡアセチル基転移酵素（ＥＲＧ１０）遺伝子のコード配列は配列番号６１で表され、コードされるアミノ酸配列は配列番号６２で表される。アセチルＣｏＡアセチル基転移酵素（ＥＲＧ１０）は、アセチルＣｏＡをアセトアセチルＣｏＡに変換する酵素である。また、アセチルＣｏＡアセチル基転移酵素（ＥＲＧ１０）の活性を有するポリペプチドは、配列番号６２で表されるアミノ酸配列において、１もしくは数個のアミノ酸が欠失、置換、付加もしくは挿入されたアミノ酸配列を含み、かつＥＲＧ１０活性を有するポリペプチドであってもよい。

【００２５】

ヤロウィア・リポリティカ由来のＡＣＬ１遺伝子のコード配列は配列番号３１で表され、コードされるアミノ酸配列は配列番号３２で表される。また、ＡＣＬ１の活性を有するポリペプチドは、配列番号３２で表されるアミノ酸配列において、１もしくは数個のアミノ酸が欠失、置換、付加もしくは挿入されたアミノ酸配列を含み、かつＡＣＬ１活性を有するポリペプチドであってもよい。さらに、ヤロウィア・リポリティカ由来のＡＣＬＹ遺伝子のコード配列は配列番号３３で表され、コードされるアミノ酸配列は配列番号３４で表される。また、ＡＣＬＹの活性を有するポリペプチドは、配列番号３４で表されるアミノ酸配列において、１もしくは数個のアミノ酸が欠失、置換、付加もしくは挿入されたアミノ酸配列を含み、かつＡＣＬＹ活性を有するポリペプチドであってもよい。これらのＡＣＬ１とＡＣＬＹをサブユニットとするＡＴＰ－クエン酸除去付加酵素（ＡＣＬ）は、クエン酸をアセチルＣｏＡに変換する酵素である。

【００２６】

キャンディダ・ユティリス由来のミトコンドリア内膜カルニチン輸送体（ＣＲＣ１）遺伝子のコード配列は配列番号２３で表され、コードされるアミノ酸配列は配列番号２４で表される。ミトコンドリア内膜カルニチン輸送体（ＣＲＣ１）は、アセチルカルニチンをミトコンドリアから細胞質へ輸送する酵素である。また、ミトコンドリア内膜カルニチン輸送体（ＣＲＣ１）の活性を有するポリペプチドは、配列番号２４で表されるアミノ酸配列において、１もしくは数個のアミノ酸が欠失、置換、付加もしくは挿入されたアミノ酸配列を含み、かつＣＲＣ１活性を有するポリペプチドであってもよい。

【００２７】

キャンディダ・ユティリス由来のカルニチンアセチル転移酵素遺伝子（ＹＡＴ１）遺伝子のコード配列は配列番号２７で表され、コードされるアミノ酸配列は配列番号２８で表

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 457 of 3077

される。カルニチンアセチル転移酵素遺伝子（ＹＡＴ１）は、アセチルカルニチンをアセチルＣｏＡに変換する酵素である。また、カルニチンアセチル転移酵素遺伝子（ＹＡＴ１）の活性を有するポリペプチドは、配列番号２８で表されるアミノ酸配列において、１もしくは数個のアミノ酸が欠失、置換、付加もしくは挿入されたアミノ酸配列を含み、かつＹＡＴ１活性を有するポリペプチドであってもよい。

【００２８】

アミノ酸の欠失、置換、付加、又は挿入は、上記ポリペプチドをコードする遺伝子を、当技術分野で公知の手法によって改変することによって行うことができる。遺伝子に変異を導入するには、Ｋｕｎｋｅｌ法又はＧａｐｐｅｄ ｄｕｐｌｅｘ法等の公知手法又はこれに準ずる方法により行うことができ、例えば部位特異的突然変異誘発法を利用した変異導入用キット、例えばＭｕｔａｎｔ－Ｋ（タカラバイオ社）やＭｕｔａｎｔ－Ｇ（タカラバイオ社）などを用いて、あるいは、タカラバイオ社のＬＡ ＰＣＲ ｉｎ ｖｉｔｒｏ Ｍｕｔａｇｅｎｅｓｉｓシリーズキット、ＫＯＤ－Ｐｌｕｓ－Ｍｕｔａｇｅｎｅｓｉｓ Ｋｉｔ（ＴＯＹＯＢＯ）などを用いて変異を導入することができる。上記の各酵素の活性は、当技術分野において公知の手法により確認することができる。

【００２９】

上記の遺伝子に代えて、それぞれの同等物を用いることもできる。この同等物は、それぞれの遺伝子と同等の機能を有することを条件に、一部のヌクレオチド残基が異なる遺伝子を意味する。このような同等物としては、それぞれのヌクレオチド配列と７０％以上、好ましくは８０％以上、より好ましくは８５％以上、さらに好ましくは９０％以上、最も好ましくは９５％以上の相同性（配列同一性）があり、かつ各酵素の活性を有するポリペプチドをコードするヌクレオチド配列を含む遺伝子が挙げられる。前記同等物としてはさらに、それぞれのヌクレオチド配列もしくはその相補配列とストリンジェントな条件下でハイブリダイズし、かつ各酵素の活性を有するポリペプチドをコードするヌクレオチド配列を含む遺伝子が挙げられる。前記同等物としてはさらに、それぞれのヌクレオチド配列において１もしくは数個のヌクレオチド残基が欠失、置換、付加、または挿入された配列を含み、かつ各酵素の活性を有するポリペプチドをコードするヌクレオチド配列を含む遺伝子が挙げられる。

【００３０】

ここで、ヌクレオチド残基の欠失、置換、付加、又は挿入は、上記配列を含む遺伝子を、当技術分野で公知の手法によって改変することによって行うことができる。遺伝子に変異を導入するには、Ｋｕｎｋｅｌ法又はＧａｐｐｅｄ ｄｕｐｌｅｘ法等の公知手法又はこれに準ずる方法により行うことができ、例えば部位特異的突然変異誘発法を利用した変異導入用キットを用いればよい。例えばＭｕｔａｎｔ－Ｋ（タカラバイオ社）やＭｕｔａｎｔ－Ｇ（タカラバイオ社）などを用いて、あるいは、タカラバイオ社のＬＡ ＰＣＲ ｉｎ ｖｉｔｒｏ Ｍｕｔａｇｅｎｅｓｉｓシリーズキット、ＫＯＤ－Ｐｌｕｓ－Ｍｕｔａｇｅｎｅｓｉｓ Ｋｉｔ（ＴＯＹＯＢＯ）などを用いて変異を導入することができる。各酵素の活性は、当技術分野において公知の手法により確認することができる。

【００３１】

相同性（配列同一性）を示す数値（％）は、塩基配列比較用プログラム：例えばＧＥＮＥＴＹＸ－ＷＩＮ７．０．０を用いて、デフォルト（初期設定）のパラメーターにより算出されるものである。すなわち、酵母染色体上の各遺伝子が、同一ではないが同等の機能、すなわち各活性を有するポリペプチドをコードする遺伝子によって相同組換え等を介して置換されていてもよい。各酵素の活性は、当技術分野において公知の手法により確認することができる。

【００３２】

ストリンジェントな条件とは、例えば、Ｒａｐｉｄ－Ｈｙｂ Ｂｕｆｆｅｒ（ＧＥヘルスケアバイオサイエンス社製）を用い、温度条件を好ましくは４０～７０℃、より好ましくは６０℃として、その他は添付のプロトコールに従って行うハイブリダイゼーション条件である。その後、例えば当業者に公知の一般的な方法を用い、２×ＳＳＣと０．１％（

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 458 of 3077

ｗ／ｖ）ＳＤＳから成り立つ溶液での５分間の洗浄、続いて１×ＳＳＣと０．１％（ｗ／ｖ）ＳＤＳから成り立つ溶液での１０分間の洗浄、さらに０．１×ＳＳＣと０．１％（ｗ／ｖ）ＳＤＳから成り立つ溶液での１０分間の洗浄を行うことを指す。ただしハイブリダイゼーション時の温度条件や、その後のメンブレンの洗浄に用いる溶液の塩濃度等の条件を適宜設定することにより、ある一定（７０％、８０％、８５％、９０％、９５％のいずれか）以上の相同性を有する塩基配列を含むＤＮＡをクローニングできる。そのようにして得られる遺伝子が、配列上は同一ではないが同等の機能、すなわち各活性を有するポリペプチドをコードする遺伝子によって相同組換え等を介して置換されていてもよい。各酵素の活性は、当技術分野において公知の手法により確認することができる。

【００３３】

有用物質を合成するための遺伝子

本発明による酵母菌株は、所望の有用物質の製造に用いるために、該物質の生産に必要となる遺伝子で形質転換されていてもよい。例えば、本発明の方法によってアセチルＣｏＡを経由する代謝産物を製造するために、本発明による酵母菌株は、アセチルＣｏＡを経由する代謝産物を合成するための遺伝子により形質転換されていてもよい。必要な遺伝子は、製造しようとする物質に応じて、当業者であれば適切に選択することができる。

【００３４】

例えば、イソプロパノールを製造しようとする場合には、さらなる遺伝的改変として、イソプロパノール合成遺伝子で酵母菌株を形質転換することができる。イソプロパノール合成遺伝子としては、特開２０１１－９７９２９号公報に記載の酵素遺伝子群など、当業者に公知の遺伝子を用いることができ、例えば、ＣｏＡ基転移酵素サブユニットＡ（ｃｔｆＡ）、同サブユニットＢ（ｃｔｆＢ）、アセト酢酸脱炭酸酵素（ａｄｃ）および二級アルコール脱水素酵素（ｓａｄｈ）の４つの遺伝子の組み合わせを用いることができる。

【００３５】

本発明の一つの実施態様によれば、本発明による酵母菌株は、さらなる遺伝的改変として、ＣｏＡ基転移酵素サブユニットＡ（ｃｔｆＡ）、ＣｏＡ転移酵素サブユニットＢ（ｃｔｆＢ）、アセト酢酸脱炭酸酵素（ａｄｃ）および二級アルコール脱水素酵素（ｓａｄｈ）の活性を有するポリペプチドをコードしている４種の遺伝子により形質転換されている。このような酵母菌株はイソプロパノールの製造に用いることができ、この製造法は、本実施態様による酵母菌株を培養することを含む。

【００３６】

これらの遺伝子は様々な生物に由来するものが知られており、その起源は特に制限されない。例えば、これらの遺伝子は、それぞれ独立に、クロストリジウム・アセトブチリカムまたはクロストリジウム・ベイジェリンキに由来するものとすることができる。それぞれの酵素の起源としては、例えば、ＣｏＡ転移酵素遺伝子サブユニットＡ（ｃｔｆＡ）および同サブユニットＢ（ｃｔｆＢ）はクロストリジウム・アセトブチリカムが好ましく、アセト酢酸脱炭酸酵素（ａｄｃ）および二級アルコール脱水素酵素（ｓａｄｈ）はクロストリジウム・ベイジェリンキが好ましい。

【００３７】

クロストリジウム・アセトブチリカム由来のＣｏＡ転移酵素サブユニットＡ（ｃｔｆＡ）のアミノ酸配列は配列番号２で表される。ＣｏＡ転移酵素サブユニットＡ（ｃｔｆＡ）は、同サブユニットＢ（ｃｔｆＢ）と一緒になって、アセトアセチルＣｏＡをアセト酢酸に変換する酵素である。配列番号２で表されるアミノ酸配列をコードする遺伝子配列は、クロストリジウム・アセトブチリカム由来の同遺伝子の配列など、当業者であれば適宜選択することができるため、特に制限されないが、好ましくは本発明による酵母菌株の種において頻度の高いコドンを使用した配列とされる。例えば、キャンディダ・ユティリスにおけるコドン頻度に適合した遺伝子配列としては、配列番号１で表されるヌクレオチド配列が挙げられる。ＣｏＡ転移酵素サブユニットＡ（ｃｔｆＡ）の活性を有するポリペプチドは、配列番号２で表されるアミノ酸配列において、１もしくは数個のアミノ酸が欠失、置換、付加もしくは挿入されたアミノ酸配列を含み、かつｃｔｆＡ活性を有するポリペプ

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 459 of 3077

チドであってもよい。

【００３８】

クロストリジウム・アセトブチリカム由来のＣｏＡ転移酵素サブユニットＢ（ｃｔｆＢ）のアミノ酸配列は配列番号４で表される。ＣｏＡ転移酵素サブユニットＢ（ｃｔｆＢ）は、同サブユニットＡ（ｃｔｆＡ）と一緒になって、アセトアセチルＣｏＡをアセト酢酸に変換する酵素である。配列番号４で表されるアミノ酸配列をコードする遺伝子配列は、クロストリジウム・アセトブチリカム由来の同遺伝子の配列など、当業者であれば適宜選択することができるため、特に制限されないが、好ましくは本発明による酵母菌株の種において頻度の高いコドンを使用した配列とされる。例えば、キャンディダ・ユティリスにおけるコドン頻度に適合した遺伝子配列としては、配列番号３で表されるヌクレオチド配列が挙げられる。ＣｏＡ転移酵素サブユニットＢ（ｃｔｆＢ）の活性を有するポリペプチドは、配列番号４で表されるアミノ酸配列において、１もしくは数個のアミノ酸が欠失、置換、付加もしくは挿入されたアミノ酸配列を含み、かつｃｔｆＢ活性を有するポリペプチドであってもよい。

【００３９】

クロストリジウム・ベイジェリンキ由来のアセト酢酸脱炭酸酵素（ａｄｃ）のアミノ酸配列は配列番号６で表される。アセト酢酸脱炭酸酵素（ａｄｃ）は、アセト酢酸をアセトンに変換する酵素である。配列番号６で表されるアミノ酸配列をコードする遺伝子配列は、クロストリジウム・ベイジェリンキ由来の同遺伝子の配列など、当業者であれば適宜選択することができるため、特に制限されないが、好ましくは本発明による酵母菌株の種において頻度の高いコドンを使用した配列とされる。例えば、キャンディダ・ユティリスにおけるコドン頻度に適合した遺伝子配列としては、配列番号５で表されるヌクレオチド配列が挙げられる。アセト酢酸脱炭酸酵素（ａｄｃ）の活性を有するポリペプチドは、配列番号６で表されるアミノ酸配列において、１もしくは数個のアミノ酸が欠失、置換、付加もしくは挿入されたアミノ酸配列を含み、かつａｄｃ活性を有するポリペプチドであってもよい。

【００４０】

クロストリジウム・ベイジェリンキ由来の二級アルコール脱水素酵素（ｓａｄｈ）のアミノ酸配列は配列番号８で表される。二級アルコール脱水素酵素（ｓａｄｈ）は、アセトンをイソプロパノールに変換する酵素である。配列番号８で表されるアミノ酸配列をコードする遺伝子配列は、クロストリジウム・ベイジェリンキ由来の同遺伝子の配列など、当業者であれば適宜選択することができるため、特に制限されないが、好ましくは本発明による酵母菌株の種において頻度の高いコドンを使用した配列とされる。例えば、キャンディダ・ユティリスにおけるコドン頻度に適合した遺伝子配列としては、配列番号７で表されるヌクレオチド配列が挙げられる。二級アルコール脱水素酵素（ｓａｄｈ）の活性を有するポリペプチドは、配列番号８で表されるアミノ酸配列において、１もしくは数個のアミノ酸が欠失、置換、付加もしくは挿入されたアミノ酸配列を含み、かつｓａｄｈ活性を有するポリペプチドであってもよい。

【００４１】

アミノ酸の欠失、置換、付加、又は挿入は、上記ポリペプチドをコードする遺伝子を、当技術分野で公知の手法によって改変することによって行うことができる。遺伝子に変異を導入するには、Ｋｕｎｋｅｌ法又はＧａｐｐｅｄ ｄｕｐｌｅｘ法等の公知手法又はこれに準ずる方法により行うことができ、例えば部位特異的突然変異誘発法を利用した変異導入用キット、例えばＭｕｔａｎｔ－Ｋ（タカラバイオ社）やＭｕｔａｎｔ－Ｇ（タカラバイオ社）などを用いて、あるいは、タカラバイオ社のＬＡ ＰＣＲ ｉｎ ｖｉｔｒｏ Ｍｕｔａｇｅｎｅｓｉｓシリーズキット、ＫＯＤ－Ｐｌｕｓ－Ｍｕｔａｇｅｎｅｓｉｓ Ｋｉｔ（ＴＯＹＯＢＯ）などを用いて変異を導入することができる。上記の各酵素の活性は、当技術分野において公知の手法により確認することができる。

【００４２】

上記の遺伝子に代えて、それぞれの同等物を用いることもできる。この同等物は、それ

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 460 of 3077

ぞれの遺伝子と同等の機能を有することを条件に、一部のヌクレオチド残基が異なる遺伝子を意味する。このような同等物としては、それぞれのヌクレオチド配列と７０％以上、好ましくは８０％以上、より好ましくは８５％以上、さらに好ましくは９０％以上、最も好ましくは９５％以上の相同性（配列同一性）があり、かつ各酵素の活性を有するポリペプチドをコードするヌクレオチド配列を含む遺伝子が挙げられる。前記同等物としてはさらに、それぞれのヌクレオチド配列もしくはその相補配列とストリンジェントな条件下でハイブリダイズし、かつ各酵素の活性を有するポリペプチドをコードするヌクレオチド配列を含む遺伝子が挙げられる。前記同等物としてはさらに、それぞれのヌクレオチド配列において１もしくは数個のヌクレオチド残基が欠失、置換、付加、または挿入された配列を含み、かつ各酵素の活性を有するポリペプチドをコードするヌクレオチド配列を含む遺伝子が挙げられる。

【００４３】

ここで、ヌクレオチド残基の欠失、置換、付加、又は挿入は、上記配列を含む遺伝子を、当技術分野で公知の手法によって改変することによって行うことができる。遺伝子に変異を導入するには、Ｋｕｎｋｅｌ法又はＧａｐｐｅｄ ｄｕｐｌｅｘ法等の公知手法又はこれに準ずる方法により行うことができ、例えば部位特異的突然変異誘発法を利用した変異導入用キットを用いればよい。例えばＭｕｔａｎｔ－Ｋ（タカラバイオ社）やＭｕｔａｎｔ－Ｇ（タカラバイオ社）などを用いて、あるいは、タカラバイオ社のＬＡ ＰＣＲ ｉｎ ｖｉｔｒｏ Ｍｕｔａｇｅｎｅｓｉｓシリーズキット、ＫＯＤ－Ｐｌｕｓ－Ｍｕｔａｇｅｎｅｓｉｓ Ｋｉｔ（ＴＯＹＯＢＯ）などを用いて変異を導入することができる。各酵素の活性は、当技術分野において公知の手法により確認することができる。

【００４４】

相同性（配列同一性）を示す数値（％）は、塩基配列比較用プログラム：例えばＧＥＮＥＴＹＸ－ＷＩＮ７．０．０を用いて、デフォルト（初期設定）のパラメーターにより算出されるものである。すなわち、酵母染色体上の各遺伝子が、同一ではないが同等の機能、すなわち各活性を有するポリペプチドをコードする遺伝子によって相同組換え等を介して置換されていてもよい。各酵素の活性は、当技術分野において公知の手法により確認することができる。

【００４５】

ストリンジェントな条件とは、例えば、Ｒａｐｉｄ－Ｈｙｂ Ｂｕｆｆｅｒ（ＧＥヘルスケアバイオサイエンス社製）を用い、温度条件を好ましくは４０～７０℃、より好ましくは６０℃として、その他は添付のプロトコールに従って行うハイブリダイゼーション条件である。その後、例えば当業者に公知の一般的な方法を用い、２×ＳＳＣと０．１％（ｗ／ｖ）ＳＤＳから成り立つ溶液での５分間の洗浄、続いて１×ＳＳＣと０．１％（ｗ／ｖ）ＳＤＳから成り立つ溶液での１０分間の洗浄、さらに０．１×ＳＳＣと０．１％（ｗ／ｖ）ＳＤＳから成り立つ溶液での１０分間の洗浄を行うことを指す。ただしハイブリダイゼーション時の温度条件や、その後のメンブレンの洗浄に用いる溶液の塩濃度等の条件を適宜設定することにより、ある一定（７０％、８０％、８５％、９０％、９５％のいずれか）以上の相同性を有する塩基配列を含むＤＮＡをクローニングできる。そのようにして得られる遺伝子が、配列上は同一ではないが同等の機能、すなわち各活性を有するポリペプチドをコードする遺伝子によって相同組換え等を介して置換されていてもよい。各酵素の活性は、当技術分野において公知の手法により確認することができる。

【００４６】

イソプロパノール以外にも、リコペンなどの特定のカロテノイド、スクアレンあるいはスクアレンから生産される特定のステロールなどの物質を生産する酵母菌株を構築することが可能である。例えば、Ｅｒｗｉｎｉａ ｕｒｅｄｏｖｏｒａ由来のＧＧＰＰ合成酵素遺伝子、フィトエン合成酵素遺伝子、フィトエン不飽和化酵素遺伝子を発現し、内在性のＨＭＧ－ＣｏＡ還元酵素遺伝子（ＨＭＧ１）を過剰発現するキャンディダ・ユティリスはリコペンを高蓄積することが報告されている（Shimada H, Kondo K, Fraser PD, Miura Y, Saito T, Misawa N. Increased carotenoid production by the food yeast Candida u

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 461 of 3077

tilis through metabolic engineering of the isoprenoid pathway. Appl Environ Microbiol. 1998 Jul;64(7):2676-80）。また、ザイモステロール-２４-メチルトランスフェラーゼ遺伝子およびエルゴスタ５，７，２４，（２８）－トリエノール－２２－脱水素酵素遺伝子の発現に欠損を持ち、ＨＭＧ１が過剰発現するように操作されたサッカロマイセス・セレビシエ組み換え株は、高濃度のスクアレンを蓄積することが知られている（特開平５－１９２１８４号公報）。これらの手法を用いることにより、リコペン、スクアレンを生産する酵母菌株を構築することが可能である。

【００４７】

酵母菌株の分子育種

本発明による酵母菌株の分子育種は、上記の遺伝子群を過剰発現させるための遺伝的改変を酵母菌株に施すこと、例えば、上記の遺伝子群を発現可能な状態で酵母菌株に導入することによって行うことができる。これら遺伝子群は、相同組換え等の技術により酵母の染色体上に組み込まれることが好ましい。このような染色体上への組込みを実現するための相同組換え用遺伝子配列の選択は、当業者において周知であり、当業者であれば必要に応じて適切な相同組換え用遺伝子配列を選択して相同組換え用ＤＮＡ断片を構成することができる。

【００４８】

上記の遺伝子群が酵母ゲノムに組込まれる染色体上の位置は特に制限されるものではないが、例えば、オロチジン５’リン酸デカルボキシラーゼ遺伝子（ＵＲＡ３）座、アルファアミノアジピン酸還元酵素遺伝子（ＬＹＳ２）座などを選択することができる。また、これらの遺伝子は、組み込まれた遺伝子座に当初から存在するプロモーター、例えばＵＲＡ３遺伝子プロモーター、ＬＹＳ２遺伝子プロモーターなどの制御下に置いてもよいが、グリセルアルデヒド－３－リン酸脱水素酵素（ＧＡＰ）遺伝子プロモーターなどの他のプロモーターの制御下に置いてもよい。

【００４９】

本発明の好ましい実施態様によれば、本発明による酵母菌株の育種に用いられる酵母はキャンディダ・ユティリスとされる。キャンディダ・ユティリスは、その倍数性が高く、胞子を形成しない。倍数性の高い株の遺伝子に変異を導入しようとする場合、１倍体の株に比べて、その変異を重度に加える必要があるが、その際に、変異を与えたい遺伝子ではない遺伝子にも変異が加えられる可能性が高まると考えられる。従って、キャンディダ・ユティリスの遺伝子に変異を導入する場合には、標的とする遺伝子のみに効率よく多重に変異を加えることができる技術を用いることが好ましい。キャンディダ・ユティリスの形質転換法としては、特開２００３－１４４１８５号公報に記載の技術が挙げられる。この文献では、利用可能なベクターとして、キャンディダ・ユティリスの染色体ＤＮＡと相同な配列と、選択マーカー遺伝子とを含んでなり、相同組換えによって異種遺伝子をキャンディダ・ユティリスの染色体ＤＮＡに組み込むことができるもの、あるいは、キャンディダ・ユティリスで自律複製機能を有するＤＮＡ配列と、選択マーカー遺伝子とを含んでなり、高い頻度でキャンディダ・ユティリスを形質転換できるものが開発されている。

【００５０】

形質転換系で利用可能な選択マーカー遺伝子としては、酵母において機能し得る薬剤耐性マーカー、好ましくはシクロヘキシミド耐性型Ｌ４１遺伝子、ジェネティシン（Ｇ４１８）耐性を付与する遺伝子、またはハイグロマイシンＢ耐性を付与する遺伝子などがある。ジェネティシン（Ｇ４１８）耐性を付与する遺伝子、またはハイグロマイシンＢ耐性を付与する遺伝子は、野生の酵母には存在しない配列であるため、標的の遺伝子座に組込まれる確率が高いと考えられている。また、これらの薬剤耐性遺伝子は、その宿主の形質に与える影響も小さいと考えられている（Ｂａｇａｎｚ Ｆら，１３（１６）：１５６３－７３．，１９９７）（Ｃｏｒｄｅｒｏ Ｏｔｅｒｏ Ｒら，Ａｐｐｌ Ｍｉｃｒｏｂｉｏｌ Ｂｉｏｔｅｃｈｎｏｌ，４６（２）：１４３－８．，１９９６）。

【００５１】

他の形質転換系として、バクテリオファージＰ１由来のＣｒｅ－ｌｏｘＰ系が挙げられ

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 462 of 3077

る。これは、２つの３４ｂｐのｌｏｘＰ配列間での部位特異的組換えシステムであり、この組換えはＣｒｅ遺伝子がコードするＣｒｅ組換え酵素によって触媒される。このシステムは、サッカロマイセス・セレビシエなどの酵母細胞においても機能することが報告されており、２つのｌｏｘＰ配列の間に配置された選択マーカー遺伝子は、ｌｏｘＰ配列間の組換えによって除去されることが知られている（Ｇｕｌｄｅｎｅｒ，Ｕ．ら，Ｎｕｃｌｅｉｃ Ａｃｉｄｓ Ｒｅｓ．，２４，２５１９－２４．，１９９６）。このシステムはクルイベロマイセス・ラクティス（Ｋｌｕｙｖｅｒｏｍｙｃｅｓ ｌａｃｔｉｓ）など、複数の酵母種で利用されている（Ｓｔｅｅｎｓｍａ，Ｈ．Ｙ．ら，Ｙｅａｓｔ，１８，４６９－７２．，２００１）。

【００５２】

物質の製造方法

本発明による酵母菌株を適当な炭素源の存在下で培養することにより、培養物中に物質（例えばイソプロパノール）を製造することができる。本発明による物質（例えばイソプロパノール）製造法によれば、培養系から物質（例えばイソプロパノール）を分離する工程を実施することにより、物質（例えばイソプロパノール）を得ることができる。なお、本発明において培養物とは、培養上清の他、培養細胞あるいは菌体、細胞もしくは菌体の破砕物を包含している。

【００５３】

本発明による酵母菌株の培養にあたっては、酵母の種類に応じて培養方法や培養条件を選択することができる。培養方法としては、例えば、試験管、フラスコあるいはジャーファーメンターを用いた液体培養法を挙げることができ、回分培養、半回分培養などの培養形式を採用できる。振盪速度および通気量は、一般的には５０～２００ｒｐｍ、好ましくは８０～１５０ｒｐｍ、あるいはこの振盪速度に相当する通気量とされる。初期の菌体量としてのOD_{600}は、一般的には０．１～３０、好ましくは０．５～１０とされる。さらに、本発明による酵母菌株の培養にあたっては、振盪速度は速い方が好ましく（つまり通気量は多い方が好ましく）、初期OD_{600}は高い方が好ましい。例えば、試験管やフラスコでの培養では、振盪速度は約１２０ｒｐｍまたはそれ以上、初期OD_{600}は約１０とすることが好ましい。

【００５４】

本発明による物質（例えばイソプロパノール）製造法において、培地の組成は、酵母が生育し、且つ目的の物質を製造できる各種栄養素を含む組成であればよく、特に限定されない。培地に含まれる資化炭素源としては、例えば、グルコースのほかにもキシロースやスクロースでも資化できれば用いることができる。本発明の好ましい実施態様によれば、炭素源はグルコースを含む。培地中のグルコース濃度は、目的の物質の生産効率を確認しながら、当業者であれば適宜調節することができる。特に、本発明による酵母菌株の培養では、培地中のグルコース濃度は、比較的低濃度に保つことが好ましい。例えば、１９２時間の培養において、グルコースの初期濃度を１００ｇ／Ｌとし、培養７２時間後に終濃度１００ｇ／Ｌとなるようにグルコース固体を添加した場合に比べて、グルコースの初期濃度を５０ｇ／Ｌとし、４８時間後、９６時間後および１４４時間後に終濃度５０ｇ／Ｌとなるようにグルコース固体を添加した場合の方が、物質の生産効率が高い。

【００５５】

培地に含まれる栄養源としては、例えば、酵母エキス、ペプトン、ホエーなどが用いられるが、ＹＰ（１０ｇ／Ｌ酵母エキス、２０ｇ／Ｌペプトン）に上記の資化炭素源を加えた培地、例えばＹＰＤ５（５０ｇ／Ｌグルコース、１０ｇ／Ｌ酵母エキス、２０ｇ／Ｌペプトン）等の培地で、さらに適宜ｐＨ調整されたものが便利である。

【００５６】

発酵温度は、用いる酵母の生育可能な範囲で選択することができる。発酵温度は、例えば、約１５℃～４５℃とすることができ、より好ましくは２０～４０℃、さらに好ましくは２５～３５℃、最も好ましくは約３０℃とする。また、発酵過程における培地のｐＨは３～８に保持することが好ましく、より好ましくはｐＨ４～７であり、必要に応じてｐＨ

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 463 of 3077

調節剤を用いることができる。用いるｐＨ調節剤としては、炭酸カルシウム、水酸化ナトリウム、水酸化カリウム等が挙げられるが、好ましくは炭酸カルシウムとされる。

【００５７】

物質（例えばイソプロパノール）製造に要する反応時間は特に限定されず、本発明の効果が認められる限り任意の反応時間で実施される。これらの条件の最適化は、当業者であれば容易に行うことができる。

【００５８】

本発明による物質（例えばイソプロパノール）の製造法においては、このようにして製造した物質（例えばイソプロパノール）成分を培地から分離・回収するが、そのための具体的な手順は、分離・回収しようとする物質に応じて、当業者であれば容易に選択することができる。

【実施例】

【００５９】

以下、実施例を挙げて本発明をより具体例に説明するが、これら実施例は本発明の技術的範囲を制限するものではない。

【００６０】

実施例１：高いイソプロパノール耐性を示すキャンディダ・ユティリス株の選抜

キャンディダ・ユティリス　ＮＢＲＣ０３９６株、ＮＢＲＣ０６１９株、ＮＢＲＣ０９８８株、ＮＢＲＣ０６３９株、ＮＢＲＣ０６２６株、ＮＢＲＣ１０８６株、ＮＢＲＣ１０７０７株、ＡＴＣＣ９２５５株、ＡＴＣＣ２８９５５株、ＡＴＣＣ４４６３８株、ＡＴＣＣ９２０６株、ＡＴＣＣ９６６２１株、ＡＴＣＣ１８２０１株、およびＡＴＣＣ１５２３９株を、ＹＰＤ２（２０ｇ／Ｌペプトン、１０ｇ／Ｌ酵母エキス、２０ｇ／Ｌグルコース）に接種し、３０℃で培養を行った。定常期の酵母培養液を１０倍ごとに段階希釈し、１０～１００ｇ／ＬのイソプロパノールをふくむＹＰＤ２寒天培地にスポットし、３０℃で培養した。３日間培養を継続し、コロニーの形成を観察した。結果を表１に記載する。

【００６１】

【表１】

表１

菌株	生育限界 ＩＰＡ濃度（ｇ／ｌ）
ＮＢＲＣ０３９６	５０
ＮＢＲＣ０６１９	６０
ＮＢＲＣ０９８８	４０
ＮＢＲＣ０６３９	５０
ＮＢＲＣ０６２６	５０
ＮＢＲＣ１０８６	６５
ＮＢＲＣ１０７０７	５５
ＡＴＣＣ９２５５	６５
ＡＴＣＣ２８９５５	５５
ＡＴＣＣ４４６３８	６０
ＡＴＣＣ９２０６	６０
ＡＴＣＣ９６６２１	５５
ＡＴＣＣ１８２０１	５０
ＡＴＣＣ１５２３９	６５

【００６２】

以上の結果より、アッセイを行った株の中では、ＮＢＲＣ１０８６株、ＡＴＣＣ９２５５株、およびＡＴＣＣ１５２３９株が、最も高い６５ｇ／ＬのＩＰＡ耐性を示すことが明らかとなった。

【００６３】

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 464 of 3077

実施例２：ピルビン酸脱炭酸酵素遺伝子が破壊されたキャンディダ・ユティリスＡＴＣＣ１５２３９株の構築

キャンディダ・ユティリスＡＴＣＣ１５２３９株において、その内因性遺伝子であるピルビン酸脱炭酸酵素遺伝子（ＰＤＣ）を破壊した。この遺伝子破壊は、Ｃｒｅ－ｌｏｘＰシステムを用いて国際公開第２０１０／０９５７５０号パンフレットと同様の方法で行った。２回の破壊により、ＰＤＣ１アレルが完全に欠損した株（ｄＰＤＣ１株）を構築した。ｄＰＤＣ１株は完全にはエタノール生産能を欠損しなかったが、エタノール生産能は著しく低下していた。

【００６４】

実施例３：クロストリジウム・アセトブチリカム由来のＣｏＡ転移酵素サブユニットＡ遺伝子、同サブユニットＢ遺伝子、クロストリジウム・ベイジェリンキ由来のアセト酢酸脱炭酸酵素遺伝子および二級アルコール脱水素酵素遺伝子を発現するベクターの構築

イソプロパノール生合成に関わる４つの遺伝子、クロストリジウム・アセトブチリカム由来のＣｏＡ転移酵素遺伝子サブユニットＡ（ｃｔｆＡ）、同サブユニットＢ（ｃｔｆＢ）、クロストリジウム・ベイジェリンキ由来のアセト酢酸脱炭酸酵素（ａｄｃ）および二級アルコール脱水素酵素（ｓａｄｈ）を効率良くキャンディダ・ユティリスで発現させるために、以下の項目を指標として、天然に存在しない新規な遺伝子配列を設計し、合成した：

（１）キャンディダ・ユティリスにおいて多用されているコドンの使用；

（２）遺伝子クローニングに不適当な制限酵素部位の排除；および

（３）染色体組み込み型発現ベクターに導入するための、５’末端におけるＸｂａＩサイトおよび３’末端におけるＢａｍＨＩサイトの付与。

【００６５】

合成されたＣｏＡ転移酵素遺伝子サブユニットＡ遺伝子、同サブユニットＢ遺伝子、アセト酢酸脱炭酸酵素遺伝子および二級アルコール脱水素酵素遺伝子のヌクレオチド配列を、それぞれ配列番号１、３、５および７に示す。また、これらのヌクレオチド配列によってコードされるアミノ酸配列を、それぞれ配列番号２、４、６および８に示す。これらの遺伝子を有するプラスミドをそれぞれＧＡ０９１４２－ｃｔｆＡ、ＧＡ０９１４２－ｃｔｆＢ、ＧＡ０９１４２－ａｄｃ、ＧＡ０９１４２－ｓａｄｈと名付けた。

【００６６】

ｐＶＴ９２は、キャンディダ・ユティリスのオロチジン５’酸脱炭酸酵素遺伝子（ＵＲＡ３）座への組み込みベクターであり、ハイグロマイシン耐性遺伝子を選択マーカーとして有している（国際公開第２０１０／０９５７５０号パンフレット）。ｐＶＴ９２は、ＢｇｌＩＩ処理してリニアＤＮＡにすることにより、染色体導入することができる。ｐＶＴ９２は、外来遺伝子を発現させるために、ＸｂａＩ／ＢａｍＨＩギャップの外側にそれぞれ、グリセルアルデヒド脱水素酵素遺伝子プロモーターとグリセリン酸３’リン酸リン酸化酵素遺伝子ターミネーターを有している。また、ｐＶＴ９２は、プロモーターの５’末端側およびターミネーターの３’末端側には、それぞれＮｈｅＩサイトおよびＳｐｅＩサイトを有している。

【００６７】

ｐＶＴ２３９は、キャンディダ・ユティリスのアルファアミノアジピン酸還元酵素遺伝子（ＬＹＳ２）座への組み込みベクターであり、ジェネティシン耐性遺伝子を選択マーカーとして有している（国際公開第２０１０／０９５７５０号パンフレット）。ｐＶＴ２３９は、ＮｏｔＩ処理してリニアＤＮＡにすることにより、染色体導入することができる。また、ｐＶＴ２３９は、ジェネティシン耐性遺伝子の５’末端側にＳｐｅＩサイトを有している。

【００６８】

各イソプロパノール合成酵素遺伝子を含む４つのベクターＧＡ０９１４２－ｃｔｆＡ、ＧＡ０９１４２－ｃｔｆＢ、ＧＡ０９１４２－ａｄｃ、およびＧＡ０９１４２－ｓａｄｈを、ＸｂａＩおよびＢａｍＨＩ処理し、アガロースゲル抽出による精製を行った後、同制

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 465 of 3077

限酵素で処理したｐＶＴ９２にライゲーションして、ｐＶＴ２９６、ｐＶＴ２９８、ｐＶＴ３００およびｐＶＴ３０２を得た。ｐＶＴ２９６およびｐＶＴ３００はＳｐｅＩ処理し、脱リン酸化を行った。また、ｐＶＴ２９８およびｐＶＴ３０２はＮｈｅＩ、ＳｐｅＩ処理を行った後、ｃｔｆＢ発現カセット、ｓａｄｈ発現カセットをそれぞれ含むＤＮＡ断片の精製を行い、それぞれｐＶＴ２９６およびｐＶＴ３００に連結して、ｐＶＴ３２１およびｐＶＴ３２５を得た。ｐＶＴ３２５をＮｈｅＩおよびＳｐｅＩで処理した後、ａｄｃ発現カセットとｓａｄｈ発現カセットを含むＤＮＡ断片の精製を行い、ｐＶＴ２３９のＳｐｅＩサイトに導入し、ｐＶＴ３４６を得た。最終的に得られたｐＶＴ３２１はｃｔｆＡとｃｔｆＢのそれぞれの発現カセットとハイグロマイシン耐性遺伝子を有し、キャンディダ・ユティリスのＵＲＡ３座に組み込まれるベクターであり、ｐＶＴ３４６はａｄｃとｓａｄｈのそれぞれの発現カセットとジェネティシン耐性遺伝子を有し、キャンディダ・ユティリスのＬＹＳ２座に組み込まれるベクターである。

【００６９】

実施例４：４つのイソプロパノール合成遺伝子を有するキャンディダ・ユティリス株の構築

ｐＶＴ３２１をＢｇｌＩＩで消化し、エタノール沈殿により濃縮した。得られたＤＮＡ断片を用いて、Ｃ．ｕｔｉｌｉｓ　ＡＴＣＣ１５２３９株、および実施例２で得られたｄＰＤＣ１株を形質転換し、６００μｇ／ｍＬハイグロマイシンを含むＹＰＤ培地に塗抹し、３０℃で２日間培養した。導入した全ての遺伝子を有する株を選抜した。

【００７０】

さらにｐＶＴ３４６をＮｏｔＩで消化し、エタノール沈殿により濃縮した。得られたそれぞれのＤＮＡ断片を用いて上記で得られた２株をそれぞれ形質転換し、６００μｇ／ｍＬハイグロマイシンと２００μｇ／ｍＬジェネティシンを含むＹＰＤ培地に塗抹し、３０℃で２日間培養した。導入した全ての遺伝子を有する株を選抜し、ＡＴＣＣ１５２３９株由来のイソプロパノール生産株をＴＭＳ２７２株、ｄＰＤＣ株由来のイソプロパノール生産株をＴＭＳ２９０株とした。

【００７１】

実施例５：イソプロパノール生産試験

実施例４で得られた形質転換体のイソプロパノール生産能の評価を、以下に示すように実施した。

【００７２】

各種形質転換体を２ｍＬ　ＹＰＤ液体培地（２０ｇ／Ｌペプトン、１０ｇ／Ｌ酵母エキス、２０ｇ／Ｌグルコース）／１４ｍＬ容試験管に接種し、１４０ｒｐｍ、３０℃で２４時間振とう培養した。得られた前培養液を１００ｍＬ　ＹＰＤ１０（２０ｇ／Ｌペプトン、１０ｇ／Ｌ酵母エキス、１００ｇ／Ｌグルコース）／２００ｍＬ容三角フラスコもしくはＹＰＤ１０Ｃ（２０ｇ／Ｌペプトン、１０ｇ／Ｌ酵母エキス、１００ｇ／Ｌグルコース、４５ｇ／Ｌ炭酸カルシウム）／２００ｍＬ容三角フラスコに接種し、初期ＯＤ６００＝１、１００ｒｐｍ、３０℃で振とう培養した。培養液から経時的にサンプリングし、０．２μｍフィルターでろ過した後に、各種成分の定量分析を行った。グルコース、酢酸、エタノールおよびグリセロールは、高速液体クロマトフラフィー（島津製作所製）、ＩＳｅｐ－ＩＯＮ３００カラム（東京化成工業製）および示唆屈折計を用いて定量した。イソプロパノールおよびアセトンは、ガスクロマトフラフィーＧＣ－２０１０　Ｐｌｕｓ（島津製作所製）を用いて定量した。

【００７３】

また、フラスコでの簡易的な中和培養としてＹＰＤ１０Ｃ（２０ｇ／Ｌペプトン、１０ｇ／Ｌ酵母エキス、１００ｇ／Ｌグルコース、４５ｇ／Ｌ炭酸カルシウム）培地を用いても同様の実験を行った。

【００７４】

非中和条件下での培養試験の結果を図２、中和条件下での培養試験の結果を図３に記載する。なお、データは３反復試行の平均で示した。

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 466 of 3077

【００７５】
図２によれば、親株であるＡＴＣＣ１５２３９株およびｄＰＤＣ１株ではイソプロパノールの生産が認められなかったのに対し、イソプロパノール合成遺伝子を導入したＴＭＳ２７２株およびＴＭＳ２９０株では、それぞれ最大で０．１６ｇ／Ｌおよび０．１３ｇ／Ｌのイソプロパノールの生産が認められた。また、ＰＤＣ１遺伝子を破壊した株では糖の消費速度および増殖能が著しく低下していた。ＴＭＳ２７２株では培養２４時間でグルコースが完全に消費され、エタノールの蓄積が認められたが、その後の培養の継続に伴ってエタノール濃度が低下した。エタノール濃度が低下している間もイソプロパノールの生産が認められたことから、ＴＭＳ２７２株は、一度生産したエタノールを二次消費することによりイソプロパノールを生産していることが考えられた。ｄＰＤＣ１株に由来するＴＭＳ２９０株ではイソプロパノール生産速度が低下したことから、酵母におけるイソプロパノール生産においては、基質である細胞質のアセチルＣｏＡがエタノールから供給される経路が最も重要であることが示唆された。エタノールの同化は酸化的反応であることから、これらの結果は、酵母によるイソプロパノール生産においてはキャンディダ・ユティリスのような呼吸活性の高いクラブツリー陰性酵母が適している可能性を示唆している。

【００７６】
図３によれば、培地に炭酸カルシウムを添加したことにより、ＴＭＳ２７２株は１．１ｇ／Ｌのイソプロパノールを生産し、非中和条件より生産量が増大した。また、ＴＭＳ２７２株は、１６．２ｇ／Ｌの酢酸を蓄積しており、何らかの理由により酢酸が蓄積したことでｐＨが低下し、イソプロパノールの生産を阻害していたことが考えられた。また、ＴＭＳ２９０株においても、中和条件下ではイソプロパノール生産量が０．９ｇ／Ｌまで増大しており、一方で３７．４ｇ／Ｌという高濃度のグリセロールを蓄積していた。

【００７７】
実施例６：過剰発現によってイソプロパノール生産性を向上させる遺伝子の選抜
過剰発現することによりイソプロパノールの生産性を向上させる遺伝子の選抜を行った。候補遺伝子が触媒する代謝経路を図１に記載する。

【００７８】
キャンディダ・ユティリス内在性のアセチルＣｏＡ加水分解酵素遺伝子（ＡＣＨ１；ヌクレオチド配列：配列番号９；アミノ酸配列：配列番号１０）、アセチルＣｏＡ合成酵素遺伝子（ＡＣＳ１；ヌクレオチド配列：配列番号１１；アミノ酸配列：配列番号１２、ＡＣＳ２；ヌクレオチド配列：配列番号１３；アミノ酸配列：配列番号１４）、細胞質局在型エタノール脱水素酵素遺伝子（ＡＤＨ２；ヌクレオチド配列：配列番号１５；アミノ酸配列：配列番号１６）、ミトコンドリア局在型エタノール脱水素酵素遺伝子（ＡＤＨ３；ヌクレオチド配列：配列番号１７；アミノ酸配列：配列番号１８）、細胞質局在型アセトアルデヒド脱水素酵素遺伝子（ＡＬＤ１；ヌクレオチド配列：配列番号１９；アミノ酸配列：配列番号２０）、ミトコンドリア局在型アセトアルデヒド脱水素酵素遺伝子（ＡＬＤ２；ヌクレオチド配列：配列番号２１；アミノ酸配列：配列番号２２）、ミトコンドリア内膜カルニチン輸送体（ＣＲＣ１；ヌクレオチド配列：配列番号２３；アミノ酸配列：配列番号２４）、ミトコンドリア局在型カルニチンアセチルCoA転移酵素（ＣＡＴ２；ヌクレオチド配列：配列番号２５；アミノ酸配列：配列番号２６）、カルニチンアセチル転移酵素遺伝子（ＹＡＴ１；ヌクレオチド配列：配列番号２７；アミノ酸配列：配列番号２８、ＹＡＴ２；ヌクレオチド配列：配列番号２９；アミノ酸配列：配列番号３０）、アセチルＣｏＡアセチル基転移酵素（ＥＲＧ１０；ヌクレオチド配列：配列番号６１；アミノ酸配列：配列番号６２）、ヤロウィア・リポリティカ由来のＡＴＰ－クエン酸除去付加酵素遺伝子（ＡＣＬ１；ヌクレオチド配列：配列番号３１；アミノ酸配列：配列番号３２と、ＡＣＬＹ；ヌクレオチド配列：配列番号３３；アミノ酸配列：配列番号３４がサブユニットを構成する遺伝子）を、それぞれのゲノムＤＮＡを鋳型としてＰＣＲにより増幅した。ＰＣＲに用いたプライマーは以下の通りである：（ＡＣＨ１；配列番号３５と配列番号３６）、（ＡＣＳ１；配列番号３７と配列番号３８）、（ＡＣＳ２；配列番号３９と配列番号４０）、（ＡＤＨ２；配列番号４１と配列番号４２）、（ＡＤＨ３；配列番号４３と配

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 467 of 3077

列番号４４）、（ＡＬＤ１；配列番号４５と配列番号４６）、（ＡＬＤ２；配列番号４７と配列番号４８）、（ＣＡＴ２；配列番号４９と配列番号５０）、（ＣＲＣ１；配列番号５１と配列番号５２）、（ＹＡＴ１；配列番号５３と配列番号５４）、（ＹＡＴ２；配列番号５５と配列番号５６）、（ＡＣＬ１；配列番号５７と配列番号５８）、（ＡＣＬＹ；配列番号５９と配列番号６０）、（ＥＲＧ１０；配列番号６３と配列番号６４）。

【００７９】

得られたＰＣＲ産物をＸｂａＩ／ＢａｍＨＩ処理して、同制限酵素で処理したｐＣＵ１５５（Keiji Kondo, , Yutaka Miura, Hidetaka Sone, Kazuo Kobayashi & Hiroshi liji ma., High-level expression of a sweet protein, monellin, in the food yeast Candida utilis., Nature Biotechnology 15, 453 - 457 (1997)）にクローニングした。ｐＣＵ１５５は、シクロヘキシミド耐性遺伝子を選抜マーカーとして有し、ＵＲＡ３座にマルチコピーインテグレーションされるベクターであり、ＸｂａＩ／ＢａｍＨＩギャップの外側にそれぞれ、グリセルアルデヒド脱水素酵素遺伝子プロモーターと同ターミネーターを有している。これまでに構築しているイソプロパノール生産株のＵＲＡ３座にはイソプロパノール合成酵素遺伝子が組み込まれているが、Ｃ．ｕｔｉｌｉｓは高次倍数体であり、ＵＲＡ３座が複数コピー存在するため、さらにＵＲＡ３座に遺伝子の導入が可能である。また、イソプロパノール合成遺伝子はＵＲＡ３座に二点組み換えにより挿入されているため、ｐＣＵ１５５の組み込みにより、導入された遺伝子の置換が起こることはない。候補遺伝子がクローン化されたプラスミドはＢｇｌＩＩで消化し、エタノール沈殿により濃縮した。得られたＤＮＡ断片を用いてＣ．ｕｔｉｌｉｓ ＴＭＳ２７２株を形質転換し、６００μｇ／ｍＬハイグロマイシン、２００μｇ／ｍＬジェネティシンおよび２０μｇ／ｍＬシクロヘキシミドを含むＹＰＤ培地に塗抹し、３０℃で４日間培養した。なお、ＡＣＬ１とＡＣＬＹのＤＮＡは混合し、同時に形質転換を行った。得られた各４クローンについて発酵試験を行い、イソプロパノール生産性を調査した。結果を図４に示す。

【００８０】

図４によると、ＡＣＳ１、ＡＣＳ２、ＣＲＣ１、ＹＡＴ１、ＥＲＧ１０、またはＡＣＬの過剰発現により、イソプロパノール生産量が１．５～４．０倍になることが明らかとなった。ＡＣＳ１、ＡＣＳ２、ＹＡＴ１、およびＡＣＬの各遺伝子はアセチルＣｏＡの合成に関わる酵素遺伝子であり、ＣＲＣ１遺伝子はアセチルＣｏＡ合成の前駆物質をミトコンドリアから細胞質に輸送するキャリアタンパク質遺伝子であり、ＥＲＧ１０遺伝子はアセチルＣｏＡからアセトアセチルＣｏＡを合成する遺伝子である。これらの結果により、酵母によるイソプロパノール高生産のためには、細胞質のアセチルＣｏＡを効率的に生産するための酵素遺伝子の発現または活性をアップレギュレートすることが重要であることが示される。

【００８１】

実施例７：ＡＣＬ１／ＡＣＬＹ、ＡＣＳ２、ＥＲＧ１０、ＣＲＣ１およびＹＡＴ１を同時に過剰発現するイソプロパノール生産株の構築と培養試験

ＡＣＬ１、ＡＣＬＹ、ＡＣＳ２、ＥＲＧ１０、ＣＲＣ１およびＹＡＴ１の各遺伝子がクローン化されたプラスミドをＢｇｌＩＩＩで消化し、エタノール沈殿により濃縮した。得られた各５μｇのＤＮＡ断片を混合し、この混合物を用いてＴＭＳ２７２株を形質転換し、６００μｇ／ｍＬハイグロマイシン、２００μｇ／ｍＬジェネティシンおよび１０μｇ／ｍＬシクロヘキシミドを含むＹＰＤ培地に塗抹し、３０℃で４日間培養した。得られた８０クローンについて発酵試験を行い、イソプロパノール生産量が最も高い株を選抜し、ＴＭＳ３７８株とした。

【００８２】

ＴＭＳ３７８株の培養試験を行った結果を図５に示す。ＴＭＳ２７２株が１．３ｇ／Ｌのイソプロパノールを生産したのに対して、ＴＭＳ３７８株は９．８ｇ／Ｌと約７．５倍のイソプロパノール生産量を示した。また、ＴＭＳ２７２株では酢酸が１８．１ｇ／Ｌ蓄積したのに対して、ＴＭＳ３７８株は４．５ｇ／Ｌと、主要な副産物である酢酸が約４分の１に低下していた。

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 468 of 3077

【００８３】

実施例８：ＴＭＳ３７８株のイソプロパノール生産における初期ＯＤ$_{600}$および振とう速度の影響

イソプロパノール生産に適したＴＭＳ３７８株の培養条件の検討を行った。ＴＭＳ３７８株を２ｍＬ ＹＰＤ液体培地（２０ｇ／Ｌペプトン、１０ｇ／Ｌ酵母エキス、２０ｇ／Ｌグルコース）／１４ｍＬ容試験管に接種し、１４０ｒｐｍ、３０℃で２４時間振とう培養した。得られた前培養液を１００ｍＬ ＹＰＤ１０Ｃ（２０ｇ／Ｌペプトン、１０ｇ／Ｌ酵母エキス、１００ｇ／Ｌグルコース、４５ｇ／Ｌ炭酸カルシウム）／３００ｍＬ容三角フラスコに接種し、初期ＯＤ$_{600}$を１もしくは１０とし、１００ｒｐｍもしくは１２０ｒｐｍ、３０℃で振とう培養した。培養７２時間後に終濃度１００ｇ／Ｌとなるようにグルコース固体を培養液に添加し、さらに１４４時間まで培養を行った。経時的にサンプリングして０．２μｍフィルターでろ過した。各種定量分析は、実施例５に記載の方法に従って行った。なお、データは３反復試行の平均で示した。結果を図６に示す。

【００８４】

図６によると、初期ＯＤ６００＝１０、１２０ｒｐｍで培養したときに最も高いイソプロパノール生産量が見られ、培養１４４時間後に最大で１８．５ｇ／Ｌのイソプロパノールが生産された。その際の対糖収率は３０．８％であった。このことから、振とう速度が速いほど（通気量が多いほど）、また、初期ＯＤが高いほど、イソプロパノール生産には好ましいことが明らかとなった。

【００８５】

実施例９：ＴＭＳ３７８株のイソプロパノール生産における添加グルコース濃度の影響

イソプロパノール生産に適したＴＭＳ３７８株の培養条件の検討を行った。ＴＭＳ３７８株を２ｍＬ ＹＰＤ液体培地（２０ｇ／Ｌペプトン、１０ｇ／Ｌ酵母エキス、２０ｇ／Ｌグルコース）／１４ｍＬ容試験管に接種し、１４０ｒｐｍ、３０℃で２４時間振とう培養した。得られた前培養液を１００ｍＬ ＹＰＤ１０Ｃ（２０ｇ／Ｌペプトン、１０ｇ／Ｌ酵母エキス、１００ｇ／Ｌグルコース、４５ｇ／Ｌ炭酸カルシウム）／３００ｍＬ容三角フラスコもしくはＹＰＤ５Ｃ（２０ｇ／Ｌペプトン、１０ｇ／Ｌ酵母エキス、５０ｇ／Ｌグルコース、４５ｇ／Ｌ炭酸カルシウム）／３００ｍＬ容三角フラスコに接種し、初期ＯＤ６００＝１、１２０ｒｐｍ、３０℃で振とう培養した。ＹＰＤ１０Ｃで培養を行ったサンプルについては、培養７２時間後に終濃度１００ｇ／Ｌとなるようにグルコース固体を添加し、さらに１９２時間まで培養を行った。また、ＹＰＤ５Ｃで培養を行ったサンプルについては、培養４８時間おきに３回終濃度５０ｇ／Ｌとなるようにグルコース固体を添加し、さらに１９２時間まで培養を行った。いずれの条件も最終のグルコース濃度は２００ｇ／Ｌである。経時的にサンプリングして０．２μｍフィルターでろ過した。各種定量分析は、実施例５に記載の方法に従って行った。なお、データは３反復試行の平均で示した。結果を図７に示す。

【００８６】

図７によると、ＹＰＤ１０Ｃで培養し、培養７２時間後に１００ｇ／Ｌのグルコースを添加した場合は、培養１２２時間後に最大で１９．２ｇ／Ｌのイソプロパノールが生産された。しかし、それ以降はイソプロパノール生産量の増大は認められず、酢酸が蓄積した（２７．２ｇ／Ｌ）。イソプロパノールの生産における最大の対糖収率は２９．０％であった。一方、ＹＰＤ５Ｃで培養し、４８時間おきに５０ｇ／Ｌのグルコースを３回添加した場合は、培養１９２時間までイソプロパノールの増大が認められ、培養１９２時間後に最大で２７．４ｇ／Ｌのイソプロパノールが生産された。また、酢酸の蓄積はほとんど認められなかった。この条件でのイソプロパノールの生産における最大の対糖収率は４１．５％であった。このことから、添加するグルコースの濃度を比較的低濃度に保つことはイソプロパノール生産において好ましいことが明らかとなった。

【００８７】

実施例１０：イソプロパノール生産量向上に寄与の大きい遺伝子の特定

同時に過剰発現することによってイソプロパノール生産量を向上させる寄与の大きい遺

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 469 of 3077

伝子を特定するために、(i)ＡＣＬ１、ＡＣＬＹおよびＡＣＳ２の組み合わせを過剰発現する株、(ii)ＡＣＬ１、ＡＣＬＹおよびＥＲＧ１０の組み合わせを過剰発現する株、(iii)ＡＣＳ２およびＥＲＧ１０の組み合わせを過剰発現する株、(iv)ＡＣＬ１、ＡＣＬＹ、ＡＣＳ２およびＥＲＧ１０の組み合わせを過剰発現する株をそれぞれ構築し、ＡＣＬ１、ＡＣＬＹ、ＡＣＳ２、ＣＲＣ１、ＥＲＧ１０およびＹＡＴ１の組み合わせを過剰発現する株（ＴＭＳ３７８株）およびそれぞれの単独遺伝子の過剰発現株との間で、イソプロパノール生産量の比較を行った。組み換え株の構築は実施例６および実施例７と同様の方法を用いて行い、組み込みたい遺伝子ＤＮＡを混合し、同時に形質転換を行った。得られた各３２クローンについて発酵試験を行い、それぞれの導入遺伝子の組み合わせにおいて平均的なイソプロパノール生産量を示す株を選抜した。各構築株を２ｍＬ ＹＰＤ液体培地（２０ｇ／Ｌペプトン、１０ｇ／Ｌ酵母エキス、２０ｇ／Ｌグルコース）／１４ｍＬ容試験管に接種し、１４０ｒｐｍ、３０℃で２４時間振とう培養した。得られた前培養液を３ｍＬ ＹＰＤ１０Ｃ（２０ｇ／Ｌペプトン、１０ｇ／Ｌ酵母エキス、１００ｇ／Ｌグルコース、４５ｇ／Ｌ炭酸カルシウム）／ガラス試験管に接種し、初期ＯＤ６００＝１、１００ｒｐｍ、３０℃で振とう培養した。経時的にサンプリングを行い、高速液体クロマトフラフィーおよびガスクロマトフラフィーＧＣ－２０１０でイソプロパノール生産量を評価した。なお、データは３反復試行の平均で示した。

【００８８】

図８によると、ＡＣＳ１、ＡＣＳ２およびＥＲＧ１０のそれぞれ単独の過剰発現により、イソプロパノール生産量が２．２～２．７倍になることが明らかとなった。(i)ＡＣＬ１、ＡＣＬＹおよびＡＣＳ２の組み合わせを過剰発現する株、ならびに(ii)ＡＣＬ１、ＡＣＬＹおよびＥＲＧ１０の組み合わせを過剰発現する株のイソプロパノール生産量は、それぞれＡＣＳ２およびＥＲＧ１０を単独で過剰発現する株と有意な差は認められなかった。一方、(iii)ＡＣＳ２およびＥＲＧ１０の組み合わせを過剰発現する株ではイソプロパノール生産の飛躍的な向上が認められ、生産量が１１．６倍に向上した。また、(iii)ＡＣＳ２およびＥＲＧ１０の組み合わせを過剰発現する株、(iv)ＡＣＬ１、ＡＣＬＹ、ＡＣＳ２およびＥＲＧ１０の組み合わせを過剰発現する株、ＴＭＳ３７８株（ＡＣＬ１、ＡＣＬＹ、ＡＣＳ２、ＣＲＣ１、ＥＲＧ１０およびＹＡＴ１の組み合わせを過剰発現する株）の株間において、イソプロパノール生産量に有意な差は認められなかった。

【００８９】

これらの結果より、同時に過剰発現した際にはＡＣＬ、ＣＲＣ１およびＹＡＴ１の過剰発現がイソプロパノールの生産性向上に与える寄与は小さく、イソプロパノール生産性を向上させるためにはＡＣＳ２とＥＲＧ１０を同時に過剰発現することが最も重要であることが示された。以上の結果は、酵母によるイソプロパノール高生産のためにはエタノールからアセチルＣｏＡを合成する代謝経路、すなわち酢酸からアセチルＣｏＡを合成する代謝経路に関わる酵素遺伝子の発現または活性を増強することが重要であることを示している。

【図面の簡単な説明】

【００９０】

【図１】図１は、酵母によるイソプロパノール合成経路を示す図である。

【図２】図２は、イソプロパノール生産遺伝子を導入した株の非中和条件での発酵試験の結果を示す図であり、培養液中のイソプロパノール濃度（パネルＡ）、グルコース濃度（パネルＢ）、エタノール濃度（パネルＣ）、酢酸濃度（パネルＤ）、グリセロール濃度（パネルＥ）およびＯＤ６００（パネルＦ）の経時変化を示している。各パネルにおいて、黒丸はＡＴＣＣ１５２３９株の結果を示し、白丸はＴＭＳ２７２株（ＡＴＣＣ１５２３９株＋ＩＰＡ合成遺伝子）の結果を示し、黒三角はＰＤＣ１破壊株の結果を示し、白三角はＴＭＳ２９０株（ＰＤＣ破壊株＋ＩＰＡ合成遺伝子）の結果を示す。

【図３】図３は、イソプロパノール生産遺伝子を導入した株の中和条件での発酵試験の結果を示す図であり、培養液中のイソプロパノール濃度（パネルＡ）、グルコース濃度（パネルＢ）、エタノール濃度（パネルＣ）、酢酸濃度（パネルＤ）およびグリセロール濃度

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 470 of 3077

（パネルＥ）の経時変化を示している。各パネルにおいて、黒丸はＡＴＣＣ１５２３９株の結果を示し、白丸はＴＭＳ２７２株（ＡＴＣＣ１５２３９株＋ＩＰＡ合成遺伝子）の結果を示し、黒三角はＰＤＣ１破壊株の結果を示し、白三角はＴＭＳ２９０株（ＰＤＣ破壊株＋ＩＰＡ合成遺伝子）の結果を示す。

【図４】図４は、ＴＭＳ２７２株において、各種遺伝子を過剰発現させた場合におけるイソプロパノール生産量を、ＴＭＳ２７２株によるイソプロパノール生産量に対する相対量として示したグラフである。

【図５】図５は、イソプロパノール生酸性を向上させる遺伝子を全て過剰発現する株の発酵試験の結果を示す図であり、培養液中のイソプロパノール濃度（パネルＡ）、グルコース濃度（パネルＢ）、エタノール濃度（パネルＣ）、酢酸濃度（パネルＤ）、グリセロール濃度（パネルＥ）およびアセトン濃度（パネルＦ）の経時変化を示している。各パネルにおいて、黒丸はＴＭＳ２７２株の結果を示し、白丸はＴＭＳ３７８株（ＴＭＳ２７２株＋ＡＣＬ１：ＡＣＬＹ：ＡＣＳ２：ＥＲＧ１０：ＣＲＣ１：ＹＡＴ１）の結果を示す。

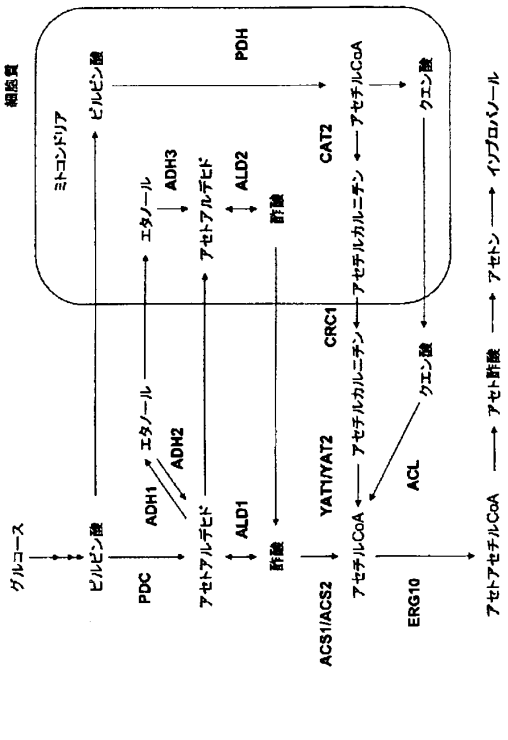
【図６】図６は、イソプロパノール生産性を向上させる遺伝子を全て過剰発現するＴＭＳ３７８株の３０℃での発酵試験（初期ＯＤ$_{600}$および振とう速度の影響）の結果を示す図であり、培養液中のイソプロパノール濃度（パネルＡ）、グルコース濃度（パネルＢ）、エタノール濃度（パネルＣ）、酢酸濃度（パネルＤ）、グリセロール濃度（パネルＥ）およびアセトン濃度（パネルＦ）の経時変化を示している。各パネルにおいて、黒丸は初期ＯＤ６００＝１かつ撹拌数＝１００ｒｐｍの条件下での結果を示し、黒三角は初期ＯＤ６００＝１かつ撹拌数＝１２０ｒｐｍの条件下での結果を示し、白丸は初期ＯＤ６００＝１０かつ撹拌数＝１００ｒｐｍの条件下での結果を示し、白三角は初期ＯＤ６００＝１０かつ撹拌数＝１２０ｒｐｍの条件下での結果を示す。

【図７】図７は、イソプロパノール生産性を向上させる遺伝子を全て過剰発現するＴＭＳ３７８株の３０℃での発酵試験（流加グルコース濃度の影響）の結果を示す図である。パネルＡは、初期グルコース濃度を１００ｇ／Ｌとし、培養７２時間後に１００ｇ／ｌのグルコースを流加して、１９２時間培養を行った結果を示す。パネルＢは、初期グルコース濃度を５０ｇ／Ｌとし、培養４８時間後、９６時間後および１４４時間後にそれぞれ５０ｇ／ｌのグルコースを流加して、１９２時間培養を行った結果を示す。各パネルにおいて、イソプロパノール濃度（黒三角）、グルコース濃度（黒丸）、エタノール濃度（白丸）および酢酸濃度（白三角）の経時変化を示している。

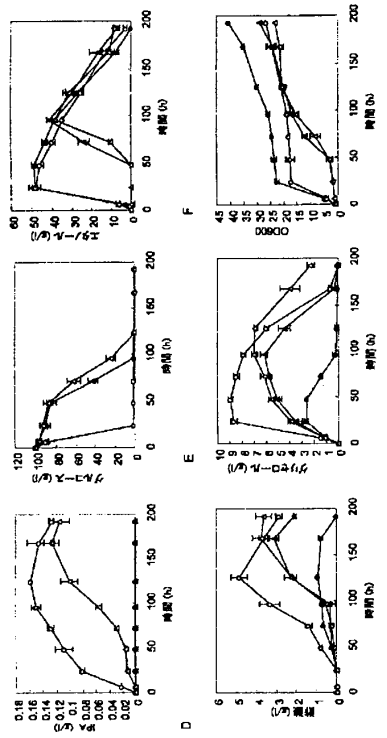
【図８】図８は、ＴＭＳ２７２株において、各種遺伝子を過剰発現させた場合におけるイソプロパノール生産量を示したグラフである。

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 471 of 3077

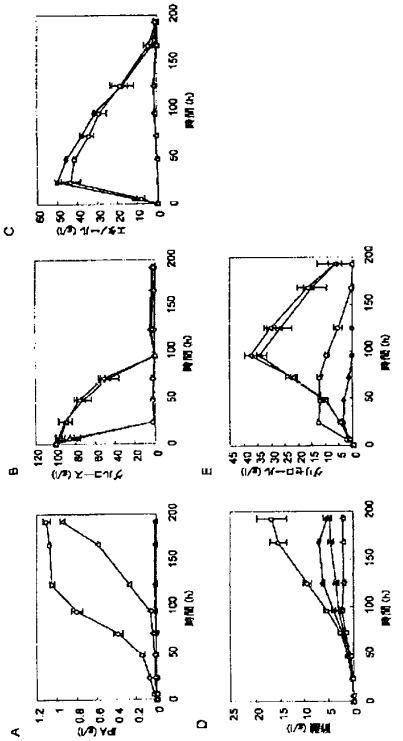
【図１】


細胞質
ミトコンドリア
グルコース
ピルビン酸
PDC
アセトアルデヒド
ADH1
ADH2
ALD1
エタノール
酢酸
ACS1/ACS2
アセチルCoA
YAT1/YAT2
CRC1
アセチルカルニチン
ERG10
ACL
クエン酸
PDH
ADH3
ALD2
CAT2
アセトアセチルCoA
CTF
アセト酢酸
ADC
アセトン
SADH
イソプロパノール

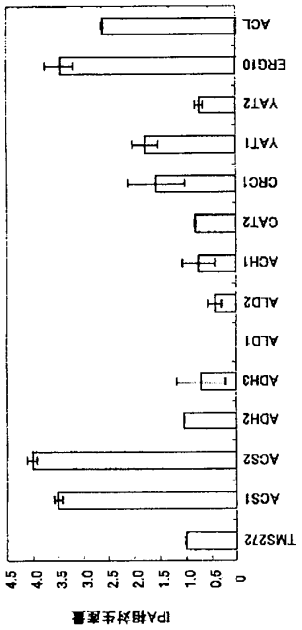
【図２】



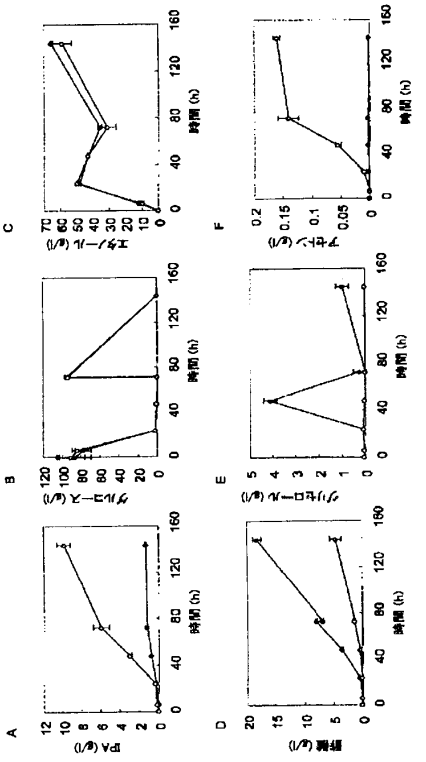
【図３】



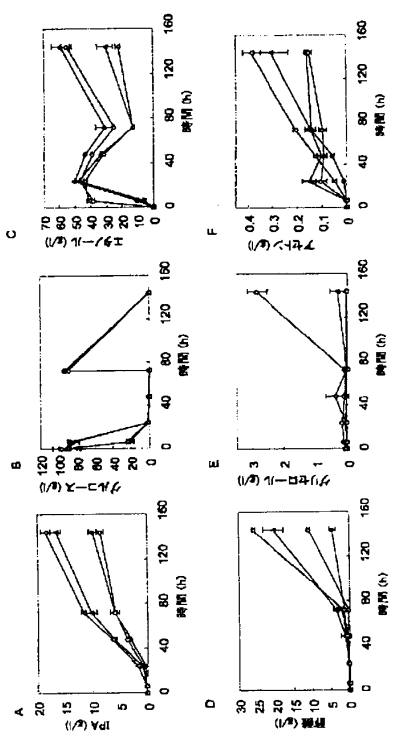
【図４】


TMS272
ACS1
ACS2
ADH2
ADH3
ALD1
ALD2
ACH1
CAT2
CRC1
YAT1
YAT2
ERG10
ACL
IPA相対生産量

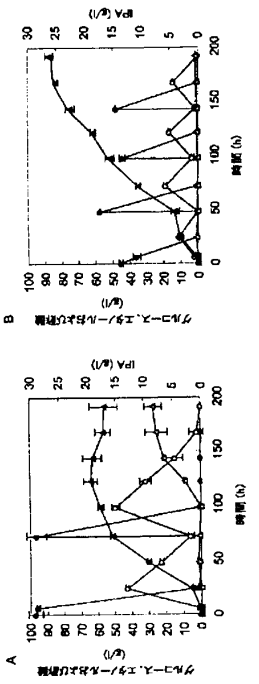
【図５】



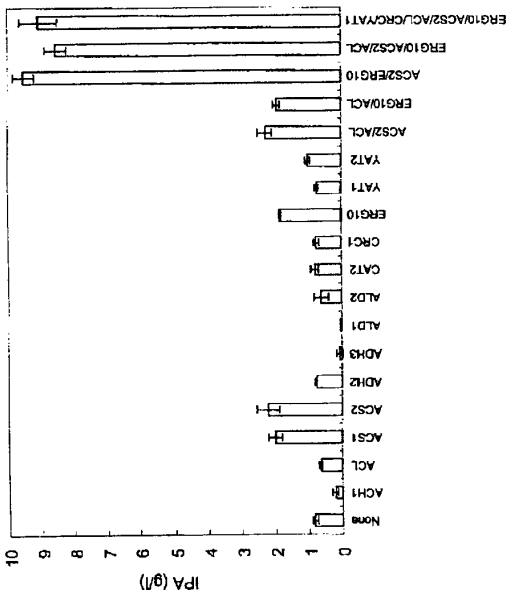
【図６】



【図７】



【図８】


IPA (g/l)
None
ACH1
ACL
ACS1
ACS2
ADH2
ADH3
ALD1
ALD2
CAT2
CRC1
ERG10
YAT1
YAT2
ACS2/ACL
ERG10/ACL
ACS2/ERG10
ERG10/ACS2/ACL
ERG10/ACS2/ACL/CRC/YAT1

【配列表】
2012254044000001.app

フロントページの続き

(72)発明者 生 嶋 茂 仁
東京都中央区新川二丁目１０番１号 キリンホールディングス株式会社内

Ｆターム(参考) 4B024 AA17 BA80 CA04 DA12 GA11
4B050 CC03 DD02 LL05
4B064 AC02 CA06 CA19 CC24 DA16
4B065 AA23Y AA73X AB01 AC14 BA02 CA05 CA54

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 475 of 3077

METHOD FOR PRODUCING SUBSTANCE USING METABOLIC PATHWAY GOING THROUGH ACETYL COA IN YEAST

ABSTRACT

PROBLEM TO BE SOLVED: To provide a yeast strain capable of highly producing metabolic products going through acetyl-CoA.SOLUTION: The yeast strain in which activity of the following enzymes: (a) an acetyl-CoA synthetase (ACS); and (b) an acetyl-CoA acetyltransferase (ERG10) are enhanced.

Pub. No.	JP2012254044A
App. No.	JP20110129527
App. No. Original	2011129527
Appl. Date	09-Jun-2011
Pub. Date	27-Dec-2012
Inventor(s)	TAMAGAWA HIDEYUKI IKUSHIMA SHIGEHITO
Assignee Norm.	KIRIN HOLDING CO LTD
Current Owner	MITSUBISHI GROUP
IPC	C12N1/19 , C12N9/10 , C12N15/09 , C12P7/04
CPC	Y02P20/52

CLAIMS

1. The following enzymes: (A) acetyl-CoA synthase (ACS); (b) and acetyl-CoA acetyltransferase (ERG10) Activity is enhanced, yeast strain.

2. The following genes: Gene encodes a polypeptide having an activity (a) acetyl-CoA synthase (ACS); expression of a gene encoding a polypeptide having an activity (b) acetyl-CoA acetyltransferase in (ERG10) and is enhanced The yeast strain according to any yl, in claim 1.

3. Polypeptide with ACS activity, The amino acid sequence represented by SEQ ID NO: 14 or (ii) SEQ ID NO: 12 or a polypeptide comprising the amino acid sequence represented by SEQ ID NO: 14, or (i) SEQ ID NO: 12, one or several amino acids are deleted, is a polypeptide having an ACS activity and comprises an amino acid sequence that has been inserted, a substituted or added, the yeast strain according to claim 2.

4. Is a polypeptide having ERG10 activity, Amino acids in the amino acid sequence represented by (ii) SEQ ID NO: 62 or a polypeptide comprising the amino acid sequence represented by (i) SEQ ID NO: 62, one or more amino acids is inserted deleted, substituted, added, or is a polypeptide having an ERG10 activity and comprises the sequence, the yeast strain according to claim 2 or 3.

5. In addition, the following genes: The gene encodes a polypeptide having an activity (c) ATP- citric acid removal enzyme addition (ACL); Selected from the group consisting of a gene encoding a polypeptide having an activity (e) carnitine acetyl transferase gene (YAT1) and; gene encoding a polypeptide having an activity (d) the inner mitochondrial membrane carnitine transporter (CRC1) Expression of one gene is increased at least a yeast strain as claimed in any one of claims 1 to 4.

6. It is intended a polypeptide having ACL activity, a polypeptide having ACLY activity and a polypeptide having the activity ACL1, Is a polypeptide having ACL1 activity, Amino acids in the amino acid sequence represented by (ii) SEQ ID NO: 32 or a polypeptide comprising the amino acid sequence represented by (i) SEQ ID NO: 32, one or more amino acids is inserted deleted, substituted, added, or It is a polypeptide having an activity ACL1 and comprises a sequence, Is a polypeptide having ACLY activity, Amino acids in the amino acid sequence represented by (ii) SEQ ID NO: 34 or a polypeptide

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 476 of 3077

comprising the amino acid sequence represented by (i) SEQ ID NO: 34, one or more amino acids is inserted deleted, substituted, added, or is a polypeptide having an ACLY activity and comprises the sequence, the yeast strain according to claim 5.

7. Is a polypeptide having CRC1 activity, Amino acids in the amino acid sequence represented by (ii) SEQ ID NO: 24 or a polypeptide comprising the amino acid sequence represented by (i) SEQ ID NO: 24, one or more amino acids is inserted deleted, substituted, added, or is a polypeptide having a CRC1 activity and comprises the sequence, the yeast strain according to claim 5 or 6.

8. Is a polypeptide having YAT1 activity, Amino acids in the amino acid sequence represented by (ii) SEQ ID NO: 28 or a polypeptide comprising the amino acid sequence represented by (i) SEQ ID NO: 28, one or more amino acids is inserted deleted, substituted, added, or is a polypeptide having a YAT1 activity and comprises the sequence, the yeast strain according to any one of claims 5 to 7.

9. A Crabtree negative yeast effect, the yeast strain according to any of claims 1 to 8.

10. Is Candida utilis, and a yeast strain as claimed in any one of claims 1 to 9.

11. Is one which has been transformed with genes for the synthesis of metabolites via the acetyl CoA, the yeast strain according to any one of claims 1 to 10.

12. 4 that encodes a polypeptide having the activity secondary alcohol dehydrogenase CoA transferase subunit A (ctfA), CoA transferase subunit B (ctfB), acetoacetic acid decarboxylase and (adc) of (sadh) The yeast strains according is to be transformed with the genes of the species to claim 11.

13. The method comprising the step of culturing the yeast strain according to any one of claims 1 to 12, to produce a material.

14. Is a metabolite through the acetyl CoA, the material A method according to claim 13.

15. Is isopropanol said substance, the method according to claim 13 or 14.

DESCRIPTION

Background of the invention

In yeast, TECHNICAL FIELD The present invention relates to a process for the production of a substance using the metabolic pathway via acetyl-CoA.

Against the background of environmental issues to the background art serious, biofuels have attracted attention as an alternative fuel for petroleum. Ethanol is well known as biofuels, it only has lower energy density than gasoline, it is necessary to be mixed with gasoline in a concentration range that is limited in order to function as a transportation fuel, as a fuel shortcomings in many cases. Also, having a higher octane number than the 1-butanol, isopropanol, a solvent having high versatility is also used as industrial raw materials. Isopropanol, which are fabricated by the hydration of propylene from naphtha, never been produced commercially from renewable sources.

Fermentative production of butanol and acetone by microorganisms, has been carried out industrially in the early 20th century. Can be produced at the same time the butanol and acetone Clostridium bacteria of some sort of strictly anaerobic bacteria (acetone-butanol fermentation). Acetone is produced as a raw material acetyl CoA present in the bacterial cell. Is converted to acetoacetyl-CoA by (thiolase) acetyl-CoA acetyltransferase, and are converted to acetoacetate by CoA transferase addition, acetyl-CoA is converted to acetone by acetoacetic acid decarboxylase enzyme. Clostridium certain bacteria has a dehydrogenase of secondary alcohols further, I want to produce isopropanol from acetone. Have reported tha

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 477 of 3077

in Clostridium bacteria 52 strains, including Clostridium Beijerinki, strains producing isopropanol 30mM of (1.8g / L) at the maximum exists Chen et al (1: Chen, J.-S., and SF Hiu Acetone-butanol-isopropanol production by Clostridium beijerinckii (synonym, Clostridium butylicum) Biotechnol Lett 1986. 8:.... 371-376). We have also introduced CoA transferase of Clostridium genus, acetoacetic acid decarboxylase, a secondary alcohol dehydrogenase in E. coli, Hanai et al. And Jojima colleagues, excess acetyl-CoA acetyltransferase of endogenous recently ... T Hanai, S. Atsumi, and JC Liao Engineered Synthetic Pathway for Isopropanol Production in Escherichia coli Appl Environ Microbiol,: 2 (non-patent literature have reported E. coli strain that can efficiently produce isopropanol by expressing ... Toru Jojima & Masayuki Inui & Hideaki Yukawa Production of isopropanol by metabolically engineered Escherichia coli Appl Microbiol Biotechnol 2008 77::; 2007, p 7814-7818 Dec. 3 non-patent literature 1219-1224). However, E. coli, various stress-resistant glucose tolerance, osmotic tolerance, salt tolerance, and organic solvent resistance is not very high in many cases, there is also a risk of phage contamination. Also, isopropanol resistance is low means that the inhibition of fermentation by product occurs. How to continue fermentation while removing from the medium the product by using, for example gas stripping process in order to avoid fermentation inhibitors are known, the use of the method not only to complication of the process of fermentation , leading to a cost increase. It can be considered as one means for solving these problems, and producing isopropanol using the microorganism having a high organic solvent resistance.

Yeast are unicellular eukaryotic microorganisms also showed a high organic solvent resistance, as well as the microorganisms such as E. coli and that can be cultured at low costs than the cells of higher organisms in terms of culture cost culture equipment, and culture time I have a feature. In particular, budding yeast is used in the production of bread and alcoholic beverages such as beer and wine, as the safety of the food has been confirmed.

In order to produce isopropanol in yeast, the yeast strain that is required to CoA transferases exogenous acetoacetate decarboxylase, and introducing a secondary alcohol dehydrogenase. Until now, the example of producing isopropanol by yeast, it was produced isopropanol by introducing the gene of the isopropyl alcohol-producing enzymes Saccharomyces cerevisiae to (Saccharomyces cerevisiae), and culturing the transformant has been reported (: Patent Publication No. 2011-97929 Patent Document 1). However, production of isopropanol by this method, only a small quantity of 100mg / L before and after the concentration in the culture medium. Yeast is a eukaryote, from that organelle is present, the supply route is more present in the cytoplasm acetyl-CoA is a substrate of isopropanol synthesis. There is no example of what yeast or is excellent in such capacity can be provided to the cytoplasm acetyl-CoA from which path or whether advantageous for isopropanol synthesis was investigated to date.

Candida utilis (Candida utilis), as well Saccharomyces cerevisiae and (Saccharomyces cerevisiae) Saccharomyces fragilis and (Saccharomyces fragilis), and has been certified as edible yeast FDA by (Federal Food and Drug Administration), as a food additive It is an edible yeast can be used safely permitted. In fact, even today, including the Japanese, it is produced in countries around the world such as the United States and Germany, Candida utilis, has been used as a food diet. In additior Candida utilis, has been used in a wide range of industries as producing strain bacterial production by deer pentose, ethyl acetate, L- glutamic acid, glutathione, and invertase. Hiroshi Shimada, Keiji Kondo, Paul D. Fraser, Yutaka: 4 (non-patent literature in recent years, the transformation method has been developed in Candida utilis, technologies various compounds are produced metabolically engineering has been developed .... Miura, Toshiko Saito, and Norihiko Misawa, Increased Carotenoid Production by the Food Yeast Candida utilis through Metabolic Engineering of the Isoprenoid Pathway, Appl Environ Microbiol, July 1998, p 2676-2680, Vol 64, 7 No.; non- Patent Document 5:. Yutaka Miura, Keiji Kondo, Toshiko Saito, Hiroshi Shimada, Paul D. Fraser, and Norihiko MisawaProduction of the Carotenoids Lycopene, β-Carotene, and Astaxanthin in the Food Yeast Candida utilis, Appl Environ Microbiol, April 1998, .. p 1226-1229, Vol 64, No. 4; 6 Non-Paten Document: Keiji Kondo,, Yutaka Miura, Hidetaka Sone, Kazuo Kobayashi & Hiroshi Iijima, High-level expression of a sweet protein, monellin, in the food yeast Candida utilis, 15, 453 Nature Biotechnology - 457 (1997); 7 Non-Patent Document:.. Yutaka Miura, Keiji Kondo, Hiroshi Shimada, Toshiko .. Saito, Katsumi Nakamura, Norihiko Misawa, Production of lycopene by the food yeast, Candida utilis that does not naturally synthesize carotenoid, Biotechnology and Bioengineering Volume 5(Issue 2-3, Pages 306-308; 8 Non-Patent Document: Shigehito IKUSHIMA, Toshio FUJII, Osamu KOBAYASHI, Satoshi

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 478 of 3077

YOSHIDA and Aruto YOSHIDA., Genetic Engineering of Candida utilis Yeast for Efficient Production of L-Lactic Acid., Bioscience, Biotechnology, and Biochemistry Vol. 73 (2009), No. 8 pp.1818-1824 ; Non-Patent Document 9:... Yi-Ren Hong, Ya-Lei Chen, Lynn Farh, Wen-Jen Yang, Chen-Hua Liao and David Shiuan, Recombinant Candida utilis for the production o biotin, Applied Microbiology and Biotechnology, Volume 71 (2006), Number 2, 211-221). In addition, Candida utilis, unlike Saccharomyces genus, from that Crabtree is (Crabtree) effect negative yeast, respiratory activity is not reduced even in the presence of glucose. Thus, by adopting the conditions for ventilating enough, it is possible to suppress generation of ethano Moreover, respiratory activity high Candida utilis, since the ethanol produced is also once again rapidly assimilated, thereby enabling the production of metabolites and cell efficiency.

When producing materials such as isopropanol in yeast, generally, generation of ethanol as a by-product becomes a problem. Patent document; No. 2010/095751 WO 2 (patent document in order to suppress the ethanol by-product, efforts to deficient (PDC) pyruvate decarboxylase gene is a key enzyme in the ethanol production have been made 3:..; 2010/095750 No. WO 10 non-patent literature: Flikweert MT, Van Der Zanden L, Janssen WM, et al Pyruvate decarboxylase: an indispensable enzyme for growth of Saccharomyces cerevisiae on glucose Yeast 12 (1996): ..; 247-257 Non-Patent Document 11: van Maris AJA, Winkler AA, Porro D, et al Homofermentative lactate production can not sustain anaerobic growth of engineered Saccharomyces cerevisiae: Possible consequence of energy-dependent lactate export Appl Environ Microbiol 70 (2004) : 2898-2905).

Chen, J.-S., and SF Hiu Acetone-butanol-isopropanol production by Clostridium beijerinckii (synonym, Clostridium butylicum Biotechnol Lett 1986. 8:.... 371-376 T. Hanai, S. Atsumi, and JC Liao. Engineered Synthetic Pathway for Isopropanol Production in Escherichia coli. Appl Environ Microbiol, Dec. 2007, p. 7814-7818 .. Toru Jojima & Masayuki Inui & Hideaki Yukawa Production of isopropanol by metabolically engineered Escherichia coli Appl Microbiol Biotechnol 2008 77: 1219-1224 Hiroshi Shimada, Keiji Kondo, Paul D. Fraser, Yutaka Miura, Toshiko Saito, and Norihiko Misawa., Increased Carotenoid Production by the Food Yeast Candida utilis through Metabolic Engineering of the Isoprenoid Pathway., Appl Environ Microbiol, July 1998, p. 2676 -2680, Vol. 64, No. 7 Yutaka Miura, Keiji Kondo, Toshiko Saito, Hiroshi Shimada, Paul D. Fraser, and Norihiko MisawaProduction of the Carotenoids Lycopene, β-Carotene, and Astaxanthin in the Food Yeast Candida utilis., Appl Environ Microbiol, April 1998, p. 1226- 1229, Vol. 64, No. 4 .. Keiji Kondo ,, Yutaka Miura, Hidetaka Sone, Kazuo Kobayashi & Hiroshi Iijima, High-level expression of a sweet protein, monellin, in the food yeast Candida utilis, Nature Biotechnology 15, 453 - 457 (1997) Yutaka Miura, Keiji Kondo, Hiroshi Shimada, Toshiko Saito, Katsumi Nakamura, Norihiko Misawa., Production of lycopene by the food yeast, Candida utilis that does not naturally synthesize carotenoid., Biotechnology and Bioengineering Volume 58 Issue 2-3, Pages 306- 308 Shigehito IKUSHIMA, Toshio FUJII, Osamu KOBAYASHI, Satoshi YOSHIDA and Aruto YOSHIDA., Genetic Engineering of Candida utilis Yeast for Efficient Production c L-Lactic Acid., Bioscience, Biotechnology, and Biochemistry Vol. 73 (2009), No. 8 pp. 1818-1824 Yi-Ren Hong, Ya-Lei Chen Lynn Farh, Wen-Jen Yang, Chen-Hua Liao and David Shiuan., Recombinant Candida utilis for the production of biotin., Applied Microbiology and Biotechnology., Volume 71 (2006), Number 2, 211-221 Flikweert MT, Van Der Zanden L, Janssen WM, et al Pyruvate decarboxylase:.. An indispensable enzyme for growth of Saccharomyces cerevisiae on glucose Yeast 1: (1996): 247-257 van Maris AJA, Winkler AA, Porro D, et al Homofermentative lactate production can not sustain anaerobic growth of engineered Saccharomyces cerevisiae:.. Possible consequence of energy-dependent lactate export Appl Environ Microbiol 70 (2004): 2898-2905 Japanese Patent Publication No. 2011-97929 No. 2010/095751 WO No. 2010/095750 WO

By enhancing the activity acetyl CoA acetyl transferase enzyme acetyl-CoA synthase in yeast (ACS) and of (ERG10), the present inventors have found that, to a significant increase in the production volume of the substance through the acetyl-Co, in yeast I have found that. The present invention is based on this finding.

Thus, yeast strains enzymatic activity according to the synthesis and metabolism of acetoacetyl-CoA and acetyl-CoA are involved as intermediates in the fermentation production of various substances is enhanced, an object of the present invention, the materials, as well as with yeast strain It is another object to provide a manufacturing process.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 479 of 3077

The yeast strain according to the present invention, the following enzymes:
(A) acetyl-CoA synthase (ACS); (b) and acetyl-CoA acetyltransferase (ERG10)
It is a yeast strain activity is enhanced.

Furthermore, the production method according to the invention is a method comprising the step of culturing a yeast strain according to the present invention, to produce the material.

In accordance with the present invention, the metabolic pathway of yeast acetoacetyl CoA and acetyl-CoA is involved is activated. Because of the wide variety of metabolites to generate also there is a wide variety of metabolic pathways which acetoacetyl-CoA and acetyl-CoA. Therefore, according to the present invention, it is possible to manufacture with high efficiency various substances by culturing the yeast strain. Also, production of the material by depending on the material to be produced, is incorporated into the yeast in an expressible form genes required enzyme, and culturing the yeast strain obtained is also possible.

DETAILED DESCRIPTION OF THE INVENTION

The yeast strain according to the invention, the following genes: a gene encoding a polypeptide having an activity (a) acetyl-CoA synthase (ACS); poly having an activity (b) acetyl-CoA acetyltransferase in (ERG10) and Expression of the gene encoding the peptides is enhanced.

As used herein, acetyl-CoA synthase and the above (ACS) gene acetyl-CoA acetyltransferase with (ERG10) gene, to prepare a yeast strain overexpressing the four genes for the synthesis of isopropanol, culturing this By the, the high concentration of isopropanol are produced very has been demonstrated. Isopropanol production pathway in the yeast strain is as shown in Figure 1. From path shown in Figure 1, a high production volume isopropanol, due to the enhancement of enzyme activity relating to synthesis and metabolism of acetoacetyl-CoA and acetyl-CoA is clear. Therefore, the present invention allows to produce a high various materials not only isopropanol, via acetoacetyl-CoA and acetyl-CoA. Also, as shown in Figure 1, ERG10 and ACS are enzymes that function in the cytoplasm.

Acetyl-CoA supplied to the cytoplasm, a core material compound having a flavonoid skeleton, compounds having a terpene terpenoid skeleton, a compound having a sterol backbone, such as fatty acids, secondary metabolites for various utilities. Thus, as long as the synthetic route of the secondary metabolites are in a state of being genetically modified artificially or native state yeast strains, it can be highly productive of these substances by the method of the present invention .

When in general, to perform the microbial production of substances, when suppressing effort is made a by-product production in many cases. For example, when producing the substance in yeast, In general, in order to suppress the formation of ethanol as a by-product, the enzyme genes involved in ethanol production, such efforts to be deficient in pyruvate decarboxylase gene (PDC) have been made .. 2010/095750 No. WO; No. 2010/095751 pamphlet (International Publication Flikweert MT, Van Der Zanden L, Janssen WM, et al Pyruvate decarboxylase: an indispensable enzyme for growth of Saccharomyces cerevisiae on glucose Yeast 12 (1996):.. 247-257; van Maris AJA, Winkler AA, Porro D, et al Homofermentative lactate production can not sustain anaerobic growth of engineered Saccharomyces cerevisiae: Possible consequence of energy-dependent lactate export Appl Environ Microbiol 70 (2004): 2898 -2905). However, herein, than the veast strain was deficient in the PDC described above. the direction of a veast strain did not lacking PDC indicates isopropanol production significantly higher have been demonstrated. Thus, the yeast strain according to the invention is a

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 480 of 3077

feature also is maintained without being deleted gene involved in ethanol production, such as PDC, isopropanol production, or its acetyl CoA in turn a result that the production volume of acetoacetyl-CoA is a metabolite can be improved remarkably, is surprising.

In this specification the enzyme activity being "increased" means that at one time at least in the course of culturing, the stronger the enzyme activity is observed compared with yeast which is not subjected to means for enhanced activity to. Enhancement of enzyme activity, can be implemented by a replacement of a higher enzyme specific activity of the gene, by modification of the enzyme gene by way engineering or protein that enhances expression of the enzyme gene. As used herein the expression of the gene being "enhanced", the expression level of that gene in yeast is increased by genetic modification of the yeast, that is, it is meant that overexpression of the gene. The method of genetic modification include, for example, (that is, a method for transfecting yeast in a form capable of expressing the gene) Methods for transforming yeast with the gene, gene present in the genome of the yeast method for enhancing the promoter activity, and a method of enhancing expression of the gene by replacing the promoter with other promoters of genes present in the genome of the yeast, and various methods are known, but preferably It is a method to transform yeast by the gene.

The yeast strain according to the invention the yeast is a yeast which has been genetically altered genes specific to overexpress. Well, yeast used for genetic modification is not particularly limited as long as it is a yeast having a metabolic pathway through the acetoacetyl-CoA or acetyl-CoA, the clubs and more preferably, is a Crabtree effect negative yeast with high respiratory activity, preferably it is a Candida utilis is a Tree negative yeast. Various strains of well-known, for example, NBRC0396 shares, NBRC0619 shares, NBRC0988 shares, NBRC0639 shares, NBRC0626 shares, NBRC1086 shares, NBRC10707 shares, ATCC9255 shares, ATCC28955 strain, ATCC44638 strains, strains of Candida utilis ATCC9206 in the art strain, ATCC96621 strain, ATCC18201 shares, and may be a ATCC15239 shares, etc., but is preferably NBRC1086 share, and ATCC15239 shares or ATCC9255 shares.

The following enzymes enzyme activity is enhanced in yeast strain caused by the gene present invention and the enzyme activity is enhanced,: (a) acetyl-CoA synthase (ACS); (b) and acetyl-CoA acetyltransferase (ERG10 is). In accordance with a preferred embodiment of the present invention, genes of these two enzymes are over-expressed in a yeast strain according to the present invention, that is, the following genes: poly have activity (a) acetyl-CoA synthase (ACS) gene encoding the peptide; expression of a gene encoding a polypeptide having an activity (b) acetyl-CoA acetyltransferase in (ERG10) and has enhanced.

The yeast strain according to the present invention, further,: a gene encoding a polypeptide having the activity (c) ATP- citric acid removal additional enzymes (ACL); (d) the inner mitochondrial membrane carnitine transporter and of (CRC1) Expression of one gene may have been increased at least is selected from the group consisting of a gene encoding a polypeptide having an activity (e) carnitine acetyl transferase gene (YAT1) and; gene encoding a polypeptide having an activity may be, are (c), (d) and the expression of all genes (e) is preferably increased.

Those derived from a variety of yeast are known genes and these enzymes, its origin is not particularly limited. Or may be the expression product and the endogenous genes of the yeast used in breeding, gene and enzyme may be a unit derived from yeast other than yeast used for breeding. For example, genes and enzymes these may be the one independently from Yarrowia lipolytica or Candida utilis, respectively. As to the origin of each enzyme, for example, acetyl-CoA synthase (ACS) Candida utilis yeast, acetyl-CoA acetyl transferase enzyme (ERG10) yeast Candida utilis, mitochondrial inner membrane

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 481 of 3077

carnitine transporter (CRC1) is ACLY and ACL1 of (Yarrowia lipolytica derived constitute a subunit Yarrowia lipolytica yeast (ACL) the yeast Candida utilis, ATP- citric acid removal adding enzyme Candida utilis yeast, carnitine acetyltransferase gen (YAT1) gene) that is preferred.

Acetyl-CoA synthase derived from Candida utilis to (ACS) gene, there is a gene ACS2 and ACS1 gene. Coding sequence o ACS1 gene is represented by SEQ ID NO: 11, the amino acid sequence encoded is shown in SEQ ID NO: 12. Coding sequence of ACS2 gene is represented by SEQ ID NO: 13, the amino acid sequence encoded is shown in SEQ ID NO: 14. Acetyl-CoA synthetase (ACS) is an enzyme that converts acetyl-CoA and acetate. May be any of ACS2 and ACS1 The ACS gene used in the present invention, but I are preferably ACS2. In addition, a polypeptide having the activity acetyl-CoA synthase (ACS) is an amino acid in the amino acid sequence represented by SEQ ID NO: 14 or SEQ ID NO: 12, one or more amino acids is inserted deleted, substituted, added, or It may be a polypeptide having an ACS activity and comprises a sequence.

Coding sequence acetyl CoA acetyl transferase enzyme derived from Candida utilis of (ERG10) gene is represented by SEQ ID NO: 61, the amino acid sequence encoded is represented by SEQ ID NO: 62. Acetyl-CoA acetyltransferase (ERG10) is an enzyme that converts acetoacetyl-CoA to acetyl-CoA. In addition, the amino acid sequence represented by SEQ ID NO: 62, a polypeptide having the activity acetyl-CoA acetyltransferase in (ERG10) is an amino acid sequence that is inserted deleted, substituted, added, or one or several amino acids It may be a polypeptide having an activity ERG10 and containing

The coding sequence of the gene ACL1 Yarrowia lipolytica derived is represented by SEQ ID NO: 31, the amino acid sequence encoded is shown in SEQ ID NO: 32. Also, in the amino acid sequence represented by SEQ ID NO: 32, a polypeptide having the activity of ACL1 poly having a ACL1 activity and comprising an amino acid sequence inserted deleted, substituted, added, or one or several amino acids may be a peptide. Moreover, the coding sequence of the gene ACLY Yarrowia lipolytica derived is represented by SEQ ID NO: 33, the amino acid sequence encoded is shown in SEQ ID NO: 34. In addition, in the amino acid sequence represented by SEQ ID NO: 34, a polypeptide having the activity of ACLY poly having a ACLY activity and comprising an amino acid sequence that has been inserted deletion, substitution, or addition of one or several amino acid residues may be a peptide. And subunit ACLY ACL1 and these ATP- citric acid removal adding enzyme (ACL) is an enzyme that converts acetyl CoA and citric acid.

The coding sequence mitochondrial inner membrane carnitine transporter derived from Candida utilis of (CRC1) gene is represented by SEQ ID NO: 23, the amino acid sequence encoded is shown in SEQ ID NO: 24. Mitochondrial inner membrane carnitine transporter (CRC1) is an enzyme that transport to the cytoplasm from mitochondria acetyl carnitine. In addition, in the amino acid sequence represented by SEQ ID NO: 24, a polypeptide having the activity mitochondrial inner membrane carnitine transporter (CRC1) is an amino acid sequence that is inserted deletion, substitution, or addition of one or several amino acid residues It may be a polypeptide having an activity CRC1 and containing.

Coding sequence carnitine acetyl transferase gene derived from Candida utilis of (YAT1) gene is represented by SEQ ID NO: 27, the amino acid sequence encoded is represented by SEQ ID NO: 28. Carnitine acetyl transferase gene (YAT1) is an enzyme that converts acetyl-CoA acetyl carnitine. In addition, the amino acid sequence represented by SEQ ID NO: 28, a polypeptide having the activity carnitine acetyl transferase gene (YAT1) comprises an amino acid sequence inserted deleted substituted, added, or one or several amino acids , and may be a polypeptide having an activity YAT1.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 482 of 3077

Amino acid deletions, substitutions, additions, or insertions can be done by modifying the techniques known in the art, a gene encoding the polypeptide. It is possible to introduce a mutation in a gene is carried out by a method or an equivalent known method Gapped duplex method, or the Kunkel method, for example mutagenesis kit that utilizes site-directed mutagenesis, Mutant-K or alternatively, using Mutant-G and (Takara Bio Inc.), and the like (Takara Bio Inc.), to introduce the mutation using LA PCR in vitro Mutagenesis series kit of Takara Bio Inc., KOD-Plus-Mutagenesis Kit and (TOYOBO) I can be. Activity of the enzyme described above can be confirmed by procedures known in the art.

Can also be substituted for the above genes, the use of equivalents of each. On condition that it has a function equivalent tc each gene, the equivalents, refers to a gene nucleotide residues of some different. The equivalents of such 70% or more an the respective nucleotide sequence, preferably 80% or more, more preferably 85% or more, more preferably 90% or more, homology of 95% or more and most preferably (sequence identity genes comprising a nucleotide sequence encoding a polypeptide having the activity of each enzyme, and the like) and is there. Furthermore, a gene comprising a nucleotide sequence encoding a polypeptide having the activity of each enzyme, and hybridized under stringent conditions with a complementary sequence thereof, or the nucleotide sequence of each can be mentioned as the equivalent. Furthermore, th nucleotide sequence encoding a polypeptide having an activity of each enzyme and comprises a sequence that has been inserted deletions, nucleotide residues of one or several substitution, addition, or in the nucleotide sequence of each as the equivalent gene, including the like.

Here, nucleotide deletions, substitutions, additions, or insertions can be done by modifying the techniques known in the art, genes containing the sequence. It is possible to introduce a mutation in a gene is carried out by a method or an equivalent known method Gapped duplex method, or the Kunkel method, for example, may be used mutagenesis kit that utilizes site-directed mutagenesis. For example, or, using Mutant-K Mutant-G and (Takara Bio Inc.), and the like (Takara Bio Inc.), using LA PCR in vitro Mutagenesis series kit of Takara Bio Inc., KOD-Plus-Mutagenesis Kit and (TOYOBO) It is possible to introduce mutations. Activity of each enzyme can be confirmed by procedures known in the art.

A number that indicates the (sequence identity) homology (percent), nucleotide sequence comparison program: are those using GENETYX-WIN7.0.0 for example, is calculated by the default parameters (initial setting). That is, each gene in the yeast chromosome, not the same, but functionally equivalent, that is optionally substituted via homologous recombination such as the gene encoding a polypeptide having the activity. Activity of each enzyme can be confirmed by procedures know in the art.

The stringent conditions, for example, using Rapid-Hyb Buffer (GE Healthcare Bio-Science Co., Ltd.), 40~70 ℃ preferably, as 60 ℃ and more preferably, high performed according to the protocol attached to other temperature conditions is the hybridization conditions. Then, for example, using common methods known to those skilled in the art, washing for 5 minutes with a solution consists of (w / v) SDS 0.1% and 2 × SSC, followed by 0.1% and 1 × SSC (refers to do a 10 minute wash with a solution consist w / v) SDS, washed for 10 minutes with a solution consists of (w / v) SDS 0.1% and 0.1 × SSC furthe However by appropriately setting the conditions of salt concentration of the solution used and the temperature condition during hybridization and washing of the membrane then one (70%, 80%, 85%, 90%, 95% certain I can be cloned DNA containing a nucleotide sequence having homology) or more. Gene obtained in this manner is not identical to sequence on but functionally equivalent, that is optionally substituted via homologous recombination such as the gene encoding a polypeptide having the activity. Activity of each enzyme can be confirmed by procedures known in the art.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 483 of 3077

For use in the production of useful substances desired, the yeast strain by gene present invention for the synthesis of useful substances may be transformed with a gene required for production of the substance. For example, in order to produce the metabolites via the acetyl-CoA by the method of the present invention, the yeast strain according to the invention may be transformed with genes for the synthesis of metabolites via acetyl CoA. Necessary genes, may be selected depending on the material to be produced, and appropriately selected by those skilled in the art.

For example, in the case to be produced and isopropanol, as an additional genetic alterations, transforming a yeast strain with isopropanol synthetic gene. The isopropanol synthetic gene, it is possible and enzyme genes described uses a gene known to those skilled in the JP 2011-97929 discloses, for example, CoA sulfotransferase subunit A (ctfA), the subunit B It is possible to use a combination of four genes secondary alcohol dehydrogenase (ctfB), acetoacetic acid decarboxylase and (adc) of (sadh).

According to one embodiment of the present invention, as a further genetic modification, CoA sulfotransferase subunit A (ctfA), CoA transferase subunit B (ctfB), the yeast strain according to the invention, acetoacetic acid decarboxylase it has been transformed by the four genes that encode a polypeptide having the activity secondary alcohol dehydrogenase and (adc) of (sadh). It can be used in the manufacture of isopropanol yeast strains such, this method of manufacture, comprising culturing the yeast strain according to the present embodiment.

Those derived from various organisms are known to these genes, their origin is not particularly limited. For example, these genes can be assumed that the independently derived from Clostridium Beijerinki or Clostridium acetobutylicum, respectively. As to the origin of each enzyme, for example, preferably Clostridium acetobutylicum is (ctfB) same subunit B CoA transferase gene subunit A and (ctfA), and secondary alcohol dehydrogenase acetoacetic acid decarboxylase and (adc Clostridium Beijerinki is preferred (sadh).

The amino acid sequence CoA transferase subunit A of Clostridium acetobutylicum derived (ctfA) is represented by SEQ ID NO: 2. CoA transferase subunit A (ctfA) is an enzyme that together with (ctfB) same subunit B, is converted to acetoacetate to acetoacetyl-CoA. It is possible such arrangement of the gene of Clostridium acetobutylicum from, appropriately selected by those skilled in the art, the gene sequence encoding the amino acid sequence represented by SEQ ID NO: 2 is not particularly limited, according to the invention preferably It is the sequences using codons frequently in species of the yeast strain. For example, gene sequences that conform to the codon frequency in Candida utilis, nucleotide sequence represented by SEQ ID NO: 1. In the amino acid sequence represented by SEQ ID NO: 2, a polypeptide having the activity CoA transferase subunit A of (ctfA) comprises an amino acid sequence inserted deleted, substituted, added, or one or several amino acid residues, and may be a polypeptide having an activity ctfA.

The amino acid sequence CoA transferase subunit B of Clostridium acetobutylicum derived (ctfB) is represented by SEQ ID NO: 4. CoA transferase subunit B (ctfB) is an enzyme that together with (ctfA) same subunit A, is converted to acetoacetate to acetoacetyl-CoA. It is possible such arrangement of the gene of Clostridium acetobutylicum from, appropriately selected by those skilled in the art, the gene sequence encoding the amino acid sequence represented by SEQ ID NO: 4 is not particularly limited, according to the invention preferably It is the sequences using codons frequently in species of the yeast strain. For example, gene sequences that conform to the codon frequency in Candida utilis, nucleotide sequence represented by SEQ ID NO: 3. In the amino acid sequence represented by SEQ ID NO: 4, a polypeptide having the activity CoA transferase subunit B of (ctfB) comprises an amino acid sequence inserted deleted, substituted, added, or one or several amino acid residues, and may be a polypeptide having an activity ctfB.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 484 of 3077

The amino acid sequence acetoacetic acid decarboxylase of Clostridium Beijerinki derived (adc) is represented by SEQ ID NO: 6. Acetoacetic acid decarboxylase (adc) is an enzyme that converts acetone and acetoacetic acid. It is possible such arrangement of the gene of Clostridium Beijerinki from, appropriately selected by those skilled in the art, the gene sequence encoding the amino acid sequence represented by SEQ ID NO: 6 is not particularly limited, according to the invention preferably It is the sequences using codons frequently in species of the yeast strain. For example, gene sequences that conform to the codon frequency in Candida utilis, nucleotide sequence represented by SEQ ID NO: 5, and the like. In the amino acid sequence represented by SEQ ID NO: 6, a polypeptide having the activity acetoacetic acid decarboxylase (adc) is and comprises the amino acid sequence inserted deleted, substituted, added, or one or several amino acids It may be a polypeptide having an activity adc.

The amino acid sequence secondary alcohol dehydrogenase of Clostridium Beijerinki derived (sadh) is represented by SEQ ID NO: 8. Secondary alcohol dehydrogenase (sadh) is an enzyme that converts isopropanol and acetone. It is possible such arrangement of the gene of Clostridium Beijerinki from, appropriately selected by those skilled in the art, the gene sequence encoding the amino acid sequence represented by SEQ ID NO: 8 is not particularly limited, according to the invention preferably It is the sequences using codons frequently in species of the yeast strain. For example, gene sequences that conform to the codon frequency in Candida utilis, nucleotide sequence represented by SEQ ID NO: 7, and the like. In the amino acid sequence represented by SEQ ID NO: 8, a polypeptide having an active secondary alcohol dehydrogenase (sadh) comprises an amino acid sequence inserted deletion, substitution, or addition of one or several amino acid residues and may be a polypeptide having an activity sadh.

Amino acid deletions, substitutions, additions, or insertions can be done by modifying the techniques known in the art, a gene encoding the polypeptide. It is possible to introduce a mutation in a gene is carried out by a method or an equivalent known method Gapped duplex method, or the Kunkel method, for example mutagenesis kit that utilizes site-directed mutagenesis, Mutant-K or alternatively, using Mutant-G and (Takara Bio Inc.), and the like (Takara Bio Inc.), to introduce the mutation using LA PCR in vitro Mutagenesis series kit of Takara Bio Inc., KOD-Plus-Mutagenesis Kit and (TOYOBO) I can be. Activity of the enzyme described above can be confirmed by procedures known in the art.

Can also be substituted for the above genes, the use of equivalents of each. On condition that it has a function equivalent to each gene, the equivalents, refers to a gene nucleotide residues of some different. The equivalents of such 70% or more an the respective nucleotide sequence, preferably 80% or more, more preferably 85% or more, more preferably 90% or more, homology of 95% or more and most preferably (sequence identity genes comprising a nucleotide sequence encoding a polypeptide having the activity of each enzyme, and the like) and is there. Furthermore, a gene comprising a nucleotide sequence encoding a polypeptide having the activity of each enzyme, and hybridized under stringent conditions with a complementary sequence thereof, or the nucleotide sequence of each can be mentioned as the equivalent. Furthermore, the nucleotide sequence encoding a polypeptide having an activity of each enzyme and comprises a sequence that has been inserted deletions, nucleotide residues of one or several substitution, addition, or in the nucleotide sequence of each as the equivalent gene, including the like.

Here, nucleotide deletions, substitutions, additions, or insertions can be done by modifying the techniques known in the art, genes containing the sequence. It is possible to introduce a mutation in a gene is carried out by a method or an equivalent known method Gapped duplex method, or the Kunkel method, for example, may be used mutagenesis kit that utilizes site-directed mutagenesis. For example, or, using Mutant-K Mutant-G and (Takara Bio Inc.), and the like (Takara Bio Inc.), using LA PCR in vitro Mutagenesis series kit of Takara Bio Inc., KOD-Plus-Mutagenesis Kit and (TOYOBO) It is possible to introduce mutations. Activity of each enzyme can be confirmed by procedures known in the art.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 485 of 3077

A number that indicates the (sequence identity) homology (percent), nucleotide sequence comparison program: are those using GENETYX-WIN7.0.0 for example, is calculated by the default parameters (initial setting). That is, each gene in the yeast chromosome, not the same, but functionally equivalent, that is optionally substituted via homologous recombination such as the gene encoding a polypeptide having the activity. Activity of each enzyme can be confirmed by procedures know in the art.

The stringent conditions, for example, using Rapid-Hyb Buffer (GE Healthcare Bio-Science Co., Ltd.), 40~70 ℃ preferably, as 60 ℃ and more preferably, high performed according to the protocol attached to other temperature conditions is the hybridization conditions. Then, for example, using common methods known to those skilled in the art, washing for 5 minutes with a solution consists of (w / v) SDS 0.1% and 2 × SSC, followed by 0.1% and 1 × SSC (refers to do a 10 minute wash with a solution consist w / v) SDS, washed for 10 minutes with a solution consists of (w / v) SDS 0.1% and 0.1 × SSC furthe However by appropriately setting the conditions of salt concentration of the solution used and the temperature condition during hybridization and washing of the membrane then one (70%, 80%, 85%, 90%, 95% certain I can be cloned DNA containing a nucleotide sequence having homology) or more. Gene obtained in this manner is not identical to sequence on but functionally equivalent, that is optionally substituted via homologous recombination such as the gene encoding a polypeptide having the activity. Activity of each enzyme can be confirmed by procedures known in the art.

Also isopropanol addition, it is possible to construct a yeast strain producing substances such as sterols particular produced a particular carotenoid, such as lycopene, squalene or squalene. For example, Candida utilis that express GGPP synthase gene of Erwinia uredovora origin, phytoene synthase gene, phytoene desaturase gene, overexpressing HMG-CoA reductas gene of endogenously (HMG1) high lycopene (Shimada H, Kondo K, Fraser PD, Miura Y, Saito T, Misawa N. Increased carotenoid production by the food yeast Candida utilis through metabolic engineering of the isoprenoid pathway. Appl Environ Microbiol. 1998 Jul which can accumulate has been reported ; 64 (7): 2676-80). Also, ergosta and Zaimosuteroru-24-methyl transferase gene 5,7,24, (28) - Saccharomyces cerevisiae that has a defect in the expression of trienol 22-dehydrogenase gene, has been operated HMG1 is to overexpress recombinant strain, (5-192184 JP) that can accumulate a high concentration of squalene are known. By using these techniques, it is possible to construct a yeast strain that produces lycopene, and squalene.

Applying the yeast strains genetically modified to overexpress the genes described above, for example, molecular breeding of yeast strains by molecular breeding invention yeast strains, yeast strains in an expressible state the gene clusters above can be done by introducing. The genes are, be incorporated into the chromosome of the yeast by the technique of homologous recombination or the like. It is well known in the art, the choice of homologous recombination for gene sequences to achieve integration into the chromosome such, and selecting the appropriate homologous recombination for gene sequence as required by those skilled in the art so that it is possible to construct a homologous recombination DNA fragment for Te.

The chromosomal location of the genes described above is integrated into the yeast genome is not particularly limited, for example, orotidine 5 'phosphate decarboxylase gene (URA3) locus, alpha amino adipic acid reductase gene (LYS2) seat ca be chosen such. In addition, these genes, promoters originally present in the locus incorporated, for example, may be place under the control of the URA3 gene promoter, such as LYS2 gene promoter, glyceraldehyde-3-phosphate dehydrogenase (may be placed under the control of other promoters such as GAP) gene promoter.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 486 of 3077

In accordance with a preferred embodiment of the present invention, the yeast used in breeding the yeast strain according t the invention is a Candida utilis. Its high ploidy, Candida utilis, does not form spores. If it is desired to introduce a mutation ir a gene of the strain with high ploidy, it is necessary compared to strain haploid, added to severe variations thereof but, in doing so, also the gene is not a gene that want to give the mutant I considered the possibility to be added mutation will increase. Therefore, in the case of introducing a mutation in a gene of Candida utilis, using a technique capable of applying mutation in multiple efficiently only the target gene is preferred. The transformation method of Candida utilis, technology include those described in JP Patent Publication No. 2003-144185. In this document, that the available vectors, and sequences homologous to the chromosomal DNA of Candida utilis, and comprises a selectable marker gene, is integrated into the chromosomal DNA of Candida utilis heterologous gene by homologous recombination or which can be, a DNA sequence having an autonomous replication in Candida utilis, which comprise a selection marker gene can transform Candida utilis at high frequency have been developed.

A selection marker gene available in the transformation system, drug resistance markers that can function in yeast, etc. gen conferring hygromycin B resistance gene or confers cycloheximide resistance L41 gene, geneticin (G418) resistance preferably I have a. Gene conferring hygromycin B resistance gene or confer (G418) resistant geneticin is considered because it is a sequence not present in the yeast wild probability which is incorporated locus of the target is high. In additior drug resistance genes, these are considered impact on the traits of the host also small: (Cordero Otero R, et al, (Baganz F e al, 13 (16). 1563-73, 1997) Appl Microbiol Biotechnol, 46 (2):. 143-8, 1996).

As transformation systems other, Cre-loxP system from bacteriophage P1, and the like. This is a site-specific recombination system in the loxP sequences of 34bp between the two, the recombination is catalysed by Cre recombinase Cre gene that encodes. The system can also function in yeast cells such as Saccharomyces cerevisiae has been reported, the selectable marker gene placed between the two loxP sites are known be removed by recombination of the loxP sequence between (Guldener, U., et al, Nucleic Acids Res., 24,2519-24., 1996) that it is. This system (Steensma, H.Y., Et al, Yeast, 18,469-72., 2001) it is being used Kluyveromyces lactis and (Kluyveromyces lactis), in multiple yeast species.

By culturing in the presence of a suitable carbon source and yeast strains by the production method of the present invention substance can be prepared (eg isopropanol) substances in the culture. According to the production method (eg isopropanol) substances according to the invention, by performing the step of separating (eg isopropanol) material from the culture system, it is possible to obtain (eg isopropanol) material. It should be noted that the culture in the present invention, in addition to the culture supernatant, I have included the debris of bacteria bacterial cells or cultured cells, or cells.

Upon culturing the yeast strain according to the present invention, it is possible to choose the culture conditions and culture method for different types of yeast. As a culture method, for example, can be employed to form the culture can be cited liqui culture method using a test tube, a jar fermenter or flask, culture batch, such as fed-batch culture. Aeration and shaking speed is a ventilation amount corresponding to this shaking speed 50~200rpm in general, or 80~150rpm, preferably. Is 0.1 t 30 in general, OD_{600} of the cell amount of the initial, and preferably 0.5 to 10. In addition, when the culture of the yeast strain according to the present invention, the faster is (preferably the greater the amount of ventilation that is) preferably shaking speed, the higher is preferable that the initial OD_{600}. For example, in a culture in a test tube or flask, shaking it speeds greater than or equal to about 120rpm, initial OD_{600} is about 10 is preferred.

In the production process (such as isopropanol) substances according to the invention, the yeast is grown, well, the composition of the medium is not particularly limited as long as it is a composition containing various nutrients can be produced and the desired material. The assimilable carbon source contained in the culture medium, for example, may be

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 487 of 3077

used as long assimilated in sucrose and xylose In addition to glucose. In accordance with a preferred embodiment of the present invention, the carbon source comprises glucose. Glucose concentration in the medium, can be confirmed while the production efficiency of the substance of interest is adjusted appropriately by those skilled in the art. In particular, the culture of the yeast strain according to the invention, glucose concentration in the medium, keeping the relatively low concentration is preferred. For example, in the culture of 192 hours, and 100g / L and the initial concentration of glucose, as compared wit the case where glucose was added solid to a final concentration of 100g / L after 72 hours culture, and 50g / L and the initia concentration of glucose , 48 hours later, those obtained when glucose was added to the solid to a final concentration of 50g / L after 144 hours and after 96 hours is higher production efficiency of the material.

The nutrient source contained in the medium, for example, yeast extract, peptone, and whey is used, but media supplemented with assimilation of carbon sources mentioned above (10g / L yeast extract, 20g / L peptone) in YP, for example YPD5 in medium (50g / L glucose, 10g / L yeast extract, 20g / L peptone) such, those appropriate pH adjustment is more convenient.

The fermentation temperature can be chosen in a range of viable yeast to be used. For example, it can be about 15 ℃ ~45 ℃, 20~40 ℃ more preferably, 25~35 ℃ more preferably the fermentation temperature is about 30 ℃ most preferably. Also, keeping the 3 to 8, more preferably is pH4~7, pH of the medium in the fermentation process can be used and the pH adjusting agent as needed. The pH adjusting agent used, calcium carbonate, sodium hydroxide, potassium hydroxide and the like, but are preferably calcium carbonate.

The reaction time required for manufacturing (eg isopropanol) material is not particularly limited, and is carried out in a reaction time insofar as the effects of the present invention is observed. Optimization of these conditions can be readily made by those skilled in the art.

In the production method (eg isopropanol) substances according to the invention, be separated and recovered from the culture medium ingredients (eg, isopropanol) material produced in this manner, the specific procedures for this reason, the substances to be separated and recovered It is possible according to and readily selected by those skilled in the art.

Will now be described in more detail examples of the present invention by examples which are not intended to limit the scope of the invention.

Example 1: NBRC0396 share selection Candida utilis Candida utilis strain exhibiting the isopropanol resistance high, NBRC0619 shares, NBRC0988 shares, NBRC0639 shares, NBRC0626 shares, NBRC1086 shares, NBRC10707 shares, ATCC9255 shares, ATCC28955 strain, ATCC44638 shares, was inoculated YPD2 (20g / L peptone, 10g / L yeast extract, 20g / L glucose) in, ATCC9206 shares, ATCC96621 strain, ATCC18201 shares, and the ATCC15239 strain, and culture was performed at 30 ℃. Serially diluted 10-fold for each yeast culture in stationary phase, were spotted in YPD2 agar medium containing isopropanol 10~100g / L, and cultured at 30 ℃. And continued for 3 days incubation, was observed the formation of colonies. I listed in Table 1 shows the results.

From the above results, among the strains assayed, that NBRC1086 strain, ATCC9255 strain and ATCC15239 strain, indicating the IPA resistance highest 65g / L revealed.

Example 2: In ATCC15239 shares build Candida utilis ATCC15239 strain Candida utilis pyruvate decarboxylase gene is destroyed, destroyed the (PDC) pyruvate decarboxylase gene and its endogenous gene. The gene disruption was performe by the same method as the first 2010/095750 WO with the Cre-loxP system. By the destruction of the two, I was constructed

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 488 of 3077

strains PDC1 allele was completely deficient in (dPDC1 shares). dPDC1 strain not lacking ethanol production ability to completely, ethanol producing ability was markedly decreased.

Building isopropanol generation of vectors expressing CoA transferase subunit A gene of Clostridium acetobutylicum derive from the same subunit B gene, a secondary alcohol dehydrogenase gene and acetoacetic acid decarboxylase gene of Clostridium Beijerinki from Example 3: secondary alcohol dehydrogenase acetoacetic acid decarboxylase four genes involved in the synthesis, CoA transferase gene subunit A of Clostridium acetobutylicum from (ctfA), the subunit B (ctfB), Clostridium Beijerinki and origin (adc) In order to be expressed in Candida utilis efficiently (sadh) enzymes, and to design a novel gene sequences as an indication that the following items do not exist in nature, were synthesized:
The use of codons that are frequently used in (1) Candida utilis;
For introducing (3) chromosome embedded expression vector, grant of BamH I site at the 3 'end and Xba I site in the 5' end and elimination of inappropriate restriction enzyme site (2) gene cloning.

I is shown in SEQ ID NO: 7 and 1, 3, 5, respectively, the nucleotide sequence CoA transferase gene subunit A gene that ha been synthesized, the B subunit gene, the secondary alcohol dehydrogenase gene and acetoacetic acid decarboxylase gene. Also, I show in 2, 4, 6 and 8 SEQ ID NOs, the amino acid sequence encoded by these nucleotide sequences. I was named, GA09142-sadh GA09142-ctfA respectively, GA09142-ctfB, GA09142-adc a plasmid carrying the gene of these.

It is a built-in vector orotidine 5 'acid decarboxylase gene of Candida utilis to (URA3) seat, pVT92 is, (the first 2010/095750 WO), which has as a selection marker hygromycin resistance gene. By employing the linear DNA by BgII II processing, pVT92 may be introduced chromosome. To express the foreign gene, respectively, pVT92 has a glycerate 3 'phosphate kinase gene terminator glyceraldehyde dehydrogenase gene promoter outside of the Xba I / BamH I gap. In addition, 5 of the promoter 'end and 3 Terminator' to the end side, pVT92 has a Spe I site and Nhe I site, respectively.

It is a built-in vector alpha amino adipic acid reductase gene of Candida utilis to (LYS2) seat, pVT239 is, (the first 2010/095750 WO), which has as a selectable marker geneticin resistance gene. By employing the linear DNA by Not I processing, pVT239 may be introduced chromosome. Also, pVT239 has a Spe I site at the 5 'end of the geneticin resistance gene.

After the BamH I processing and Xba I, the GA09142-sadh and, GA09142-adc, four vectors containing each isopropanol synthase gene GA09142-ctfA, GA09142-ctfB, it was purified by agarose gel extraction, with the same restriction enzyme Th ligated into pVT92 which has been treated to give a pVT302 and pVT296, pVT298, pVT300. pVT300 and pVT296 and Spe I process, I went dephosphorylation. Also, after Nhe I, and Spe I treatment is performed purification of DNA fragments containing each ctfB expression cassette, the sadh expression cassette, and ligated into pVT300 and pVT296 respectively, pVT302 and pVT298 was obtained pVT325 and pVT321. After treatment with Spe I and Nhe I and pVT325, and Purification of a DNA fragment containing the sadh expression cassette and adc expression cassette is introduced into the Spe I site of pVT239, to obtain a pVT346. PVT321 finally obtained is a vector having a hygromycin resistance gene expression cassette of each of ctfB and ctfA, are integrated into the URA3 locus of Candida utilis, pVT346 is the expression cassette of each of sadh and adc It is a vector having a geneticin resistance gene is incorporated into the LYS2 locus of Candida utilis.

Example 4 was digested with BgII II building pVT321 of Candida utilis strains with isopropanol synthetic gene of four, was concentrated by ethanol precipitation. Using the DNA fragment obtained, ATCC15239 strain C. utilis, and was transformed dPDC1 strain obtained in Example 2 was spread on YPD medium containing 600μg / mL hygromycin, and incubated 2 days at 30 ℃ . Were selected a strain having all genes introduced.

Was digested with Not I and pVT346 was further concentrated by ethanol precipitation. Transforming each of the two strains obtained above by using a DNA fragment of each of the obtained was spread on YPD medium containing 200μg / mL geneticin and 600μg / mL hygromycin, and incubated 2 days at 30 ℃. And select a strain with all genes were introduced, TMS272 shares, was the TMS290 shares of isopropanol-producing strain of dPDC strain isopropanol production strains derived from strain ATCC15239.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 489 of 3077

Was carried out as shown below, the evaluation of the isopropanol-producing ability of the transformants obtained with isopropanol production test Example 4; Example 5.

Inoculated (20g / L peptone, 10g / L yeast extract, 20g / L glucose) in / 14mL volume test tube 2mL YPD liquid medium the various transformants were cultured with shaking for 24 hours at, 30 ℃ 140rpm. 100mL YPD10 YPD10C (20g / L peptone, 10g / L yeast extract, 100g / L glucose) or / 200mL Erlenmeyer flask (20g / L peptone, 10g / L yeast extract, 100g / L glucos pre-culture solution obtained, was inoculated to 45g / L calcium carbonate) / 200mL Erlenmeyer flask, and cultured with shaking OD600 = 1,100rpm initial, at 30 ℃. After sampling with time from the culture was filtered through a 0.2μm filter, and subjected to quantitative analysis of the various components. Glucose, glycerol, acetic acid, and ethanol, were quantified using the suggested refractometer high performance liquid chromatographic (manufactured by Shimadzu Corporation), ISep ION300 column (manufactured by Tokyo Chemical Industry Co., Ltd.). Acetone and isopropanol, were quantified using gas chromatographic GC-2010 Plus the (manufactured by Shimadzu Corporation).

Also, I did a similar experiment using (20g / L peptone, 10g / L yeast extract, 100g / L glucose, 45g / L calcium carbonate) a a medium YPD10C neutralizing culture simple in the flask.

I described in Figure 3 shows the results Figure 2, of the culture of the neutralization test conditions the results of the culture test in the non-neutralizing conditions. In addition, I was shown by the average of three repeated trials data.

According to Figure 2, the production of isopropanol while not observed in dPDC1 strain and ATCC15239 strain, which is th parent strain, the TMS290 strains and TMS272 strains introduced isopropanol synthetic genes, 0 and 0.16g / L at each up production of isopropanol .13g / L was observed. In addition, proliferative capacity and rate of consumption of sugar was decreased significantly in the strain that destroyed PDC1 gene. Glucose is completely consumed in the 24 hours of culture, the accumulation of ethanol was observed in the TMS272 share, but ethanol concentration was decreased with the continuation of subsequent culture. Since the production of isopropanol was observed while the ethanol concentration is reduced, the TMS272 strain be producing isopropanol by the secondary consume ethanol was produced at a time is considered. From the isopropanol production rate is reduced in the TMS290 strain derived from dPDC1 strain it in isopropanol production in yeast pathway acetyl-CoA in the cytoplasm, a substrate is fed from ethanol are the most importan was suggested. Since the assimilation of ethanol is an oxidation reaction, these results suggest the possibility that Crabtree negative yeast high respiratory activity such as Candida utilis is suitable in isopropanol production by yeast.

According to Figure 3, by the addition of calcium carbonate to the medium, to produce isopropanol 1.1g / L, production was increased from the non-neutralizing conditions TMS272 strain. Also, TMS272 strain has accumulated acetic acid 16.2g / L, the pH is lowered by acetic acid was accumulated for some reason, it was inhibiting the production of isopropanol was considered. In addition, in the TMS290 share, isopropanol production volume has increased to 0.9g / L in neutralizing conditions, I had accumulated a glycerol concentration as high as 37.4g / L on the other hand.

Followed by selection of the gene to improve the productivity of isopropanol by the over-expression of selected genes for improving the productivity of isopropanol by overexpression Example 6. I described in Figure 1 the metabolic pathway of the candidate gene is catalyst.

Array:; amino acid sequence SEQ ID NO: 11: nucleotide sequence; (ACS1 (;:;; SEQ ID NO: 9 amino acid sequence SEQ ID NO: 10 nucleotide sequence ACH1), acetyl-CoA synthase gene acetyl-CoA hydrolase gene of Candida utilis endogenous SEQ ID NO: 14), (ADH2 cytoplasmic localization types ethanol dehydrogenase gene; nucleotide sequence:; amino acid sequence SEQ ID NO: 13: nucleotide sequence; 12, ACS2 number SEQ ID NO: 15; amino acid sequence SEQ ID NO: 16), mitochondrial localization (SEQ ID NO: 18;:; amino acid sequence SEQ ID NO: 17 nucleotide sequence ADH3), cytoplasmic localization type acetaldehyde dehydrogenase gene (ALD1; nucleotide sequence:;: SEQ ID NO: 20 amino acid sequence SEQ ID NO: 19) type ethanol dehydrogenase gene , mitochondria localized acetaldehyde dehydrogenase gene (ALD2; nucleotide sequence:,: SEQ ID NO: 22 amino acid sequence SEQ ID NO: 21), (CRC1 mitochondrial inner membrane carnitine transporter; nucleotide sequence:;: SEQ ID NO: 24 amino acid sequence SEQ ID NO: 23), (CAT2 mitochondrial

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 490 of 3077

localization carnitine-acetyl-CoA transferase; nucleotide sequence:; SEQ ID NO: 26), (YAT1 carnitine acetyl transferase gene; nucleotide sequence: the amino acid sequence SEQ ID NO: 25; SEQ ID NO: 28 amino acid sequence SEQ ID NO: 2 , YAT2; nucleotide sequence: SEQ ID NO: 61; amino acid sequence:;:; nucleotide sequence (ERG10 acetyl-CoA acetyltransferase SEQ ID NO: 30), the amino acid sequence SEQ ID NO: 29 SEQ ID NO: 62), of Yarrowia lipolytica from ATP- citric each, (SEQ ID NO: 34 genes constituting the subunit to SEQ ID NO: 32, ACLY;:;; nucleotide sequence:;:; amino acid sequence SEQ ID NO: 31 amino acid sequence SEQ ID NO: 33 nucleotide sequence ACL1) the acid removal adding enzyme gene I was amplified by PCR as a template genomic DNA. SEQ ID NO: 36 and SEQ ID NO: 35), (ACS1;; Primers used for PCR ;(ACH1 are as follows: SEQ ID NO: 38 and SEQ ID NO: 37), (ACS2; SEQ ID NO: 40 and SEQ ID NO: 39), (ADH2 ; SEQ ID NO: 42 and SEQ ID NO: 41), (ADH3; SEQ ID NO: 44 and SEQ ID NO: 43), (ALD1; SEQ ID NO: 46 and SEQ ID NO: 45), (ALD2; SEQ ID NO: 48 and SEQ ID NO: 47), (CAT2; array SEQ ID NO: 50 and number 49), (CRC1; SEQ ID NO: 52 to SEQ ID NO: 51), (YAT1; SEQ ID NO: 54 to SEQ ID NO: 53), (YAT2; SEQ ID NO: 56 to SEQ ID NO: 55), (ACL1 SEQ ID NO: 57 SEQ ID NO: 58), (ACLY; SEQ ID NO: 60 to SEQ ID NO: 59), (ERG10; SEQ ID NO: 64 and SEQ ID NO: 63)

PCU155 (Keiji Kondo,, Yutaka Miura, Hidetaka Sone, Kazuo Kobayashi & Hiroshi Iijima., High-level expression of a sweet protein, monellin that with Xba I / BamH I treated PCR product was obtained, which was treated with the same restriction enzyme , in the food yeast Candida utilis, 15, 453 Nature Biotechnology - I was cloned into the 457 (1997)).. have as a selection marker cycloheximide resistance gene, is a vector that is multi-copy integration in the URA3 locus, respectively, pCU155 has the same terminator and glyceraldehyde dehydrogenase gene promoter on the outside of the Xba I / BamH I gap have. Isopropanol synthase gene is incorporated into the URA3 locus of isopropanol production strains are building so far, C.utilis is higher polyploid, because the URA3 locus multiple copies exist, the URA3 locus further gene transfer is possible. Also, since the isopropanol synthetic gene is inserted by the two-point recombination URA3 locus, no incorporatior of pCU155, replacement of the introduced genes occurs. was digested with BgII II plasmid candidate gene has been cloned. and concentrated by ethanol precipitation. Transforming the TMS272 strain C.utilis using a DNA fragment obtained was spread on YPD medium containing 20μg / mL cycloheximide 600μg / mL hygromycin, and 200μg / mL geneticin and culture for 4 days at 30 °C. In addition, DNA of ACLY and ACL1 were mixed, I transformation was carried out at the same time. The fermentation test carried out for each 4 clones obtained were investigated isopropanol productivity. I The results are shown in FIG.

According to Figure 4, due to overexpression of the ACL or ACS1, ACS2, CRC1, YAT1, ERG10,, isopropanol production volume will be 1.5 to 4.0 times has been revealed. Each gene of the ACL is an enzyme genes involved in the synthesis of acetyl-CoA and ACS1, ACS2, YAT1,, CRC1 gene is a carrier protein gene that exported to the cytoplasm from the mitochondria the precursor of acetyl-CoA synthesis, ERG10 gene acetyl CoA and a gene of synthesizing acetoacetyl-CoA. I These results, in order of isopropanol high production by yeast, to up-regulate the expression or activity of an enzyme gene for efficiently producing acetyl-CoA in the cytoplasm is important is shown.

Culture test building and isopropanol-producing strains overexpressing simultaneously YAT1 and, ACS2, ERG10, CRC1 ACL1 / ACLY: 7 Example

Was digested with BgII II plasmids each gene of YAT1 and, ACS2, ERG10, CRC1 ACL1, ACLY has been cloned, and concentrated by ethanol precipitation. The mixed DNA fragments of each 5μg obtained were transformed TMS272 strain by using this mixture was spread on YPD medium containing 10μg / mL cycloheximide 600μg / mL hygromycin, and 200μg / m geneticin at 30 °C I was cultured for 4 days. The fermentation test conducted on 80 clones obtained, and select a strain having the highest production of isopropanol, was the TMS378 strain.

Are shown in Figure 5 the result of the culture test of TMS378 shares. TMS272 shares whereas the production of isopropanol 1.3g / L, TMS378 shares showed isopropanol production volume of about 7.5 times and 9.8g / L. Moreover, whereas ethyl accumulated 18.1g / L The strain TMS272, TMS378 strain and 4.5g / L, acetate is a major by-product was reduced to about one-quarter.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 491 of 3077

And examined the culture conditions of TMS378 strains suitable for the production of isopropanol impact speed shaking and initial OD 600 in isopropanol production of TMS378 share: Example 8. Inoculated (20g / L peptone, 10g / L yeast extract, 20 / L glucose) in / 14mL volume test tube 2mL YPD liquid medium TMS378 strain was cultured with shaking for 24 hours at, 3(℃ 140rpm. Were inoculated (20g / L peptone, 10g / L yeast extract, 100g / L of glucose, 45g / L calcium carbonate) in / 300mL Erlenmeyer flask 100mL YPD10C a preculture obtained was 10 or 1 initial OD 600, were cultured with shaking at, 3(℃ 120rpm or 100rpm. Was added to the culture medium of glucose solids to a final concentration of 100g / L at 72 hours of culture, and cultured up to 144 hours. Was filtered through a 0.2μm filter by sampling over time. Various quantitative analysis, I was performed as described in Example 5. In addition, I was shown by the average of three repeated trials data. Figure 6 shows the results.

According to Figure 6, the highest isopropanol production was observed when cultured at an initial OD600 = 10,120rpm, isopropanol 18.5g / L was produced up to 144 hours of culture. Yield based on sugar at that time was 30.8%. From this, alsc (as the aeration amount is large) as the speed is higher, the higher the initial OD is higher, it is preferred that the isopropano production revealed shaking.

I have examined the culture conditions of TMS378 strains suitable for the production of impact isopropanol addition glucose concentration in isopropanol production of TMS378 share: Example 9. Inoculated (20g / L peptone, 10g / L yeast extract, 20 / L glucose) in / 14mL volume test tube 2mL YPD liquid medium TMS378 strain was cultured with shaking for 24 hours at, 3(℃ 140rpm. 100mL YPD10C YPD5C (20g / L peptone, 10g / L yeast extract, 100g / L glucose, 45g / L calcium carbonate) or 300mL Erlenmeyer flask (20g / L peptone, 10g / L yeast extract pre-culture solution obtained , was inoculated 50g / L glucose, to 45g / L calcium carbonate) / 300mL Erlenmeyer flask, and cultured with shaking OD600 = 1,120rpm early, at 30 ℃. For samples that were cultured in YPD10C, and glucose was added solid to a final concentration of 100g / L in 72 hours of culture, and cultured up to 192 hours. As for samples and cultured in YPD5C, and glucose was added to the solid to be a 3 Kaitsui concentration 50g / L at 48 hour intervals the culture, and cultured up to 192 hours. Glucose concentration in the final is a 200g / L any of the conditions. Was filtered through a 0.2μm filter by sampling over time. Various quantitative analysis, I was performed as described in Example 5. In addition, I was shown by the average of three repeated trials data. Figure 7 shows the results.

According to Figure 7, when cultured in YPD10C, followed by the addition of glucose 100g / L after 72 h incubation, isopropanol 19.2g / L was produced up to 122 hours of culture. However, the subsequent increase in isopropanol productior was not observed, acetic acid is accumulated (27.2g / L). Yield based on sugar of the maximum in the production of isopropanol was 29.0%. On the other hand, if it is cultured in YPD5C, was added 3 times of glucose 50g / L in 48-hour intervals, increase of isopropanol was observed up to 192 hours of culture, isopropanol of 27.4g / L production up to 192 hours after culture has been. In addition, there was no recognized these are little accumulation of acetic acid. Yield based oı sugar of the maximum in the production of isopropanol under these conditions was 41.5%. Therefore, preferable revealed ir isopropanol production keeping the relatively low concentration of glucose concentration to be added.

Example 10: In order to identify genes having a large contribution to improving the isopropanol production by over-expression at the same time specific gene large contribution isopropanol production increase, overexpression combinations ACS2 (i) and ACL1, ACLY strains overexpressing strain overexpressing lines over-expressing strains to be a combination of ERG10 (ii) and ACL1, ACLY, combination of ERG10 and (iii) ACS2, combination of ERG10 and, ACS2 (iv) ACL1, ACLY between the overexpression strains of single gene of each stock to build, respectively, overexpressing a combination of YAT1 and, ACS2, CRC1, ERG10 ACL1, ACLY and (TMS378 shares), a comparison was made of isopropanol production volume. Was performed by using the same method as construction of recombinant strain in Example 7 and Example 6, was mixed with genomic DNA to be built, and transformation was carried out at the same time. The fermentation test carried out for each 32 resulting clones were selected lines displaying isopropanol production average in the combination of the transgene, respectively. Inoculated (20g / L peptone, 10g / L yeast extract, 20g / L glucose) in / 14mL volume test tube 2mL YPD liquid medium build each strain was cultured with shaking for 24 hours at, 30 ℃ 140rpm. Inoculated (20g / L peptone, 10g / L yeast extract, 100g / L glucose, 45g / L calcium carbonate) in / glass test tube 3mL YPD10C pre-culture solution

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 492 of 3077

obtained, OD600 = 1,100rpm early, at 30 ℃ were cultured with shaking. Sampling was carried out over time, to evaluate the amount of isopropanol production by gas chromatographic GC-2010 and high-performance liquid chromatographic. In addition, I was shown by the average of three repeated trials data.

According to Figure 8, the overexpression of a single, isopropanol production is 2.2 to 2.7 times in each of the ERG10 and ACS1, ACS2 revealed. Isopropanol production volume of a strain overexpressing a combination of ERG10 (ii) and ACL1, ACLY and strain overexpressing a combination of ACS2 (i) and ACL1, ACLY includes a strain overexpressing alone ERG10 and ACS2, respectively significant difference was not observed. On the other hand, a dramatic improvement of isopropanol production observed in strains overexpressing a combination of ERG10 and (iii) ACS2, production volume was improved to 11.6 times. In addition, line over-expressing strains overexpressing a combination of ERG10 and (iii) ACS2, a combination c ERG10 and, ACS2 (iv) ACL1, ACLY, the combination of YAT1 and, ACS2, CRC1, ERG10 (ACL1, ACLY TMS378 shares In strains of the strain) overexpressing, no significant difference was observed in the isopropanol production volume.

Contribution overexpression of YAT1 and, CRC1 ACL gives improve productivity isopropanol is small, it can be simultaneously overexpressed the ERG10 and ACS2 in order to improve the isopropanol productivity when the results of these were simultaneously overexpressed But it is the most important has been shown. The above results, the metabolic pathways for the synthesis of acetyl-CoA from ethanol, namely to enhance the expression or activity of genes involved in metabolic pathways for the synthesis of acetyl-CoA from acetic acid is important for isopropanol high production by yeast I have shown that.

Figure 1 shows the synthetic route of isopropanol by yeast. Is a graph showing the results of a fermentation test for non-neutralizing conditions of strain introduced isopropanol production gene, isopropanol concentration in the medium (panel A) glucose concentration (panel B), ethanol concentration Figure 2 (panel shows the time course of (panel F) OD600 C), acid concentration (panel D), and glycerol concentration (panel E). In each panel, filled circles shows the results of the ATCC15239 strain, the white circles indicate the results TMS272 strain (ATCC15239 shares + IPA synthetic gene), black triangle shows the results of the PDC1 disrupted strain, TMS290 shares open triangles (PDC disrupted strain + IPA I shows the results of the synthetic gene). Is a graph showing the results of a fermentation test in the neutralization conditions of the strains introduced isopropanol production gene, isopropanol concentration in the medium (panel A), glucose concentration (panel B), the ethanol concentration Figure 3 (panels C I shows the time-dependent change glycerol concentration and acetic acid concentration (panel D) of (panel E),). In each panel, filled circles shows the results of the ATCC15239 strain, the white circles indicate the results TMS272 strain (ATCC15239 shares + IPA synthetic gene), black triangle shows the results of the PDC1 disrupted strain, TMS290 shares open triangles (PDC disrupted strain + IPA I shows the results of the synthetic gene). Figure 4 is a graph in TMS272 strain showed a relative amount with respect to isopropanol production by strain TMS272, isopropanol production in case of being over-expressed the various genes. Is a graph showing the results of the fermentation test of the strain overexpressing all genes for improving the isopropanol production acidic isopropanol concentration in the medium (panel A), glucose concentration (panel B), the ethanol concentration Figure 5 (Panel shows changes over time acetone concentration C), acetic acid concentration (panel D), glycerol concentration (panel E) of (panel F). In each panel, filled circles shows the results of the TMS272 shares, open circles (TMS272 shares + ACL1 YAT1; ACLY: ACS2: ERG10:: CRC1) TMS378 shares shows the results of. Is a graph showing the results of (effect of shaking speed and initial OD_{600}) fermentation test at 30 ℃ the TMS378 strains that overexpress all the genes for improving the isopropanol productivity, 6, isopropanol concentration in the culture medium shows changes over time acetone concentration (panel A), glucose concentration (panel B), ethanol concentration (panel C), acetic acid concentration (panel D), glycerol concentratior (panel E) of (panel F). In each panel, solid circle shows the results under the conditions of stirring rate = 100rpm, and initial OD600 = 1, the black triangles show the results under the condition of stirring rate = 120rpm, and initial OD600 = 1, the ope circles are the initial OD600 = shows the results under the conditions of stirring rate = 100rpm and 10, white triangle shows the results under the conditions of stirring rate = 120rpm and initial OD600 = 10. Figure 7 shows the results of (effect of fed glucose) fermentation test at 30 ℃ the TMS378 strain overexpressing all genes for improving the productivity of isopropanol And 100g / L initial glucose concentration, and fed glucose 100g / l after 72 h incubation, Panel A shows the result of the 19

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 493 of 3077

hours of culture. It is a 50g / L initial glucose concentration, 48 hours after culture in fed glucose 50g / l respectively after 14 hours and 96 hours later, Panel B shows the result of the 192 hours of culture. In each panel, and shows the time course concentration of acetic acid concentration of isopropanol (closed triangles), glucose concentration (closed circles), and ethanol concentration (open circles) (white triangles). 8 is a graph in TMS272 strain showed isopropanol production in case of being over-expressed the various genes.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 494 of 3077

PATENT COOPERATION TREATY

PCT

INTERNATIONAL PRELIMINARY REPORT ON PATENTABILITY

(Chapter II of the Patent Cooperation Treaty)

(PCT Article 36 and Rule 70)

Applicant's or agent's file reference **SS118-1026**	**FOR FURTHER ACTION**	See Form PCT/IPEA/416
International application No. **PCT/US2018/041579**	International filing date *(day/month/year)* **11 July 2018 (11.07.2018)**	Priority date *(day/month/year)* **13 July 2017 (13.07.2017)**

International Patent Classification (IPC) or both national classification and IPC
IPC: C12N 1/16, 5/10, 9/88, 9/96, 15/09, 15/52; C07H 21/04; C07K 14/415
CPC: See Supplemental Sheet

Applicant **RADICI CHIMICA S.P.A.**

1. This report is the international preliminary examination report, established by this International Preliminary Examining Authority under Article 35 and transmitted to the applicant according to Article 36.
2. This REPORT consists of a total of **6** sheets, including this cover sheet.
3. This report is also accompanied by ANNEXES, comprising:
   a. ☒ *(sent to the applicant and to the International Bureau)* a total of 48 sheets, as follows:
      ☒ sheets of the description, claims and/or drawings which have been amended and/or sheets containing rectifications authorized by this Authority, unless those sheets were superseded or cancelled, and any accompanying letters (see Rules 46.5, 66.8, 70.16, 91.2, and Section 607 of the Administrative Instructions).
      ☐ sheets containing rectifications, where the decision was made by this Authority not to take them into account because they were not authorized by or notified to this Authority at the time when this Authority began to draw up this report, and any accompanying letters (Rules 66.4*bis*, 70.2(e), 70.16 and 91.2).
      ☐ superseded sheets and any accompanying letters, where this Authority either considers that the superseding sheets contain an amendment that goes beyond the disclosure in the international application as filed, or the superseding sheets were not accompanied by a letter indicating the basis for the amendments in the application as filed, as indicated in item 4 of Box No. I and the Supplemental Box (see Rule 70.16(b)).
   b. ☐ *(sent to the International Bureau only)* a total of (indicate type and number of electronic carrier(s)) _____ containing a sequence listing, in the form of an Annex C/ST.25 text file, as indicated in the Supplemental Box Relating to Sequence Listing (see paragraph 3*ter* of Annex C of the Administrative Instructions).
4. This report contains indications relating to the following items:
   - ☑ Box No. I Basis of the report
   - ☐ Box No. II Priority
   - ☑ Box No. III Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
   - ☐ Box No. IV Lack of unity of invention
   - ☑ Box No. V Reasoned statement under Article 35(2) with regard to novelty, inventive step and industrial applicability; citations and explanations supporting such statement
   - ☐ Box No. VI Certain documents cited
   - ☑ Box No. VII Certain defects in the international application
   - ☐ Box No. VIII Certain observations on the international application

Date of submission of the demand	Date of completion of this report
13 May 2019 (13.05.2019)	**02 December 2019 (02.12.2019)**
Name and mailing address of the IPEA/US **COMMISSIONER FOR PATENTS MAIL STOP PCT, ATTN: IPEA/US P.O. BOX 1450 ALEXANDRIA, VA 22313-1450, UNITED STATES OF AMERICA** Facsimile No. **(571)273-8300**	Authorized officer **MANJUNATH RAO** Telephone No. **571-272-0939**

Form PCT/IPEA/409 (cover sheet) (January 2015)

INTERNATIONAL PRELIMINARY REPORT ON PATENTABILITY

International application No.
PCT/US2018/041579

Box No. I Basis of the report

1. With regard to the **language**, this report is based on:
   - [x] the international application in the language in which it was filed.
   - [ ] a translation of the international application into _____ which is the language of a translation furnished for the purposes of:
     - [ ] international search (Rules 12.3(a) and 23.1(b))
     - [ ] publication of the international application (Rule 12.4(a))
     - [ ] international preliminary examination (Rules 55.2(a) and/or 55.3(a) and (b))
2. With regard to the **elements** of the international application, this report is based on *(replacement sheets which have been furnished to the receiving Office in response to an invitation under Article 14 are referred to in this report as "originally filed" and are not annexed to this report)*:
   - [ ] the international application as originally filed/furnished, or
   - [x] the description:
     pages **1-550** as originally filed/furnished.
     pages* _____ received by this Authority on _____
     pages* _____ received by this Authority on _____
   - [x] the claims:
     Nos. _____ as originally filed/furnished.
     Nos.* _____ as amended (together with any statement) under Article 19
     Nos.* **1-155** received by this Authority on **10 June 2019 (10.06.2019)**
   - [x] the drawings:
     pages **1/59-59/59** as originally filed/furnished.
     pages* _____ received by this Authority on _____
     pages* _____ received by this Authority on _____
   - [ ] a sequence listing - see Supplemental Box Relating to Sequence Listing.
3. [x] The amendments have resulted in the cancellation of:
   - [ ] the description, pages _____
   - [x] the claims, Nos. **5,44,45,84-86,149**
   - [ ] the drawings, sheets/figs _____
   - [ ] the sequence listing *(specify)*: _____
4. [ ] This report has been established as if (some of) the amendments annexed to this report and listed below had not been made, since either they are considered to go beyond the disclosure as filed, or they were not accompanied by a letter indicating the basis for the amendments in the application as filed, as indicated in the Supplemental Box (Rules 70.2(c) and (c-*bis*)):
   - [ ] the description, pages _____
   - [ ] the claims, Nos. _____
   - [ ] the drawings, sheets/figs _____
   - [ ] the sequence listing *(specify)*: _____
5. [ ] This report has been established:
   - [ ] taking into account the **rectification of an obvious mistake** authorized by or notified to this Authority under Rule 91 (Rules 66.1(d-*bis*) and 70.2(e)).
   - [ ] without taking into account the **rectification of an obvious mistake** authorized by or notified to this Authority under Rule 91(Rules 66.4*bis* and 70.2(e)).
6. With regard to top-up searches (Rules 66.1*ter* and 70.2(f)):
   - [x] A top-up search was carried out by this Authority on **03 September 2019 (03.09.2019)**
     - [ ] Additional relevant documents have been discovered during the top-up search.
   - [ ] No top-up search was carried out by this Authority because it would serve no useful purpose.
7. [ ] Supplementary international search report(s) from Authority(ies) _____ has/ have been received and taken into account in establishing this report (Rule 45*bis*.8(b) and (c)).

** If item 4 applies, some or all of those sheets may be marked "superseded."*

Form PCT/IPEA/409 (Box No. I) (January 2015)

INTERNATIONAL PRELIMINARY REPORT ON PATENTABILITY

International application No.
PCT/US2018/041579

Box No. III Non-establishment of opinion with regard to novelty, inventive step and industrial applicability

The questions whether the claimed invention appears to be novel, to involve an inventive step (to be non obvious), or to be industrially applicable have not been examined in respect of:

☐ the entire international application.

☑ claims Nos. **113 and 114**

because:

☐ the said international application, or the said claims Nos. _____ relate to the following subject matter which does not require an international preliminary examination (*specify*):

☑ the description, claims or drawings *(indicate particular elements below)* or said claims Nos. **113 and 114** are so unclear that no meaningful opinion could be formed (*specify*):

Claims 113 and 114 are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

☐ the claims, or said claims Nos. _____ are so inadequately supported by the description that no meaningful opinion could be formed (*specify*):

☐ no international search report has been established for said claims Nos. _____

☐ a meaningful opinion could not be formed without the sequence listing; the applicant did not, within the prescribed time limit:

☐ furnish a sequence listing in the form of an Annex C/ST.25 text file, and such listing was not available to the International Preliminary Examining Authority in the form and manner acceptable to it; or the sequence listing furnished did not comply with the standard provided for in Annex C of the Administrative Instructions.

☐ furnish a sequence listing on paper or in the form of an image file complying with the standard provided for in Annex C of the Administrative Instructions, and such listing was not available to the International Preliminary Examining Authority in the form and manner acceptable to it; or the sequence listing furnished did not comply with the standard provided for in Annex C of the Administrative Instructions.

☐ pay the required late furnishing fee for the furnishing of a sequence listing in response to an invitation under Rules 13*ter*.1(a) or (b) and 13*ter*.2.

☐ See Supplemental Box for further details.

INTERNATIONAL PRELIMINARY REPORT ON PATENTABILITY

International application No.
PCT/US2018/041579

Box No. V **Reasoned statement under Article 35(2) with regard to novelty, inventive step and industrial applicability; citations and explanations supporting such statement**

1. Statement

Novelty (N)	Claims	1-112 and 115-155	YES
	Claims		NO
Inventive step (IS)	Claims	1-112 and 115-155	YES
	Claims		NO
Industrial applicability (IA)	Claims	1-112 and 115-155	YES
	Claims		NO

2. Citations and explanations (Rule 70.7)

---------- EXPLANATION ----------

The claims possess novelty and an inventive step under PCT Article 33(2) & 33(3) as the prior art does not teach nor suggest a genetically modified microorganism, such as a yeast/*Candida* having one or more heterologous terpene biosynthesis enzymes selected from those in claim 1, wherein the expression of said enzymes is provided by a nucleic acid that is responsive to fatty acid or alkane induction (the former for cytochrome P450).

The claims possess industial applicability under PCT Article 33(4) as production of terpenes are important in many industrial fields such as foods, cosmetics, pharmaceutical, etc.

INTERNATIONAL PRELIMINARY REPORT ON PATENTABILITY

International application No.
PCT/US2018/041579

Supplemental Box

In case **the space in any of the preceding boxes is not sufficient.**

Continuation of CPC:
C12N 1/16, 5/10, 9/88, 9/96, 15/09, 15/52, 15/63, 15/70, 15/74, 15/82, 15/8249; C07H 21/04; C07K 14/415; C12P 7/02, 7/26

INTERNATIONAL PRELIMINARY REPORT ON PATENTABILITY

International application No.
PCT/US2018/041579

Box No. VII Certain defects in the international application

The following defects in the form or contents of the international application have been noted:

In claims 140-142, beginning of second lines, a space should exist between "encoding" and the following word.

What is claimed is:

1. A genetically modified microorganism, comprising:

one or more heterologous nucleic acids encoding one or more terpene biosynthesis polypeptides, wherein expression of at least one of the heterologous nucleic acids is regulated by a nucleic acid that provides for fatty acid or alkane induction of expression of the terpene biosynthesis polypeptide:

- wherein the one or more heterologous nucleic acids that encode one or more terpene biosynthesis polypeptides is chosen from terpene synthase, phytoene synthase, geranylgeranyl diphosphate synthase, phytoene desaturase, lycopene cyclase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase, zeaxanthin glucosyltransferase and valencene synthase, or
- wherein the one or more heterologous nucleic acids is a cytochrome P450 reductase and is regulated by a nucleic acid that provides for alkane induction of expression of the cytochrome P450 reductase.

2. The microorganism of claim 1, wherein the microorganism is a fungus.

3. The microorganism of claim 2, wherein the fungus is a yeast.

4. The microorganism of claim 3, wherein the yeast is chosen from *Candida* spp, *Yarrowia* spp, *Rhodotorula* spp, *Rhodosporidium* spp, *Cryptococcus* spp, *Trichosporon* spp, *Lipomyces* spp, and *Blastobotrys* spp.

5. The microorganism of claim 4, wherein the fatty acid is a saturated fatty acid or an unsaturated fatty acid.

6. The microorganism of claim 5, wherein the fatty acid is chosen from one or more of oleic acid, palmitoleic acid, erucic acid, linoleic acid, palmitic acid, caproic acid, enanthic acid, caprylic acid pelargonic acid, capric acid, undecylic acid, lauric acid, myristic acid, pentadecanoic acid, margaric acid, stearic acid arachidic acid, behenic acid, tridecylic acid, and linolenic acid.

7. The microorganism of claim 4, wherein the alkane is chosen from one or more of hexane, heptane, nonane, decane, undecane, dodecane, tridecane, tetradecane, pentadecane, hexadecane, heptadecane, and octadecane

8. The microorganism of claim 7, wherein the nucleic acid that provides for fatty acid or alkane induction of expression of a terpene biosynthesis polypeptide comprises a fatty acid response element or an alkane response element.

9. The microorganism of claim 8, wherein the fatty acid response element comprises an oleic acid response element.

10. The microorganism of claim 8, wherein the alkane response element comprises an alkane response element 1 (ARE1) sequence.

11. The microorganism of claim 10, wherein the nucleic acid that provides for fatty acid induction of expression of a terpene biosynthesis polypeptide comprises a promoter region chosen from promoter regions of genes encoding hydratase-dehydrogenase-epimerase (HDE), acyl co-A oxidase (POX), acyl co-A thiolase (POT), peroxin (PEX) and peroxisomal adenine nucleotide transporter protein (ANT1).

12. The microorganism of claim 11, wherein the microorganism is *Candida viswanathii*.

13. The microorganism of claim 11, wherein the microorganism is *Blastobotrys adeninivorans*.

14. The microorganism of claim 13, wherein the one or more heterologous nucleic acids encode phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase.

15. The microorganism of claim 14, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a promoter region of the gene encoding a hydratase-dehydrogenase-epimerase (HDE).

16. The microorganism of claim 14, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a promoter region of the gene encoding a *Candida* hydratase-dehydrogenase-epimerase (HDE).

17. The microorganism of claim 14, 15 or 16, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a terminator region of the gene encodingan acyl co-A oxidase 4 (POX4).

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 502 of 3077

18. The microorganism of claim 14, 15 or 16, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a terminator region of the gene encodinga *Candida*acyl co-A oxidase 4 (POX4).

19. The microorganism of claim 13, wherein the one or more heterologous nucleic acids encode valencene synthase.

20. The microorganism of claim 19, wherein expression of the heterologous nucleic acids encoding valencene synthase is regulated by a promoter region of the gene encoding a hydratase-dehydrogenase-epimerase (HDE).

21. The microorganism of claim 19, wherein expression of the heterologous nucleic acids encoding valencene synthase is regulated by a promoter region of the gene encoding a *Candida* hydratase-dehydrogenase-epimerase (HDE).

22. The microorganism of claim 19, 20 or 21, wherein expression of the heterologous nucleic acids encoding valencene synthase is regulated by a terminator region of the gene encodingan acyl co-A oxidase 4 (POX4).

23. The microorganism of claim 19, 20 or 21, wherein expression of the heterologous nucleic acids encoding valencene synthase is regulated by a terminator region of the gene encodinga *Candida*acyl co-A oxidase 4 (POX4).

24. The microorganism of claim 13, wherein the one or more heterologous nucleic acids encode geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase.

25. The microorganism of claim 24, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase is regulated by a promoter region of the gene encoding a hydratase-dehydrogenase-epimerase (HDE).

26. The microorganism of claim 24, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 503 of 3077

cyclase/phytoene synthase is regulated by a promoter region of the gene encoding a *Candida* hydratase-dehydrogenase-epimerase (HDE).

27. The microorganism of claim 24, 25 or 26, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase is regulated by a terminator region of the gene encodingan acyl co-A oxidase 4 (POX4).

28. The microorganism of claim 24, 25 or 26, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase is regulated by a terminator region of the gene encodinga *Candida*acyl co-A oxidase 4 (POX4).

29. The microorganism of claim 28, wherein the one or more heterologous nucleic acids encoding the one or more terpene biosynthesis polypeptides are endogenously expressed in a microorganism chosen from *Cronobacter* spp, *Callitropsis* spp, *Xanthophyllomyces* spp, *Agrobacterium* spp, and *Pantoea* spp.

30. The microorganism of claim 29, wherein the amount and/or activity of aRas2 protein has been decreased.

31. The microorganism of claim 30, wherein the microorganism has been genetically modified to reduce or eliminate expression of an endogenous *RAS2* gene.

32. The microorganism of claim 31, wherein the amount and/or activity of an Faa1 protein has been decreased.

33. The microorganism of claim 32, wherein the microorganism has been genetically modified to reduce or eliminate expression of an endogenous *FAA1* gene.

34. The microorganism of claim 33, wherein the amounts and/or activities of one or more proteins chosen from acetyl-CoA C-acetyltransferase, HMG-CoA synthase, HMG-CoA reductase, mevalonate kinase, phosphomevalonate kinase, diphosphomevalonate decarboxylase, and isopentyl diphosphate delta isomerase,have been increased.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 504 of 3077

35. The microorganism of claim 34,wherein the microorganism has been genetically modified to increase expression of one or more endogenous genes chosen from *ERG10*, *ERG13*, *HMG1*, *ERG12*, *ERG8*, *MVD1*, and *IDI1*.

36. The microorganism of claim 35, wherein the amount and/or activity of dimethylallyltranstransferase or farnesyl diphosphate synthetase has been increased.

37. The microorganism of claim 36, wherein the microorganism has been genetically modified to increase expression of an endogenous *ERG20* gene.

38. The microorganism of claim 35, wherein expression of *ERG10*, *ERG13*, *HMG1*, *ERG12*, *ERG8*, *MVD1*, and/or *IDI1*is regulated by a promoter region of the gene encoding oleate-induced peroxisomal protein (POX18).

39. The microorganism of claim 35 or 38, wherein expression of *ERG10*, *ERG13*, *HMG1*, *ERG12*, *ERG8*, *MVD1*, and/or *IDI1*is regulated by a terminator region of the gene encoding oleate-induced peroxisomal protein (POX18).

40. The microorganism of claim 37, wherein expression of*ERG20*is regulated by a promoter region of the gene encoding oleate-induced peroxisomal protein (POX18).

41. The microorganism of claim 37 or 40, wherein expression of*ERG20*is regulated by a terminator region of the gene encoding oleate-induced peroxisomal protein (POX18).

42. A genetically modified microorganism, comprising:

one or more heterologous nucleic acids encoding one or more terpene biosynthesis polypeptides, and

a genetic modification that alters the expression of a polypeptide providing for transport of acetyl-carnitine in the microorganism,

wherein the expression of at least one of the heterologous nucleic acids is regulated by a nucleic acid that provides for glucose, fatty acid or alkane induction of expression of the terpene biosynthesis polypeptide.

43. The microorganism of claim 42, wherein the microorganism is a fungus.

44. The microorganism of claim 43, wherein the fungus is a yeast.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 505 of 3077

45. The microorganism of claim 44, wherein the yeast is chosen from *Candida* spp, *Yarrowia* spp, *Rhodotorula* spp, *Rhodosporidium* spp, *Cryptococcus* spp, *Trichosporon* spp, *Lipomyces* spp, and *Blastobotrys* spp.

46. The microorganism of claim 42, wherein the one or more heterologous nucleic acids encode one or more terpene biosynthesis polypeptides chosen from terpene synthase, phytoene synthase, geranylgeranyl diphosphate synthase, phytoene desaturase, lycopene cyclase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase, zeaxanthin glucosyltransferase, valencene synthase, and cytochrome p450 reductase.

47. The microorganism of claim 46, wherein the genetic modification reduces expression of the polypeptide providing for transport of acetyl-carnitine in the microorganism relative to a microorganism that does not have the genetic modification.

48. The microorganism of claim 47, wherein the genetic modification is a disruption, deletion or knockout of (i) a polynucleotide that encodes a polypeptide providing for transport of acetyl-carnitine, or (ii) a promoter operably linked to a polynucleotide that encodes a polypeptide providing for transport of acetyl-carnitine, whereby endogenous activity of a polypeptide providing for transport of acetyl-carnitine is reduced or abolished.

49. The microorganism of claim 47, further comprising a nucleic acid encoding a polypeptide providing for transport of acetyl-carnitine wherein expression of the nucleic acid encoding a polypeptide providing for transport of acetyl-carnitine is regulated by a promoter that provides for reduced expression relative to endogenous expression.

50. The microorganism of claim 49, wherein the genetic modification comprises replacing the promoter of an endogenous gene encoding the polypeptide providing for transport of acetyl-carnitine in the microorganism with a promoter that provides for reduced expression of the polypeptide in the microorganism relative to a microorganism that does not have the genetic modification.

51. The microorganism of claim 49 or 50, wherein the promoter that provides for reduced expression relative to endogenous expression is a promoter for a glucose-6-phosphate isomerase (G6PI) gene.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 506 of 3077

52. The microorganism of claim 49 or 50, wherein the promoter that provides for reduced expression relative to endogenous expression is a promoter for a *Candida* glucose-6-phosphate isomerase (G6PI) gene.

53. The microorganism of claim 42, wherein the polypeptide providing for transport of acetyl-carnitine in the microorganism is an acetyl-carnitine translocase (CRC).

54. The microorganism of claim 53, wherein the acetyl-carnitine translocase (CRC) is acetyl-carnitine translocase 1 (CRC1).

55. The microorganism of claim 42, wherein the microorganism is *Candida viswanathii.*

56. The microorganism of claim 42, wherein the microorganism is *Blastobotrys adeninivorans.*

57. The microorganism of claim 42, wherein the one or more heterologous nucleic acids encode phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase.

58. The microorganism of claim 57, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a promoter region of the gene encoding a hydratase-dehydrogenase-epimerase (HDE).

59. The microorganism of claim 57, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a promoter region of the gene encoding a *Candida* hydratase-dehydrogenase-epimerase (HDE).

60. The microorganism of claim 57, 58 or 59, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a terminator region of the gene encodingan acyl co-A oxidase 4 (POX4).

61. The microorganism of claim 57, 58 or 59, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a terminator region of the gene encodinga *Candida*acyl co-A oxidase 4 (POX4).

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 507 of 3077

62. The microorganism of claim 42, wherein the one or more heterologous nucleic acids encode geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase.

63. The microorganism of claim 62, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase is regulated by a promoter region of the gene encoding a hydratase-dehydrogenase-epimerase (HDE).

64. The microorganism of claim 62, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase is regulated by a promoter region of the gene encoding a *Candida* hydratase-dehydrogenase-epimerase (HDE).

65. The microorganism of claim 62, 63 or 64, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase is regulated by a terminator region of the gene encodingan acyl co-A oxidase 4 (POX4).

66. The microorganism of claim 62, 63 or 64, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase is regulated by a terminator region of the gene encodinga *Candida*acyl co-A oxidase 4 (POX4).

67. The microorganism of claim 42, wherein the one or more heterologous nucleic acids encoding the one or more terpene biosynthesis polypeptides are endogenously expressed in a microorganism chosen from *Cronobacter* spp, *Callitropsis* spp, *Xanthophyllomyces* spp, *Agrobacterium* spp, and *Pantoea* spp.

68. The microorganism of claim 42, wherein the amount and/or activity of aRas2 protein has been decreased.

69. The microorganism of claim 42, wherein the microorganism has been genetically modified to reduce or eliminate expression of an endogenous *RAS2* gene.

70. The microorganism of claim 42, wherein the amount and/or activity of an Faa1 protein has been decreased.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 508 of 3077

71. The microorganism of claim 42, wherein the microorganism has been genetically modified to reduce or eliminate expression of an endogenous *FAA1* gene.

72. The microorganism of claim 42, wherein the amounts and/or activities of one or more proteins chosen from acetyl-CoA C-acetyltransferase, HMG-CoA synthase, HMG-CoA reductase, mevalonate kinase, phosphomevalonate kinase, diphosphomevalonate decarboxylase, and isopentyl diphosphate delta isomerase, have been increased.

73. The microorganism of claim 42, wherein the microorganism has been genetically modified to increase expression of one or more endogenous genes chosen from *ERG10*, *ERG13*, *HMG1*, *ERG12*, *ERG8*, *MVD1*, and *IDI1*.

74. The microorganism of claim 42, wherein the amount and/or activity of dimethylallyltranstransferase or farnesyl diphosphate synthetase has been increased.

75. The microorganism of claim 42, wherein the microorganism has been genetically modified to increase expression of an endogenous *ERG20* gene.

76. The microorganism of claim 73, wherein expression of *ERG10*, *ERG13*, *HMG1*, *ERG12*, *ERG8*, *MVD1*, and/or *IDI1* is regulated by a promoter region of the gene encoding oleate-induced peroxisomal protein (POX18).

77. The microorganism of claim 73 or 76, wherein expression of *ERG10*, *ERG13*, *HMG1*, *ERG12*, *ERG8*, *MVD1*, and/or *IDI1* is regulated by a terminator region of the gene encoding oleate-induced peroxisomal protein (POX18).

78. The microorganism of claim 75, wherein expression of *ERG20* is regulated by a promoter region of the gene encoding oleate-induced peroxisomal protein (POX18).

79. The microorganism of claim 75 or 78, wherein expression of *ERG20* is regulated by a terminator region of the gene encoding oleate-induced peroxisomal protein (POX18).

80. A genetically modified *Candida viswanathii* yeast, comprising one or more heterologous nucleic acids encoding one or more terpene biosynthesis polypeptides, wherein expression of at least one of the heterologous nucleic acids is regulated by a nucleic acid that provides for glucose, fatty acid or alkane induction of expression of the terpene biosynthesis polypeptide and

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 509 of 3077

- wherein the one or more heterologous nucleic acids encoding one or more terpene biosynthesis polypeptides is chosen from terpene synthase, phytoene synthase, geranylgeranyl diphosphate synthase, phytoene desaturase, lycopene cyclase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase, zeaxanthin glucosyltransferase and valencene synthase, or
- wherein the one or more heterologous nucleic acids is a cytochrome P450 reductase and is regulated by a nucleic acid that provides for glucose or alkane induction of expression of the cytochrome P450 reductase.

81. The genetically modified *Candida viswanathii* yeast of claim 80, wherein the one or more heterologous nucleic acids encode phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase

82. The genetically modified *Candida viswanathii* yeast of claim 81, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a promoter region of the gene encoding a *Candida* hydratase-dehydrogenase-epimerase (HDE).

83. The genetically modified *Candida viswanathii* yeast of claim 82, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a terminator region of the gene encoding a *Candida*acyl co-A oxidase 4 (POX4).

84. The genetically modified *Candida viswanathii* yeast of claim 81, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a promoter region of the gene encoding a *Candida* glyceraldehyde-3-phosphate dehydrogenase (GPD).

85. The genetically modified *Candida viswanathii* yeast of claim 80, wherein the one or more heterologous nucleic acids encode valencene synthase.

86. The genetically modified *Candida viswanathii* yeast of claim 85, wherein expression of the heterologous nucleic acids encoding valencene synthase is regulated by a promoter region of the gene encoding a *Candida* hydratase-dehydrogenase-epimerase (HDE).

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 510 of 3077

87. The genetically modified *Candida viswanathii* yeast of claim 86, wherein expression of the heterologous nucleic acids valencene synthase is regulated by a terminator region of the gene encoding a *Candida*acyl co-A oxidase 4 (POX4).

88. The genetically modified *Candida viswanathii* yeast of claim 85, wherein expression of the heterologous nucleic acids encoding valencene synthase is regulated by a promoter region of the gene encoding a *Candida* glyceraldehyde-3-phosphate dehydrogenase (GPD).

89. The genetically modified *Candida viswanathii* yeast of claim 80, wherein the one or more heterologous nucleic acids encode geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase.

90. The genetically modified *Candida viswanathii* yeast of claim 89, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase is regulated by a promoter region of the gene encoding a *Candida* hydratase-dehydrogenase-epimerase (HDE).

91. The genetically modified *Candida viswanathii* yeast of claim 90, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase is regulated by a terminator region of the gene encoding a *Candida*acyl co-A oxidase 4 (POX4).

92. The genetically modified *Candida viswanathii* yeast of claim 89, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase is regulated by a promoter region of the gene encoding a *Candida* glyceraldehyde-3-phosphate dehydrogenase (GPD).

93. The genetically modified *Candida viswanathii* yeast of claim 80, wherein the one or more heterologous nucleic acids encode geranylgeranyl diphosphate synthase, phytoene desaturase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase,and cytochrome p450 reductase.

94. The genetically modified *Candida viswanathii* yeast of claim 93, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase,and cytochrome p450 reductase is regulated by a promoter region of the gene encoding a *Candida* hydratase-dehydrogenase-epimerase (HDE).

95. The genetically modified *Candida viswanathii* yeast of claim 94, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase,and cytochrome p450 reductase is regulated by a terminator region of the gene encoding a *Candida*acyl co-A oxidase 4 (POX4).

96. The genetically modified *Candida viswanathii* yeast of claim 93, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase,and cytochrome p450 reductase is regulated by a promoter region of the gene encoding a *Candida* glyceraldehyde-3-phosphate dehydrogenase (GPD).

97. The genetically modified *Candida viswanathii* yeast of claim 80, wherein the one or more heterologous nucleic acids encoding the one or more terpene biosynthesis polypeptides are endogenously expressed in a microorganism chosen from *Cronobacter* spp, *Callitropsis* spp, *Xanthophyllomyces* spp, *Agrobacterium* spp, and *Pantoea* spp.

98. The genetically modified *Candida viswanathii* yeast of claim 80, further comprising a genetic modification that alters the expression of one or more nucleic acids encoding one or more endogenous polypeptides.

99. The genetically modified *Candida viswanathii* yeast of claim 98, wherein the one or more endogenous polypeptides are chosen from polypeptides having one or more of the following activities: acyl-CoA synthetase activity, acyl-CoA oxidase activity, ATP-binding cassette transporter activity, carnitine acetyltransferase activity, transport of acetyl-carnitine, acyl-CoA thioesterase activity, acyl-CoA hydrolase activity, aldehyde dehydrogenase activity, monooxygenase activity, or monooxgenase reductase activity.

100. The genetically modified *Candida viswanathii* yeast of claim 80, wherein the amount and/or activity of aRas2 protein has been decreased.

101. The genetically modified *Candida viswanathii*yeast of claim 80, wherein the yeast has been genetically modified to reduce or eliminate expression of an endogenous *RAS2* gene.

102. The genetically modified *Candida viswanathii* yeast of claim 80, wherein the amount and/or activity of an Faa1 protein has been decreased.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 512 of 3077

103. The genetically modified *Candida viswanathii* yeast of claim 80, wherein the yeast has been genetically modified to reduce or eliminate expression of an endogenous *FAA1* gene.

104. The genetically modified *Candida viswanathii* yeast of claim 80, wherein the amounts and/or activities of one or more proteins chosen from acetyl-CoA C-acetyltransferase, HMG-CoA synthase, HMG-CoA reductase, mevalonate kinase, phosphomevalonate kinase, diphosphomevalonate decarboxylase, and isopentyl diphosphate delta isomerase, have been increased.

105. The genetically modified *Candida viswanathii* yeast of claim 80, wherein the yeast has been genetically modified to increase expression of one or more endogenous genes chosen from *ERG10*, *ERG13*, *HMG1*, *ERG12*, *ERG8*, *MVD1*, and *IDI1*.

106. The genetically modified *Candida viswanathii* yeast of claim 80, wherein the amount and/or activity of dimethylallyltranstransferase or farnesyl diphosphate synthetase has been increased.

107. The genetically modified *Candida viswanathii* yeast of claim 80, wherein the yeast has been genetically modified to increase expression of an endogenous *ERG20* gene.

108. The genetically modified *Candida viswanathii* yeast of claim 105, wherein expression of *ERG10*, *ERG13*, *HMG1*, *ERG12*, *ERG8*, *MVD1*, and/or *IDI1* is regulated by a promoter region of the gene encoding oleate-induced peroxisomal protein (POX18).

109. The genetically modified *Candida viswanathii* yeast of claim 105 or 108, wherein expression of *ERG10*, *ERG13*, *HMG1*, *ERG12*, *ERG8*, *MVD1*, and/or *IDI1*is regulated by a terminator region of the gene encoding oleate-induced peroxisomal protein (POX18).

110. The genetically modified *Candida viswanathii*yeast of claim 107, wherein expression of*ERG20* is regulated by a promoter region of the gene encoding oleate-induced peroxisomal protein (POX18).

111. The genetically modified *Candida viswanathii* yeast of claim 107 or 110, wherein expression of*ERG20* is regulated by a terminator region of the gene encoding oleate-induced peroxisomal protein (POX18).

112. A method for producing a terpene comprising:

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 513 of 3077

contacting the genetically modified microorganism of any one of claims 1 to 41 with a feedstock comprising a carbon source, and

culturing the microorganism under conditions in which the terpenes are produced from the feedstock.

113. A method for producing a terpene comprising:

contacting the genetically modified microorganism of any one of claims 42 to 79 with a feedstock comprising a carbon source, and

culturing the microorganism under conditions in which the terpenes are produced from the feedstock.

114. A method for producing a terpene comprising:

contacting the genetically modified *Candida viswanathii*yeast of any one of claims 80 to 111 with a feedstock comprising a carbon source, and

culturing the microorganism under conditions in which the terpenes are produced from the feedstock.

115. The method of claim 112, wherein the feedstock comprises one or more fatty acids.

116. The method of claim 115, wherein the feedstock comprises one or more saturated fatty acids or one or more unsaturated fatty acids; or one or more saturated fatty acids and one or more unsaturated fatty acids.

117. The method of claim 116, wherein the feedstock comprises one or more fatty acids chosen from oleic acid, palmitoleic acid, erucic acid, linoleic acid, palmitic acid, caproic acid, enanthic acid, caprylic acid pelargonic acid, capric acid, undecylic acid, lauric acid, myristic acid, pentadecanoic acid, margaric acid, stearic acid arachidic acid, behenic acid, tridecylic acid, and linolenic acid.

118. The method of claim 117 wherein the feedstock comprises oleic acid.

119. The method of claim 112, wherein the feedstock comprises one or more sugars.

120. The method of 119, wherein the feedstock comprises glucose.

121. The method of claim 112, wherein the feedstock comprises one or more alkane hydrocarbons.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 514 of 3077

122. The method of claim 121, wherein the feedstock comprises one or more alkane hydrocarbons chosen from one or more ofhexane, heptane, nonane, decane, undecane, dodecane, tridecane, tetradecane, pentadecane, hexadecane, heptadecane, and octadecane.

123. The method of claim 121, wherein the feedstock comprises one or more alkane hydrocarbons chosen from C10 to C14 alkane hydrocarbons, nonane and octadecane.

124. The method of claim 112, wherein the feedstock comprises one or more vegetable oils, derivatives thereof, or byproducts thereof.

125. The method of claim 124, wherein the feedstock comprises crude palm oil (CPO).

126. The method of claim 124, wherein the feedstock comprises palm oil fatty acid distillate (PFAD).

127. The method of claim 112, wherein the terpene is chosen from lycopene, beta carotene and astaxanthin.

128. The method of claim 127, wherein the yield or titer of the terpene is between about 0.001 g/L to about 2.0 g/L.

129. The method of claim 127, wherein the yield or titer of the terpene is between about 0.05 g/L to about 0.5 g/L.

130. The method of claim 127, wherein the yield or titer of the terpene is between about 0.001 g/L to about 0.2 g/L.

131. The method of claim 128, wherein the yield or titer of lycopene is between about 0.03 g/L to about 2.0 g/L.

132. The method of claim 129, wherein the yield or titer of lycopene is between about 0.05 g/L to about 0.5 g/L.

133. The method of claim 130, wherein the yield or titer of lycopene is between about 0.001 g/L to about 0.2 g/L.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 515 of 3077

134. The method of claim 128, wherein the yield or titer of beta carotene is between about 0.1 g/L to about 0.2 g/L.

135. The method of claim 130, wherein the yield or titer of beta carotene is between about 0.01 g/L to about 0.1 g/L.

136. The method of claim 130, wherein the yield or titer of beta carotene is between about 0.001 g/L to about 0.01 g/L.

137. The method of claim 128, wherein the yield or titer of astaxanthin is between about 0.005 g/L to about 0.01 g/L.

138. The method of claim 130, wherein the yield or titer of astaxanthin is between about 0.001 g/L to about 0.005 g/L.

139. The method of claim 112, further comprising isolating the terpene.

140. A genetically modified yeast comprising one or more heterologous nucleic acids encodingphytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase, wherein the amount and/or activity of a Ras2 protein has been decreased.

141. A genetically modified yeast comprising one or more heterologous nucleic acids encodinggeranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase, wherein the amount and/or activity of a Ras2 protein has been decreased.

142. A genetically modified yeast comprising one or more heterologous nucleic acids encodinggeranylgeranyl diphosphate synthase, phytoene desaturase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase,and cytochrome p450 reductase, wherein the amount and/or activity of a Ras2 protein has been decreased.

143. The genetically modified yeast of claim 142, wherein the yeast has been genetically modified to reduce or eliminate expression of an endogenous *RAS2* gene.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 516 of 3077

144. The genetically modified yeast of claim 142, wherein the amount and/or activity of an Faa1 protein has been decreased.

145. The genetically modified yeast of claim 142, wherein the yeast has been genetically modified to reduce or eliminate expression of an endogenous *FAA1* gene.

146. The genetically modified yeast of claim 142, wherein the amounts and/or activities of one or more proteins chosen from acetyl-CoA C-acetyltransferase, HMG-CoA synthase, HMG-CoA reductase, mevalonate kinase, phosphomevalonate kinase, diphosphomevalonate decarboxylase, and isopentyl diphosphate delta isomerase,have been increased.

147. The genetically modified yeast of claim 142, wherein the yeast has been genetically modified to increase expression of one or more endogenous genes chosen from *ERG10*, *ERG13*, *HMG1*, *ERG12*, *ERG8*, *MVD1*, and *IDI1*.

148. The genetically modified yeast of claim 142, wherein the amount and/or activity of dimethylallyltranstransferase or farnesyl diphosphate synthetase has been increased.

149. The genetically modified yeast of claim 142, wherein the yeast has been genetically modified to increase expression of an endogenous *ERG20* gene.

150. The genetically modified yeast of claim 142, wherein the yeast is *Candida viswanathii.*

151. The genetically modified yeast of claim 142, wherein the yeast is *Blastobotrys adeninivorans.*

152. Use of the genetically modified yeast of any one of claims 140 to 151 for the production of a terpene.

153. Use of the genetically modified yeast of any one of claims 140 and 143-151 for the production of lycopene.

154. Use of the genetically modified yeast of claim 141 and 143-151 for the production of beta carotene.

155. Use of the genetically modified yeast of any one of claims 142-151 for the production of astaxanthin.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 517 of 3077

Atty. Docket No. SS118-1026

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application	:	Radici Chimica S.p.A.
International Application No.	:	PCT/US2018/41579
International Filing Date	:	11 July 2018
For	:	**BIOLOGICAL METHODS FOR PREPARING TERPENES**

ISA/US Mail Stop PCT
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450
Attn: ISA/US Authorized Officer Shane Thomas

RESPONSE TO WRITTEN OPINION

Dear Sir,

Reference is made to the Demand under Article 31 of the PCT for the above-captioned case. The following is in reply to the outstanding Written Opinion of the ISA dated September 13th, 2018.

Arguments are herein provided to overcome the objections raised in the above cited written opinion.

1. Amendments to the claims:

Claims 6-42, 49-82, 86-145 and 150-162 were believed not in accordance with PCT Rule 6.4(a), and were thus not examined.

A new set of claims 1-155 is provided to replace the currently pending set of claims 1-162. Claims 1-155 are respectfully believed to comply with PCT Rule 6.4(a).

Claim 1: has been amended to include the limitations of pending claim 5, and to specify that when the heterologous nucleic acid is cytochrome P450 reductase, it is regulated by a nucleic acid that provides for alkane induction.

Claims 2-4 have not been amended.

Claim 5 has been deleted.

Claims 6-41 have been renumbered and formally amended to correct dependency.

Claim 42 has been amended to include the limitations of claims 44 and 45.

Pending claims 44 and 45 have been deleted.

Claims 43-79 have been renumbered and formally amended to correct dependency.

Claim 80 has been amended to include the limitations of claims 84-86 and to specify that when the heterologous nucleic acid is cytochrome P450 reductase, it is regulated by a nucleic acid that provides for glucose or alkane induction.

Pending claims 84-86 have been deleted.

Claims 81-155 have been renumbered and formally amended to correct dependency.

Claim 149 has been deleted.

New claims 1-155 are respectfully believed not contain any additional subject matter with respect to the application as filed.

2. NOVELTY (Re Box No. V, 2)

Claims 1-4, 5/1-4 and 83-84 are believed to lack novelty under PCT Art 33(2) as being anticipated by WO 2014/100504 (hereinafter Verdezyne).

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 519 of 3077

Claim 1 has been amended by incorporating the limitations of claim 5, and is now directed to

"A genetically modified microorganism, comprising:

one or more heterologous nucleic acids encoding one or more terpene biosynthesis polypeptides, wherein expression of at least one of the heterologous nucleic acids is regulated by a nucleic acid that provides for fatty acid or alkane induction of expression of the terpene biosynthesis polypeptide:

- wherein the one or more heterologous nucleic acids that encode one or more terpene biosynthesis polypeptides is chosen from terpene synthase, phytoene synthase, geranylgeranyl diphosphate synthase, phytoene desaturase, lycopene cyclase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase, zeaxanthin glucosyltransferase and valencene synthase, or

- wherein the one or more heterologous nucleic acids is a cytochrome P450 reductase and is regulated by a nuclcic acid that provides for alkane induction of expression of the cytochrome P450 reductase."

2.1 Novelty of claim 1 in view of Verdezyne

Verdezyne describes "biological methods for producing a fatty dicarboxylic acid and engineered microorganisms capable of such production" (abstract). Verdezyne describes a cytochrome P450 reductse polypeptide (Page 27, third paragraph) and POX4 promoters (page 41, 1st paragraph.).

Verdezyne does not describe microorganisms encoding a nucleic acid encoding one or more terpene biosynthesis polypeptides chosen from the group consisting of terpene synthase, phytoene synthase, geranylgeranyl diphosphate synthase, phytoene

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 520 of 3077

desaturase, lycopene cyclase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase, zeaxanthin glucosyltransferase and valencene synthase, nor does Verdezyne describe microorganisms encoding a cytochrome P450 reductase which is regulated by a nucleic acid that provides for alkane induction of expression.

Claim 1 is thus believed to be novel over Verdezyne.

Conclusions

The novelty objection raised in the outstanding Written Opinion is believed to be overcome in so far as the prior art document Verdezyne does not disclose the specific microorganisms described in amended claim 1.

Since pending claims 2-42 (new claims 2-41) depend on claim 1 and include all its features, new claims 2-41 are also believed to be novel over the prior art document Verdezyne.

2.2 Novelty of claim 42 in view of Verdezyne

Pending claims 44 and 45 are recognized as being novel (Re Box No.V) over the cited prior art.

New claim 42, which previously not examined, incorporates all the limitations of pending claims 44 and 45, and is thus also believed to be novel.

Conclusions

New claim 42 is respectfully believed to be novel over the cited prior art documents.

Since new claims 43-79 depend on claim 42 and include all its features, new claims 43-79 are also believed to be novel over the prior art documents.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 521 of 3077

2.3 Novelty of claim 83 in view of Verdezyne

The Examiner recognizes the novelty of pending claim 85 (Re Box No.V) over the cited prior art.

New claim 80 (pending claim 83) relates to a genetically modified *Candida viswanathii* yeast.

New claim 80 has been amended by incorporating all the limitations of pending claims 84-86.

New claim 80 is now believed to be novel over Verdezyne.

Conclusions

New claim 80 is respectfully believed to be novel over the cited prior art documents.

Since new claims 81-111 depend on claim 80 and include all its features, new claims 81-111 are also believed to be novel over the prior art documents.

2.4 Novelty of new claims 112 - 139

Claims 112 – 139 are method claims which all relate to a method for producing a terpene.

Independent claim 112 relates to a method for producing a terpene by contacting the genetically modified microorganism of claims 1-41 with a feedstock comprising a carbon source.

Since, as indicated in item 2.1, the microorganism of claims 1-41 is novel over the cited prior art, it is respectfully believed that the method for producing a terpene with the microorganism of claims 1-41 is also novel.

Independent claim 113 relates to a method for producing a terpene by contacting the genetically modified microorganism of claims 42-79 with a feedstock comprising a carbon source.

Since, as indicated in item 2.2, the microorganism of claims 42-79 is novel over the cited prior art, it is respectfully believed that the method for producing a terpene with the microorganism of claims 42-79 is also novel over the same prior art.

Independent claim 114 relates to a method for producing a terpene by contacting the genetically modified *Candida viswanathii* of claims 80-111 with a feedstock comprising a carbon source.

Since, as indicated in item 2.3, the microorganism of claims 80-111 is novel over the cited prior art, it is respectfully believed that the method for producing a terpene with the genetically modified *Candida viswanathii* of claims 80-111 is also novel over the same prior art.

Since claims 116-139 are all dependent on claim 112 and have all the limitations thereof, dependent claims 116-139 are also believed to be novel over the cited prior art.

2.5 Novelty of pending claims 146, 147 and 148 in view of Kirin

Pending claims 146-148 (new claims (140, 141 and 142) are believed to lack novelty in view of JP 2012-254044 (hereinafter Kirin).

The Examiner recognizes the novelty of pending claim 149 (Re Box No.V) over the cited prior art.

New claims 140-142 have been amended to include the limitations of pending claim 149 and are thus now believed to be novel over Kirin.

Conclusions

New claims 140-142 are respectfully believed to be novel over the cited prior art document.

Since new claims 143-151 depend on claim 142 and include all its features, new claims 143-151 are also believed to be novel over the prior art documents.

2.6 Novelty of new claims 152 - 155

Claims 152 – 155 are use claims which all relate to the use of a genetically modified yeast.

New independent claim 152 relates to the use of the genetically modified yeast of claims 140-151 for producing a terpene.

Since the genetically modified yeast of claims 140-151 is respectfully believed to be novel, as indicated in item 2.5 above, new claim 152 is also believed to be novel.

New independent claim 153 relates to the use of the genetically modified yeast of claims 140, 143-151 for producing a lycopene.

Since the genetically modified yeast of claims 140, 143-151 is respectfully believed to be novel, as indicated in item 2.5 above, new claim 153 is also believed to be novel.

New independent claim 154 relates to the use of the genetically modified yeast of claims 141, 143-151 for producing beta carotene.

Since the genetically modified yeast of claims 141, 143-151 is respectfully believed to be novel, as indicated in item 2.5 above, new claim 154 is also believed to be novel.

New independent claim 155 relates to the use of the genetically modified yeast of claims 142-151 for producing astaxanthin.

Since the genetically modified yeast of claims 142-151 is respectfully believed to be novel, as indicated in item 2.5 above, new claim 155 is also believed to be novel.

3. INVENTIVE STEP (Re Box No. V, 2)

According to the Examiner, the subject matter of claims 44, 46/44, 47/46/44, and 48/47/46/44 lack an inventive step as being obvious over Kirin in view of Verdezyne.

The Applicant respectfully disagrees and will discuss the reasoning in support of inventiveness of claims 1-155 below.

A. Independent claim 1

Kirin was taken to be the closest prior art, for the subject matter of claims 44, 46/44, 47/46/44, and 48/47/46/44.

Kirin describes the provision of a yeast strain capable of producing a metabolite through acetyl-CoA (abstract) and a process for the production of substances utilizing metabolic pathways via acetyl CoA in yeast (par. [0001]).

The subject matter of claim 1 differs from what is known through D1 by the distinguishing features of relating to a genetically modified microorganism, comprising a nucleic acid encoding a terpene biosynthesis polypeptide which is regulated by a nucleic acid that provides for fatty acid or alkane induction of expression.

The objective technical problem addressed by the present invention in the light of the closest prior art (Kirin) is that of obtaining an improved genetically modified microorganism which allows for the production of terpenes. This problem is solved by

the present invention with results which are surprisingly very much higher and ameliorated than the expectations.

The invention of claim 1 solves this problem by providing the microorganism which is characterized by a specific combination of terpene biosynthesis polypeptides and nucleic acids which provide for the induction of expression of the terpene biosynthesis polypeptides.

This specific combination allows for the production of lycopene, B-carotene and astaxanthin, when the microorganisms are grown in the presence of specific substrates.

Evidence that the problem is solved can be found for example in Table 22 (page 467), which shows the results of the analysis of carotenoid production carried out with some of the genetically modified microorganisms.

Examples 67, 68 and 69 (pages 484-486) describe the construction of some of the most performing strains which produce lycopene from an oleic acid substrate. The performance of these strains is shown in Table 22, last 6 rows from the bottom.

There is no suggestion in Kirin that a yeast strain capable of producing a metabolite through acetyl-CoA could produce lycopene from oleic acid.

In fact, Kirin describes yeast strains for the production of isopropanol. As can be seen in Figures 1-7, the results of the analysis carried out in Kirin is that of obtaining a yeast strain which is capable of producing isopropanol when different genes are expressed. In particular, Figure 8 shows a graphical representation of the combination of genes which provides for the highest isopropanol production.

The Examiner indicates that Kirin discloses the microorganism of claim 43 but does not disclose "wherein expression of at least one of the heterologous nucleic acids is

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 526 of 3077

regulated by a nucleic acid that provides for fatty acid or alkane induction of expression of the terpene biosynthesis polypeptide".

The Examiner fails to indicate where Kirin discloses the microorganism of claim 43.

Paragraph [0046] of Kirin states that "it is possible to construct yeast strains that produce certain carotenoids, such as lycopene, specific sterols produced from squalene". In the same paragraph Kirin further states that "Candida utilis expressing a phytoene synthase gene, a phytoene desaturase gene and overexpressing an endogenous HMG-CoA reductase gene (HMG1) has been reported to highly accumulate lycopene (Shimada H, Kondo K, Fraser PD, Miura Y, Saito T, Misawa N. Increased carotenoid production by the food yeast Candida utilis through metabolic engineering of the isoprenoid pathway. Appl Environ Microbiol. 1998 Jul;64(7):2676 - 80)".

It is not clear where the microorganism of claim 43 can be found in the description of Kirin.

The skilled person, faced with the objective technical problem of the present invention, i.e. obtaining a genetically modified microorganism which allows for an improved terpene biosynthesis, and having in mind the disclosure of the Kirin document, would not arrive at the solution of claim 1.

In view of the above, the subject matter of new independent claim 1 is inventive over Kirin.

Combination of Kirin and Verdezyne

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 527 of 3077

Even if the skilled person, starting from Kirin and having in mind the teachings therein comprised, and faced with the objective technical problem described above, were to try to combine Kirin and Verdezyne, he would not arrive at the claimed solution.

Verdezyne describes biological methods and engineered microorganisms for producing a fatty dicarboxylic acid.

Verdezyne does not describe genetically modified microorganisms for the production of terpenes.

The invention of claim 1 is hence not obvious over Kirin or Verdezyne, either taken alone or in combination.

B. Independent claims 42, 80 and 140-142

Like claim 1, also independent claims 42, 80 and 140-142 describe genetically modified microorganisms, have the same distinguishing features of claim 1 and differ from Kirin and from Verdezyne for the same reasoning indicated for claim 1.

Independent claims 42, 80 and 140-142 are thus also believed to be inventive over the cited prior art documents, either taken alone or in combination.

C. Independent claims 112-114, and 152-155

Independent claims 112-114 and 152-155 relate to methods for producing and uses of the genetically modified microorganisms described in claims 1-41.

Independent claims 112-114 and 152-155 are thus also believed to be inventive over Kirin and Verdezyne, either taken alone or in combination.

D. Dependent claims 2-41, 43-79, 81-111, 115-139, 143-151

Since the dependent claims include all the features of the independent claims they depend on, they are respectfully believed to be inventive over the cited documents.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 528 of 3077

Atty. Docket No. SS118-1026

*****

It is respectfully believed that the present response provides basis for the issuance of a positive International Preliminary Report on Patentability of the instant application.

Should any additional items remain not fully attended, it is respectfully requested either the issuance of a second Written Opinion according to R.66.4(b) PCT explaining the Examiner's position, and a further opportunity for amendments and remarks to be filed.

Respectfully submitted

SILVIA SALVADORI, P.C.

June 10, 2019

By: /Silvia Salvadori/
Silvia Salvadori, Reg. No. 48,265
270 Madison Avenue, 8th Floor
New York, NY 10016
Phone: (212) 897-1938
Fax: (646) 219-6529

Attachments:

New claims 1-155 clean and marked-up version.

P021168WO_20190608_AMD claims marked up

What is claimed is:

1. A genetically modified microorganism, comprising:

one or more heterologous nucleic acids encoding one or more terpene biosynthesis polypeptides, wherein expression of at least one of the heterologous nucleic acids is regulated by a nucleic acid that provides for fatty acid or alkane induction of expression of the terpene biosynthesis polypeptide;

wherein the one or more heterologous nucleic acids that encode one or more terpene biosynthesis polypeptides is chosen from terpene synthase, phytoene synthase, geranylgeranyl diphosphate synthase, phytoene desaturase, lycopene cyclase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase, zeaxanthin glucosyltransferase and valencene synthase, or

wherein the one or more heterologous nucleic acids is a cytochrome P450 reductase and is regulated by a nucleic acid that provides for alkane induction of expression of the cytochrome P450 reductase.

2. The microorganism of claim 1, wherein the microorganism is a fungus.

3. The microorganism of claim 2, wherein the fungus is a yeast.

4. The microorganism of claim 3, wherein the yeast is chosen from *Candida* spp, *Yarrowia* spp, *Rhodotorula* spp, *Rhodosporidium* spp, *Cryptococcus* spp, *Trichosporon* spp, *Lipomyces* spp, and *Blastobotrys* spp.

~~5. The microorganism of any one of claims 1 to 4, wherein the one or more heterologous nucleic acids encode one or more terpene biosynthesis polypeptides chosen from terpene synthase, phytoene synthase, geranylgeranyl diphosphate synthase, phytoene desaturase, lycopene cyclase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase, zeaxanthin glucosyltransferase, valencene synthase, and cytochrome p450 reductase.~~

6. The microorganism ~~of any one~~ of claims ~~1 to 5~~, wherein the fatty acid is a saturated fatty acid or an unsaturated fatty acid.

7. The microorganism ~~of any one~~ of claims ~~1 to 6~~, wherein the fatty acid is chosen from one or more of oleic acid, palmitoleic acid, erucic acid, linoleic acid, palmitic acid, caproic acid, enanthic acid, caprylic acid pelargonic acid, capric acid, undecylic acid, lauric acid, myristic acid,

pentadecanoic acid, margaric acid, stearic acid arachidic acid, behenic acid, tridecylic acid, and linolenic acid.

87. The microorganism ~~of any one~~ of claims ~~1 to 54~~, wherein the alkane is chosen from one or more of hexane, heptane, nonane, decane, undecane, dodecane, tridecane, tetradecane, pentadecane, hexadecane, heptadecane, and octadecane

88. The microorganism ~~of any one~~ of claims ~~1 to 87~~, wherein the nucleic acid that provides for fatty acid or alkane induction of expression of a terpene biosynthesis polypeptide comprises a fatty acid response element or an alkane response element.

~~89~~. The microorganism of claim 88, wherein the fatty acid response element comprises an oleic acid response element.

~~1010~~. The microorganism of claim 88, wherein the alkane response element comprises an alkane response element 1 (ARE1) sequence.

~~1111~~. The microorganism ~~of any one~~ of ~~claimsclaim~~ ~~1 to 1110~~, wherein the nucleic acid that provides for fatty acid induction of expression of a terpene biosynthesis polypeptide comprises a promoter region chosen from promoter regions of genes encoding hydratase-dehydrogenase-epimerase (HDE), acyl co-A oxidase (POX), acyl co-A thiolase (POT), peroxin (PEX) and peroxisomal adenine nucleotide transporter protein (ANT1).

~~1212~~. The microorganism ~~of any one~~ of ~~claimsclaim~~ ~~1 to 1211~~, wherein the microorganism is *Candida viswanathii.*

~~1313~~. The microorganism ~~of any one~~ of ~~claimsclaim~~ ~~1 to 1211~~, wherein the microorganism is *Blastobotrys adeninivorans.*

~~1414~~. The microorganism ~~of any one~~ of ~~claimsclaim~~ ~~1 to 1413~~, wherein the one or more heterologous nucleic acids encode phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase.

~~1515~~. The microorganism of claim ~~1414~~, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a promoter region of the gene encoding a hydratase-dehydrogenase-epimerase (HDE).

~~4716~~. The microorganism of claim ~~1614~~, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a promoter region of the gene encoding a *Candida* hydratase-dehydrogenase-epimerase (HDE).

~~4817~~. The microorganism of claim ~~15, 16 or 1714, 15 or 16~~, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a terminator region of the gene encodingan acyl co-A oxidase 4 (POX4).

~~4918~~. The microorganism of claim 14, 15 or ~~1615, 16 or 17~~, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a terminator region of the gene encodinga *Candida*acyl co-A oxidase 4 (POX4).

~~2019~~. The microorganism ~~of any one of claims~~claim ~~1 to 1413~~, wherein the one or more heterologous nucleic acids encode valencene synthase.

~~2120~~. The microorganism of claim ~~2019~~, wherein expression of the heterologous nucleic acids encoding valencene synthase is regulated by a promoter region of the gene encoding a hydratase-dehydrogenase-epimerase (HDE).

~~2221~~. The microorganism of claim ~~2019~~, wherein expression of the heterologous nucleic acids encoding valencene synthase is regulated by a promoter region of the gene encoding a *Candida* hydratase-dehydrogenase-epimerase (HDE).

~~2322~~. The microorganism of claim 19, 20 or ~~2120, 21 or 22~~, wherein expression of the heterologous nucleic acids encoding valencene synthase is regulated by a terminator region of the gene encodingan acyl co-A oxidase 4 (POX4).

~~2423~~. The microorganism of claim ~~20, 21 or 2219, 20 or 21~~, wherein expression of the heterologous nucleic acids encoding valencene synthase is regulated by a terminator region of the gene encodinga *Candida*acyl co-A oxidase 4 (POX4).

P021168WO_20190608_AMD claims marked up

2624. The microorganism ~~of any one of claims~~claim 1 ~~to 1413~~, wherein the one or more heterologous nucleic acids encode geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase.

2625. The microorganism of claim 2624, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase is regulated by a promoter region of the gene encoding a hydratase-dehydrogenase-epimerase (HDE).

2726. The microorganism of claim 2624, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase is regulated by a promoter region of the gene encoding a *Candida* hydratase-dehydrogenase-epimerase (HDE).

2827. The microorganism of claim ~~25, 26 or 2724~~, 25 or 26, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase is regulated by a terminator region of the gene encodingan acyl co-A oxidase 4 (POX4).

2928. The microorganism of claim 24, 25 or ~~2625, 26 or 27~~, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase is regulated by a terminator region of the gene encodinga *Candida*acyl co-A oxidase 4 (POX4).

3029. The microorganism ~~of any one of claims~~claim 1 ~~to 2928~~, wherein the one or more heterologous nucleic acids encoding the one or more terpene biosynthesis polypeptides are endogenously expressed in a microorganism chosen from *Cronobacter* spp, *Callitropsis* spp, *Xanthophyllomyces* spp, *Agrobacterium* spp, and *Pantoea* spp.

3130. The microorganism ~~of any one of claims~~claim 1 ~~to 3029~~, wherein the amount and/or activity of aRas2 protein has been decreased.

3231. The microorganism ~~of any one of claims~~claim 1 ~~to 3130~~, wherein the microorganism has been genetically modified to reduce or eliminate expression of an endogenous *RAS2* gene.

3332. The microorganism ~~of any one of claims~~claim 1 ~~to 3231~~, wherein the amount and/or activity of an Faa1 protein has been decreased.

3433. The microorganism ~~of any one of~~ ~~claims~~claim ~~1 to 3332~~, wherein the microorganism has been genetically modified to reduce or eliminate expression of an endogenous *FAA1* gene.

3534. The microorganism ~~of any one of~~ ~~claims~~claim ~~1 to 3433~~, wherein the amounts and/or activities of one or more proteins chosen from acetyl-CoA C-acetyltransferase, HMG-CoA synthase, HMG-CoA reductase, mevalonate kinase, phosphomevalonate kinase, diphosphomevalonate decarboxylase, and isopentyl diphosphate delta isomerase,have been increased.

3635. The microorganism ~~of any one of~~ ~~claims~~claim ~~1 to 3534~~,wherein the microorganism has been genetically modified to increase expression of one or more endogenous genes chosen from *ERG10*, *ERG13*, *HMG1*, *ERG12*, *ERG8*, *MVD1*, and *IDI1*.

3736. The microorganism ~~of any one of~~ ~~claims~~claim ~~1 to 3635~~, wherein the amount and/or activity of dimethylallyltranstransferase or farnesyl diphosphate synthetase has been increased.

3837. The microorganism ~~of any one of~~ ~~claims~~claim ~~1 to 3736~~, wherein the microorganism has been genetically modified to increase expression of an endogenous *ERG20* gene.

3938. The microorganism of claim ~~3635~~, wherein expression of *ERG10*, *ERG13*, *HMG1*, *ERG12*, *ERG8*, *MVD1*, and/or *IDI1*is regulated by a promoter region of the gene encoding oleate-induced peroxisomal protein (POX18).

4039. The microorganism of claim ~~36 or 39~~ 35 or 38, wherein expression of *ERG10*, *ERG13*, *HMG1*, *ERG12*, *ERG8*, *MVD1*, and/or *IDI1*is regulated by a terminator region of the gene encoding oleate-induced peroxisomal protein (POX18).

4140. The microorganism of claim ~~3837~~, wherein expression of*ERG20*is regulated by a promoter region of the gene encoding oleate-induced peroxisomal protein (POX18).

4241. The microorganism of claim ~~38 or 41~~ 37 or 40, wherein expression of*ERG20*is regulated by a terminator region of the gene encoding oleate-induced peroxisomal protein (POX18).

4342. A genetically modified microorganism, comprising:

one or more heterologous nucleic acids encoding one or more terpene biosynthesis polypeptides, and

a genetic modification that alters the expression of a polypeptide providing for transport of acetyl-carnitine in the microorganism.

wherein the expression of at least one of the heterologous nucleic acids is regulated by a nucleic acid that provides for glucose, fatty acid or alkane induction of expression of the terpene biosynthesis polypeptide.

~~44. The microorganism of claim 43, wherein expression of at least one of the heterologous nucleic acids is regulated by a nucleic acid that provides for fatty acid or alkane induction of expression of the terpene biosynthesis polypeptide.~~

~~45. The microorganism of claim 43, wherein expression of at least one of the heterologous nucleic acids is regulated by a nucleic acid that provides for glucose induction of expression of the terpene biosynthesis polypeptide.~~

4643. The microorganism of claim 4243, 44 or 45, wherein the microorganism is a fungus.

4744. The microorganism of claim 4346, wherein the fungus is a yeast.

4845. The microorganism of claim 4447, wherein the yeast is chosen from *Candida* spp, *Yarrowia* spp, *Rhodotorula* spp, *Rhodosporidium* spp, *Cryptococcus* spp, *Trichosporon* spp, *Lipomyces* spp, and *Blastobotrys* spp.

4946. The microorganism ~~of any one of claims~~claim 4243 to 48, wherein the one or more heterologous nucleic acids encode one or more terpene biosynthesis polypeptides chosen from terpene synthase, phytoene synthase, geranylgeranyl diphosphate synthase, phytoene desaturase, lycopene cyclase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase, zeaxanthin glucosyltransferase, valencene synthase, and cytochrome p450 reductase.

5047. The microorganism ~~of any one of claims~~claim 4643 to 49, wherein the genetic modification reduces expression of the polypeptide providing for transport of acetyl-carnitine in the microorganism relative to a microorganism that does not have the genetic modification.

5148. The microorganism of claim 4750, wherein the genetic modification is a disruption, deletion or knockout of (i) a polynucleotide that encodes a polypeptide providing for transport of acetyl-carnitine, or (ii) a promoter operably linked to a polynucleotide that encodes a polypeptide

providing for transport of acetyl-carnitine, whereby endogenous activity of a polypeptide providing for transport of acetyl-carnitine is reduced or abolished.

~~5248~~. The microorganism of claim ~~4750~~, further comprising a nucleic acid encoding a polypeptide providing for transport of acetyl-carnitine wherein expression of the nucleic acid encoding a polypeptide providing for transport of acetyl-carnitine is regulated by a promoter that provides for reduced expression relative to endogenous expression.

~~5350~~. The microorganism of claim ~~4952~~, wherein the genetic modification comprises replacing the promoter of an endogenous gene encoding the polypeptide providing for transport of acetyl-carnitine in the microorganism with a promoter that provides for reduced expression of the polypeptide in the microorganism relative to a microorganism that does not have the genetic modification.

~~5451~~. The microorganism of claim ~~52 or 5349 or 50~~, wherein the promoter that provides for reduced expression relative to endogenous expression is a promoter for a glucose-6-phosphate isomerase (G6PI) gene.

~~5552~~. The microorganism of claim ~~49 or 5053 or 53~~, wherein the promoter that provides for reduced expression relative to endogenous expression is a promoter for a *Candida* glucose-6-phosphate isomerase (G6PI) gene.

~~5653~~. The microorganism ~~of any one of claimsclaim 43 to 5542~~, wherein the polypeptide providing for transport of acetyl-carnitine in the microorganism is an acetyl-carnitine translocase (CRC).

~~5754~~. The microorganism of claim ~~5356~~, wherein the acetyl-carnitine translocase (CRC) is acetyl-carnitine translocase 1 (CRC1).

~~5855~~. The microorganism ~~of any one of claimsclaim 43 to 5742~~, wherein the microorganism is *Candida viswanathii*.

~~5956~~. The microorganism ~~of any one of claimsclaim 43 to 5742~~, wherein the microorganism is *Blastobotrys adeninivorans*.

~~6057~~. The microorganism ~~of any one of claimsclaim 43 to 5942~~, wherein the one or more heterologous nucleic acids encode phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase.

6158. The microorganism of claim 5760, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a promoter region of the gene encoding a hydratase-dehydrogenase-epimerase (HDE).

6259. The microorganism of claim 5760, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a promoter region of the gene encoding a *Candida* hydratase-dehydrogenase-epimerase (HDE).

6360. The microorganism of claim ~~60, 61 or 62~~57, 58 or 59, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a terminator region of the gene encodingan acyl co-A oxidase 4 (POX4).

6461. The microorganism of claim 57, 58 or ~~5960, 61 or 62~~, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a terminator region of the gene encodinga *Candida*acyl co-A oxidase 4 (POX4).

6562. The microorganism ~~of any one of claimsclaim 4243 to 50~~, wherein the one or more heterologous nucleic acids encode geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase.

6663. The microorganism of claim 6265, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase is regulated by a promoter region of the gene encoding a hydratase-dehydrogenase-epimerase (HDE).

6764. The microorganism of claim 6266, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase is regulated by a promoter region of the gene encoding a *Candida* hydratase-dehydrogenase-epimerase (HDE).

6865. The microorganism of claim 62, 63 or ~~6465, 66 or 67~~, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase

and bifunctional lycopene cyclase/phytoene synthase is regulated by a terminator region of the gene encodingan acyl co-A oxidase 4 (POX4).

6866. The microorganism of claim 62, 63 or 5465, 66 or 67, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase is regulated by a terminator region of the gene encodinga *Candida*acyl co-A oxidase 4 (POX4).

7067. The microorganism ~~of any one of claims~~claim 4243 to 69, wherein the one or more heterologous nucleic acids encoding the one or more terpene biosynthesis polypeptides are endogenously expressed in a microorganism chosen from *Cronobacter* spp, *Callitropsis* spp, *Xanthophyllomyces* spp, *Agrobacterium* spp, and *Pantoea* spp.

7168. The microorganism ~~of any one of claims~~claim 4243 to 70, wherein the amount and/or activity of aRas2 protein has been decreased.

7269. The microorganism ~~of any one of claims~~claim 4243 to 71, wherein the microorganism has been genetically modified to reduce or eliminate expression of an endogenous *RAS2* gene.

7370. The microorganism ~~of any one of claims~~claim 4243 to 72, wherein the amount and/or activity of an Faa1 protein has been decreased.

7471. The microorganism ~~of any one of claims~~claim 4243 to 73, wherein the microorganism has been genetically modified to reduce or eliminate expression of an endogenous *FAA1* gene.

7572. The microorganism ~~of any one of claims~~claim 4243 to 74, wherein the amounts and/or activities of one or more proteins chosen from acetyl-CoA C-acetyltransferase, HMG-CoA synthase, HMG-CoA reductase, mevalonate kinase, phosphomevalonate kinase, diphosphomevalonate decarboxylase, and isopentyl diphosphate delta isomerase, have been increased.

7673. The microorganism ~~of any one of claims~~claim 4243 to 75, wherein the microorganism has been genetically modified to increase expression of one or more endogenous genes chosen from *ERG10*, *ERG13*, *HMG1*, *ERG12*, *ERG8*, *MVD1*, and *IDI1*.

7774. The microorganism ~~of any one of claims~~claim 4243 to 76, wherein the amount and/or activity of dimethylallyltranstransferase or farnesyl diphosphate synthetase has been increased.

~~75~~. The microorganism ~~of any one of claims~~claim ~~42~~43 ~~to 77~~, wherein the microorganism has been genetically modified to increase expression of an endogenous *ERG20* gene.

~~76~~. The microorganism of claim ~~73~~, wherein expression of *ERG10*, *ERG13*, *HMG1*, *ERG12*, *ERG8*, *MVD1*, and/or *IDI1* is regulated by a promoter region of the gene encoding oleate-induced peroxisomal protein (POX18).

~~77~~. The microorganism of claim 73 or ~~76~~ ~~or 79~~, wherein expression of *ERG10*, *ERG13*, *HMG1*, *ERG12*, *ERG8*, *MVD1*, and/or *IDI1* is regulated by a terminator region of the gene encoding oleate-induced peroxisomal protein (POX18).

~~78~~. The microorganism of claim ~~75~~, wherein expression of *ERG20* is regulated by a promoter region of the gene encoding oleate-induced peroxisomal protein (POX18).

~~79~~. The microorganism of claim 75 or ~~78~~ ~~or 81~~, wherein expression of *ERG20* is regulated by a terminator region of the gene encoding oleate-induced peroxisomal protein (POX18).

~~80~~. A genetically modified *Candida viswanathii* yeast, comprising one or more heterologous nucleic acids encoding one or more terpene biosynthesis polypeptides, wherein expression of at least one of the heterologous nucleic acids is regulated by a nucleic acid that provides for glucose, fatty acid or alkane induction of expression of the terpene biosynthesis polypeptide and

- wherein the one or more heterologous nucleic acids encoding one or more terpene biosynthesis polypeptides is chosen from terpene synthase, phytoene synthase, geranylgeranyl diphosphate synthase, phytoene desaturase, lycopene cyclase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase, zeaxanthin glucosyltransferase and valencene synthase, or

- wherein the one or more heterologous nucleic acids is a cytochrome P450 reductase and is regulated by a nucleic acid that provides for glucose or alkane induction of expression of the cytochrome P450 reductase.

~~81. The genetically modified *Candida viswanathii* yeast of claim 80, wherein expression of at least one of the heterologous nucleic acids is regulated by a nucleic acid that provides for fatty acid or alkane induction of expression of the terpene biosynthesis polypeptide.~~

P021168WO_20190608_AMD claims marked up

~~85. The genetically modified *Candida viswanathii* yeast of claim 83, wherein expression of at least one of the heterologous nucleic acids is regulated by a nucleic acid that provides for glucose induction of expression of the terpene biosynthesis polypeptide.~~

~~86. The genetically modified *Candida viswanathii* yeast of claim 83, 84 or 85, wherein the one or more heterologous nucleic acids encode one or more terpene biosynthesis polypeptides chosen from terpene synthase, phytoene synthase, geranylgeranyl diphosphate synthase, phytoene desaturase, lycopene cyclase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase, zeaxanthin glucosyltransferase, valencene synthase, and cytochrome p450 reductase.~~

8781. The genetically modified *Candida viswanathii* yeast ~~of any one of claims~~claim 8083 to~~86~~, wherein the one or more heterologous nucleic acids encode phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase

8882. The genetically modified *Candida viswanathii* yeast of claim 8187, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a promoter region of the gene encoding a *Candida* hydratase-dehydrogenase-epimerase (HDE).

8983. The genetically modified *Candida viswanathii* yeast of claim 8288, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a terminator region of the gene encoding a *Candida*acyl co-A oxidase 4 (POX4).

9084. The genetically modified *Candida viswanathii* yeast of claim 8187, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a promoter region of the gene encoding a *Candida* glyceraldehyde-3-phosphate dehydrogenase (GPD).

9185. The genetically modified *Candida viswanathii* yeast ~~of any one of claims~~claim 8083 to~~86~~, wherein the one or more heterologous nucleic acids encode valencene synthase.

9286. The genetically modified *Candida viswanathii* yeast of claim 8591, wherein expression of the heterologous nucleic acids encoding valencene synthase is regulated by a promoter region of the gene encoding a *Candida* hydratase-dehydrogenase-epimerase (HDE).

P021168WO_20190608_AMD claims marked up

9387. The genetically modified *Candida viswanathii* yeast of claim 8682, wherein expression of the heterologous nucleic acids valencene synthase is regulated by a terminator region of the gene encoding a *Candida*acyl co-A oxidase 4 (POX4).

9488. The genetically modified *Candida viswanathii* yeast of claim 8684, wherein expression of the heterologous nucleic acids encoding valencene synthase is regulated by a promoter region of the gene encoding a *Candida* glyceraldehyde-3-phosphate dehydrogenase (GPD).

9589. The genetically modified *Candida viswanathii* yeast ~~of any one~~ of ~~claimsclaim~~ 8683-8686, wherein the one or more heterologous nucleic acids encode geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase.

9690. The genetically modified *Candida viswanathii* yeast of claim 8685, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase is regulated by a promoter region of the gene encoding a *Candida* hydratase-dehydrogenase-epimerase (HDE).

9791. The genetically modified *Candida viswanathii* yeast of claim 8685, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase is regulated by a terminator region of the gene encoding a *Candida*acyl co-A oxidase 4 (POX4).

9892. The genetically modified *Candida viswanathii* yeast of claim 8685, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase is regulated by a promoter region of the gene encoding a *Candida* glyceraldehyde-3-phosphate dehydrogenase (GPD).

9993. The genetically modified *Candida viswanathii* yeast ~~of any one~~ of ~~claimsclaim~~ 8683-8686, wherein the one or more heterologous nucleic acids encode geranylgeranyl diphosphate synthase, phytoene desaturase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase,and cytochrome p450 reductase.

10094. The genetically modified *Candida viswanathii* yeast of claim 8688, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase,and cytochrome p450 reductase is regulated by a promoter region of the gene encoding a *Candida* hydratase-dehydrogenase-epimerase (HDE).

~~10495~~. The genetically modified *Candida viswanathii* yeast of claim ~~94~~100, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase,and cytochrome p450 reductase is regulated by a terminator region of the gene encoding a *Candida*acyl co-A oxidase 4 (POX4).

~~10296~~. The genetically modified *Candida viswanathii* yeast of claim ~~93~~98, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase,and cytochrome p450 reductase is regulated by a promoter region of the gene encoding a *Candida* glyceraldehyde-3-phosphate dehydrogenase (GPD).

~~10397~~. The genetically modified *Candida viswanathii* yeast ~~of any one of claims~~claim ~~80~~83 to ~~102~~, wherein the one or more heterologous nucleic acids encoding the one or more terpene biosynthesis polypeptides are endogenously expressed in a microorganism chosen from *Cronobacter* spp, *Callitropsis* spp, *Xanthophyllomyces* spp, *Agrobacterium* spp, and *Pantoea* spp.

~~10498~~. The genetically modified *Candida viswanathii* yeast ~~of any one of claims~~claim ~~80~~83 to ~~103~~, further comprising a genetic modification that alters the expression of one or more nucleic acids encoding one or more endogenous polypeptides.

~~10599~~. The genetically modified *Candida viswanathii* yeast of claim ~~98~~104, wherein the one or more endogenous polypeptides are chosen from polypeptides having one or more of the following activities: acyl-CoA synthetase activity, acyl-CoA oxidase activity, ATP-binding cassette transporter activity, carnitine acetyltransferase activity, transport of acetyl-carnitine, acyl-CoA thioesterase activity, acyl-CoA hydrolase activity, aldehyde dehydrogenase activity, monooxygenase activity, or monooxgenase reductase activity.

~~106~~100. The genetically modified *Candida viswanathii* yeast ~~of any one of claims~~claim ~~80~~83 to ~~105~~, wherein the amount and/or activity of aRas2 protein has been decreased.

~~107~~101. The genetically modified *Candida viswanathii*yeast ~~of any one of claims~~claim ~~80~~83 to ~~106~~, wherein the yeast has been genetically modified to reduce or eliminate expression of an endogenous *RAS2* gene.

~~108~~102. The genetically modified *Candida viswanathii* yeast ~~of any one of claims~~claim ~~80~~83 ~~to 107~~, wherein the amount and/or activity of an Faa1 protein has been decreased.

~~109~~103. The genetically modified *Candida viswanathii* yeast ~~of any one of claims~~claim ~~80~~83 ~~to 108~~, wherein the yeast has been genetically modified to reduce or eliminate expression of an endogenous *FAA1* gene.

~~110~~104. The genetically modified *Candida viswanathii* yeast ~~of any one of claims~~claim ~~80~~83 ~~to 109~~, wherein the amounts and/or activities of one or more proteins chosen from acetyl-CoA C-acetyltransferase, HMG-CoA synthase, HMG-CoA reductase, mevalonate kinase, phosphomevalonate kinase, diphosphomevalonate decarboxylase, and isopentyl diphosphate delta isomerase, have been increased.

~~111~~105. The genetically modified *Candida viswanathii* yeast ~~of any one of claims~~claim ~~80~~83 ~~to 110~~, wherein the yeast has been genetically modified to increase expression of one or more endogenous genes chosen from *ERG10*, *ERG13*, *HMG1*, *ERG12*, *ERG8*, *MVD1*, and *IDI1*.

~~112~~106. The genetically modified *Candida viswanathii* yeast ~~of any one of claims~~claim ~~80~~83 ~~to 111~~, wherein the amount and/or activity of dimethylallyltranstransferase or farnesyl diphosphate synthetase has been increased.

~~113~~107. The genetically modified *Candida viswanathii* yeast ~~of any one of claims~~claim ~~80~~83 ~~to 112~~, wherein the yeast has been genetically modified to increase expression of an endogenous *ERG20* gene.

~~114~~108. The genetically modified *Candida viswanathii* yeast of claim ~~111~~105, wherein expression of *ERG10*, *ERG13*, *HMG1*, *ERG12*, *ERG8*, *MVD1*, and/or *IDI1* is regulated by a promoter region of the gene encoding oleate-induced peroxisomal protein (POX18).

~~115~~109. The genetically modified *Candida viswanathii* yeast of claim 105 or ~~108~~~~111 or 114~~, wherein expression of *ERG10*, *ERG13*, *HMG1*, *ERG12*, *ERG8*, *MVD1*, and/or *IDI1* is regulated by a terminator region of the gene encoding oleate-induced peroxisomal protein (POX18).

~~116~~110. The genetically modified *Candida viswanathii* yeast of claim 107~~113~~, wherein expression of *ERG20* is regulated by a promoter region of the gene encoding oleate-induced peroxisomal protein (POX18).

~~147~~111. The genetically modified *Candida viswanathii* yeast of claim 107 or 110~~143 or 146~~, wherein expression of *ERG20* is regulated by a terminator region of the gene encoding oleate-induced peroxisomal protein (POX18).

~~148~~112. A method for producing a terpene comprising:

contacting the genetically modified microorganism of any one of claims 1 to ~~41~~42 with a feedstock comprising a carbon source, and

culturing the microorganism under conditions in which the terpenes are produced from the feedstock.

~~149~~113. A method for producing a terpene comprising:

contacting the genetically modified microorganism of any one of claims 42 to 79~~43 to 82~~ with a feedstock comprising a carbon source, and

culturing the microorganism under conditions in which the terpenes are produced from the feedstock.

~~150~~114. A method for producing a terpene comprising:

contacting the genetically modified *Candida viswanathii*yeast of any one of claims 80 to 111~~83 to 147~~ with a feedstock comprising a carbon source, and

culturing the microorganism under conditions in which the terpenes are produced from the feedstock.

~~151~~115. The method ~~of any one of~~ claim~~claim~~ 112~~148 to 150~~, wherein the feedstock comprises one or more fatty acids.

~~152~~116. The method of claim 115~~151~~, wherein the feedstock comprises one or more saturated fatty acids or one or more unsaturated fatty acids; or one or more saturated fatty acids and one or more unsaturated fatty acids.

~~153~~117. The method of claim 116~~151 or 152~~, wherein the feedstock comprises one or more fatty acids chosen from oleic acid, palmitoleic acid, erucic acid, linoleic acid, palmitic acid, caproic acid, enanthic acid, caprylic acid pelargonic acid, capric acid, undecylic acid, lauric acid, myristic acid, pentadecanoic acid, margaric acid, stearic acid arachidic acid, behenic acid, tridecylic acid, and linolenic acid.

~~154~~118. The method of claim 117~~153~~ wherein the feedstock comprises oleic acid.

~~125~~119. The method ~~of any one of claims~~claim ~~112~~118 ~~to 120~~, wherein the feedstock comprises one or more sugars.

~~126~~120. The method of ~~119~~125, wherein the feedstock comprises glucose.

~~127~~121. The method ~~of any one of claims~~claim ~~112~~118 ~~to 120~~, wherein the feedstock comprises one or more alkane hydrocarbons.

~~128~~122. The method of claim ~~121~~127, wherein the feedstock comprises one or more alkane hydrocarbons chosen from one or more ofhexane, heptane, nonane, decane, undecane, dodecane, tridecane, tetradecane, pentadecane, hexadecane, heptadecane, and octadecane.

~~129~~123. The method of claim ~~121~~127, wherein the feedstock comprises one or more alkane hydrocarbons chosen from C10 to C14 alkane hydrocarbons, nonane and octadecane.

~~130~~124. The method ~~of any one of claims~~claim ~~112~~118 ~~to 120~~, wherein the feedstock comprises one or more vegetable oils, derivatives thereof, or byproducts thereof.

~~131~~125. The method of claim ~~124~~130, wherein the feedstock comprises crude palm oil (CPO).

~~132~~126. The method of claim ~~124~~130, wherein the feedstock comprises palm oil fatty acid distillate (PFAD).

~~133~~127. The method ~~of any one of claims~~claim ~~112~~118 ~~to 132~~, wherein the terpene is chosen from lycopene, beta carotene and astaxanthin.

~~134~~128. The method ~~of any one of claims~~claim ~~127~~118 ~~to 133~~, wherein the yield or titer of the terpene is between about 0.001 g/L to about 2.0 g/L.

~~135~~129. The method ~~of any one of claims~~claim ~~127~~118 ~~to 133~~, wherein the yield or titer of the terpene is between about 0.05 g/L to about 0.5 g/L.

~~136~~130. The method ~~of any one of claims~~claim ~~127~~118 ~~to 133~~, wherein the yield or titer of the terpene is between about 0.001 g/L to about 0.2 g/L.

~~137~~131. The method of claim ~~128~~134, wherein the yield or titer of lycopene is between about 0.03 g/L to about 2.0 g/L.

~~138~~132. The method of claim 129~~135~~, wherein the yield or titer of lycopene is between about 0.05 g/L to about 0.5 g/L.

~~139~~133. The method of claim 130~~136~~, wherein the yield or titer of lycopene is between about 0.001 g/L to about 0.2 g/L.

~~140~~134. The method of claim 129~~134~~, wherein the yield or titer of beta carotene is between about 0.1 g/L to about 0.2 g/L.

~~141~~135. The method of claim 130~~136~~, wherein the yield or titer of beta carotene is between about 0.01 g/L to about 0.1 g/L.

~~142~~136. The method of claim 130~~136~~, wherein the yield or titer of beta carotene is between about 0.001 g/L to about 0.01 g/L.

137~~143~~. The method of claim 129~~134~~, wherein the yield or titer of astaxanthin is between about 0.005 g/L to about 0.01 g/L.

~~144~~138. The method of claim ~~136~~130, wherein the yield or titer of astaxanthin is between about 0.001 g/L to about 0.005 g/L.

~~145~~139. The method ~~of any one of claims~~claim 112~~113~~ to ~~144~~, further comprising isolating the terpene.

~~146~~140. A genetically modified yeast comprising one or more heterologous nucleic acids encodingphytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase, wherein the amount and/or activity of a Ras2 protein has been decreased.

~~147~~141. A genetically modified yeast comprising one or more heterologous nucleic acids encodinggeranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase, wherein the amount and/or activity of a Ras2 protein has been decreased.

~~148~~142. A genetically modified yeast comprising one or more heterologous nucleic acids encodinggeranylgeranyl diphosphate synthase, phytoene desaturase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin

synthase,and cytochrome p450 reductase, wherein the amount and/or activity of a Ras2 protein has been decreased.

~~148. The genetically modified yeast of claim 146, 147 or 148, wherein the amount and/or activity of a Ras2 protein has been decreased.~~

~~149~~143. The genetically modified yeast ~~of any one~~ of ~~claims~~claim ~~142~~146 ~~to 148~~, wherein the yeast has been genetically modified to reduce or eliminate expression of an endogenous *RAS2* gene.

~~150~~144. The genetically modified yeast ~~of any one~~ of ~~claims~~claim ~~142~~146 ~~to 149~~, wherein the amount and/or activity of an Faa1 protein has been decreased.

~~151~~145. The genetically modified yeast ~~of any one~~ of ~~claims~~claim ~~142~~146 ~~to 150~~, wherein the yeast has been genetically modified to reduce or eliminate expression of an endogenous *FAA1* gene.

~~152~~146. The genetically modified yeast ~~of any one~~ of ~~claims~~claim ~~142~~146 ~~to 151~~, wherein the amounts and/or activities of one or more proteins chosen from acetyl-CoA C-acetyltransferase, HMG-CoA synthase, HMG-CoA reductase, mevalonate kinase, phosphomevalonate kinase, diphosphomevalonate decarboxylase, and isopentyl diphosphate delta isomerase,have been increased.

~~153~~147. The genetically modified yeast ~~of any one~~ of ~~claims~~claim ~~142~~146 ~~to 152~~, wherein the yeast has been genetically modified to increase expression of one or more endogenous genes chosen from *ERG10*, *ERG13*, *HMG1*, *ERG12*, *ERG8*, *MVD1*, and *IDI1*.

~~154~~148. The genetically modified yeast ~~of any one~~ of ~~claims~~claim ~~142~~146 ~~to 153~~, wherein the amount and/or activity of dimethylallyltranstransferase or farnesyl diphosphate synthetase has been increased.

~~155~~149. The genetically modified yeast ~~of any one~~ of ~~claims~~claim ~~142~~146 ~~to 154~~, wherein the yeast has been genetically modified to increase expression of an endogenous *ERG20* gene.

~~156~~150. The genetically modified yeast ~~of any one~~ of ~~claims~~claim ~~142~~146 ~~to 155~~, wherein the yeast is *Candida viswanathii*.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 547 of 3077

P021168WO_20190608_AMD claims marked up

~~158~~151. The genetically modified yeast ~~of any one of~~ claim~~claim~~ 142~~146 to 156~~, wherein the yeast is *Blastobotrys adeninivorans*.

~~159~~152. Use of the genetically modified yeast of any one of claims ~~146 to 158~~140 to 151 for the production of a terpene.

~~160~~153. Use of the genetically modified yeast of any one of claims 140~~146~~ and ~~146 to 159~~143-151 for the production of lycopene.

~~161~~154. Use of the genetically modified yeast ~~of any one of~~ claim~~claim~~ 141~~147~~ and ~~146 to 159~~143-151 for the production of beta carotene.

~~162~~155. Use of the genetically modified yeast of any one of claims ~~146 to 159~~142-151 for the production of astaxanthin.

PATENT COOPERATION TREATY

From the INTERNATIONAL SEARCHING AUTHORITY

To: Kari A Dickinson
Grant IP, Inc.
c/o CPA Global
900 Second Avenue S., Suite 600
Minneapolis, MN 55402
United States of America

PCT

NOTIFICATION OF TRANSMITTAL OF THE INTERNATIONAL SEARCH REPORT AND THE WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY, OR THE DECLARATION

(PCT Rule 44.1)

	Date of mailing *(day/month/year)* 13 SEP 2018
Applicant's or agent's file reference VAL-3002-PC	**FOR FURTHER ACTION** See paragraphs 1 and 4 below
International application No. PCT/US18/41579	International filing date *(day/month/year)* 11 July 2018 (11.07.2018)
Applicant Verdezyne (ABC), LLC	

1. ☒ The applicant is hereby notified that the international search report and the written opinion of the International Searching Authority have been established and are transmitted herewith.

   Filing of amendments and statement under Article 19:
   The applicant is entitled, if he so wishes, to amend the claims of the international application (see Rule 46):

   When? The time limit for filing such amendments is normally two months from the date of transmittal of the international search report.

   How? Directly to the International Bureau of WIPO preferably through ePCT or on paper to, 34 chemin des Colombettes 1211 Geneva 20, Switzerland, Facsimile No.: +41 22 338 82 70

   For more detailed instructions, see *PCT Applicant's Guide*, International Phase, paragraphs 9.004 – 9.011.

2. ☐ The applicant is hereby notified that no international search report will be established and that the declaration under Article 17(2)(a) to that effect and the written opinion of the International Searching Authority are transmitted herewith.

3. ☐ **With regard to any protest** against payment of (an) additional fee(s) under Rule 40.2, the applicant is notified that:
   ☐ the protest together with the decision thereon has been transmitted to the International Bureau together with any request to forward the texts of both the protest and the decision thereon to the designated Offices.
   ☐ no decision has been made yet on the protest; the applicant will be notified as soon as a decision is made.

4. **Reminders**

   The applicant may **submit comments on an informal basis on the written opinion of the International Searching Authority** to the International Bureau. These comments will be made available to the public after international publication. The International Bureau will send a copy of such comments to all designated Offices unless an international preliminary examination report has been or is to be established.

   Shortly after the expiration of **18 months from the priority date, the international application will be published** by the International Bureau. If the applicant wishes to avoid or postpone publication, a notice of withdrawal of the international application, or of the priority claim, must reach the International Bureau before the completion of the technical preparations for international publication (Rules 90*bis*.1 and 90*bis*.3).

   Within **19 months** from the priority date, but only in respect of some designated Offices, a demand for international preliminary examination must be filed if the applicant wishes to postpone the entry into the national phase **until 30 months** from the priority date (in some Offices even later); otherwise, the applicant must, **within 20 months** from the priority date, perform the prescribed acts for **entry into the national phase** before those designated Offices. In respect of other designated Offices, the time limit of **30 months** (or later) will apply even if no demand is filed within 19 months. For details about the applicable time limits, Office by Office, see www.wipo.int/pct/en/texts/time_limits.html and the *PCT Applicant's Guide*, National Chapters.

   Within **22 months from the priority date, the applicant may request that a supplementary international search be carried out** by a different International Searching Authority that offers this service (Rule 45*bis*.1). The procedure for requesting supplementary international search is described in the *PCT Applicant's Guide*, International Phase, paragraphs 8.006-8.032.

Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Authorized officer Shane Thomas PCT Helpdesk: 571-272-4300 Telephone No. PCT OSP: 571-272-7774

Form PCT/ISA/220 (July 2017)

PATENT COOPERATION TREATY

PCT

INTERNATIONAL SEARCH REPORT

(PCT Article 18 and Rules 43 and 44)

Applicant's or agent's file reference VAL-3002-PC	**FOR FURTHER ACTION**	see Form PCT/ISA/220 as well as, where applicable, item 5 below.
International application No. PCT/US18/41579	International filing date *(day/month/year)* 11 July 2018 (11.07.2018)	(Earliest) Priority Date *(day/month/year)* 13 July 2017 (13.07.2017)
Applicant Verdezyne (ABC), LLC		

This international search report has been prepared by this International Searching Authority and is transmitted to the applicant according to Article 18. A copy is being transmitted to the International Bureau.

This international search report consists of a total of 4 sheets.

☐ It is also accompanied by a copy of each prior art document cited in this report.

1. **Basis of the report**
   a. With regard to the **language**, the international search was carried out on the basis of:
      ☒ the international application in the language in which it was filed.
      ☐ a translation of the international application into _____ which is the language of a translation furnished for the purposes of international search (Rules 12.3(a) and 23.1(b)).
   b. ☐ This international search report has been established taking into account the **rectification of an obvious mistake** authorized by or notified to this Authority under Rule 91 (Rule 43.6*bis*(a)).
   c. ☐ With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, see Box No. I.

2. ☒ **Certain claims were found unsearchable** (see Box No. II).

3. ☐ **Unity of invention is lacking** (see Box No. III).

4. With regard to the **title**,
   ☒ the text is approved as submitted by the applicant.
   ☐ the text has been established by this Authority to read as follows:

5. With regard to the **abstract**,
   ☒ the text is approved as submitted by the applicant.
   ☐ the text has been established, according to Rule 38.2, by this Authority as it appears in Box No. IV. The applicant may, within one month from the date of mailing of this international search report, submit comments to this Authority.

6. With regard to the **drawings**,
   a. the figure of the **drawings** to be published with the abstract is Figure No. 1
      ☒ as suggested by the applicant.
      ☐ as selected by this Authority, because the applicant failed to suggest a figure.
      ☐ as selected by this Authority, because this figure better characterizes the invention.
   b. ☐ none of the figures is to be published with the abstract.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 550 of 3077

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US18/41579

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☒ Claims Nos.: 6-42, 49-82, 86-145, 150-162
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 551 of 3077

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US18/41579

A. CLASSIFICATION OF SUBJECT MATTER

IPC - C07H 21/04; C07K 14/415; C12N 1/16, 1/19, 1/21, 5/10, 9/02, 9/88, 9/96, 15/09, 15/52 (2018.01)

CPC - C12N 1/16, 5/10, 9/88, 9/96, 15/00, 15/09, 15/52, 15/63, 15/70, 15/74, 15/82, 15/8249; C07H 21/04; C07K 14/415; C12P 7/02, 7/26

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
See Search History document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
See Search History document

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
See Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X -- Y	WO 2014/100504 A2 (VERDEZYNE, INC.) 26 June 2014; page 9, 1st paragraph; page 10, 1st paragraph; page 27, 3rd paragraph; page 41, 1st paragraph; page 143, 2nd paragraph	1-4, 5/1-4, 83-84 --- 44, 46/44, 47/46/44, 48/47/46/44, 85, 149/146-148
X -- Y	JP 2012-254044 A (KIRIN HOLDING CO LTD) 27 December 2012; page 5, 6th paragraph; page 6, 5th paragraph; page 11, 3rd paragraph; page 12, 1st paragraph; claim 10	43, 46/43, 47/46/43, 48/47/46/43, 146-148 --- 44-45, 46/44-45, 47/46/44-45, 48/47/46/44-45
Y	US 2014/0113343 A1 (DAVIET, L et al.) 24 April 2014; paragraphs [0038], [0044]	45, 46/45, 47/46/45, 48/47/46/45, 85
Y	US 2003/0044946 A1 (LONGO, VD) 06 March 2003; paragraph [0093]	149/146-148
A	US 2015/0010978 A1 (SAPPHIRE ENERGY, INC.) 08 January 2015; entire document	1-4, 5/1-4, 43-45, 46/43-45, 47/46/43-45, 48/47/46/43-45, 83-85, 146-148, 149/146-148

☒ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
4 September 2018 (04.09.2018)	13 SEP 2018
Name and mailing address of the ISA/ Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Authorized officer Shane Thomas PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (January 2015)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US18/41579

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2013/0066035 A1 (BURGARD, AP et al.) 14 March 2013; entire document	1-4, 5/1-4, 43-45, 46/43-45, 47/46/43-45, 48/47/46/43-45, 83-85, 146-148, 149/146-148
A	US 2013/0302861 A1 (YISSUM RESEARCH DEVELOPMENT COMPANY OF THE HEBREW UNIVERSITY OF JERUSALEM LTD.) 14 November 2013; entire document	1-4, 5/1-4, 43-45, 46/43-45, 47/46/43-45, 48/47/46/43-45, 83-85, 146-148, 149/146-148
P, X	WO 2018/094110 A2 (THE BOARD OF TRUSTEES OF THE LELAND STANFORD JUNIOR UNIVERSITY) 24 May 2018; entire document	1-4, 5/1-4, 43-45, 46/43-45, 47/46/43-45, 48/47/46/43-45, 83-85, 146-148, 149/146-148

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 553 of 3077

PATENT COOPERATION TREATY

From the
INTERNATIONAL SEARCHING AUTHORITY

To: Kari A Dickinson
Grant IP, Inc.
c/o CPA Global
900 Second Avenue S., Suite 600
Minneapolis, MN 55402
United States of America

PCT

WRITTEN OPINION OF THE
INTERNATIONAL SEARCHING AUTHORITY

(PCT Rule 43*bis*.1)

Date of mailing *(day/month/year)*	13 SEP 2018
Applicant's or agent's file reference VAL-3002-PC	**FOR FURTHER ACTION** See paragraph 2 below

International application No.	International filing date *(day/month/year)*	Priority date *(day/month/year)*
PCT/US18/41579	11 July 2018 (11.07.2018)	13 July 2017 (13.07.2017)

International Patent Classification (IPC) or both national classification and IPC
IPC - C07H 21/04; C07K 14/415; C12N 1/16, 1/19, 1/21, 5/10, 9/02, 9/88, 9/96, 15/09, 15/52 (2018.01)
CPC - C12N 1/16, 5/10, 9/88, 9/96, 15/00, 15/09, 15/52, 15/63, 15/70, 15/74, 15/82, 15/8249; C07H 21/04; C07K 14/415; C12P 7/02, 7/26

Applicant Verdezyne (ABC), LLC

1. This opinion contains indications relating to the following items:

- ☒ Box No. I Basis of the opinion
- ☐ Box No. II Priority
- ☒ Box No. III Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
- ☐ Box No. IV Lack of unity of invention
- ☒ Box No. V Reasoned statement under Rule 43*bis*.1(a)(i) with regard to novelty, inventive step and industrial applicability; citations and explanations supporting such statement
- ☒ Box No. VI Certain documents cited
- ☐ Box No. VII Certain defects in the international application
- ☐ Box No. VIII Certain observations on the international application

2. **FURTHER ACTION**

If a demand for international preliminary examination is made, this opinion will be considered to be a written opinion of the International Preliminary Examining Authority ("IPEA") except that this does not apply where the applicant chooses an Authority other than this one to be the IPEA and the chosen IPEA has notified the International Bureau under Rule 66.1*bis*(b) that written opinions of this International Searching Authority will not be so considered.

If this opinion is, as provided above, considered to be a written opinion of the IPEA, the applicant is invited to submit to the IPEA a written reply together, where appropriate, with amendments, before the expiration of 3 months from the date of mailing of Form PCT/ISA/220 or before the expiration of 22 months from the priority date, whichever expires later.

For further options, see Form PCT/ISA/220.

Name and mailing address of the ISA/US	Date of completion of this opinion	Authorized officer
Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	4 September 2018 (04.09.2018)	Shane Thomas PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/237 (cover sheet) (January 2015)

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US18/41579

Box No. I Basis of this opinion

1. With regard to the **language**, this opinion has been established on the basis of:

   ☒ the international application in the language in which it was filed.

   ☐ a translation of the international application into _____ which is the language of a translation furnished for the purposes of international search (Rules 12.3(a) and 23.1(b)).

2. ☐ This opinion has been established taking into account the **rectification of an obvious mistake** authorized by or notified to this Authority under Rule 91 (Rule 43*bis*.1(a)).

3. ☐ With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, this opinion has been established on the basis of a sequence listing:

   a. ☐ forming part of the international application as filed:

      ☐ in the form of an Annex C/ST.25 text file.

      ☐ on paper or in the form of an image file.

   b. ☐ furnished together with the international application under PCT Rule 13*ter*.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.

   c. ☐ furnished subsequent to the international filing date for the purposes of international search only:

      ☐ in the form of an Annex C/ST.25 text file (Rule 13*ter*.1(a)).

      ☐ on paper or in the form of an image file (Rule 13*ter*.1(b) and Administrative Instructions, Section 713).

4. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

5. Additional comments:

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 555 of 3077

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US18/41579

Box No. III Non-establishment of opinion with regard to novelty, inventive step and industrial applicability

The questions whether the claimed invention appears to be novel, to involve an inventive step (to be non obvious), or to be industrially applicable have not been examined in respect of:

☐ the entire international application.

☒ claims Nos. 6-42, 49-82, 86-145, 150-162

because:

☐ the said international application, or the said claims Nos. _____ relate to the following subject matter which does not require an international search *(specify)*:

☒ the description, claims or drawings *(indicate particular elements below)* or said claims Nos. 6-42, 49-82, 86-145, 150-162 are so unclear that no meaningful opinion could be formed *(specify)*:

Claim Nos. 6-42, 49-82, 86-145, and 150-162 are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

☐ the claims, or said claims Nos. _____ are so inadequately supported by the description that no meaningful opinion could be formed *(specify)*:

☒ no international search report has been established for said claims Nos. 6-42, 49-82, 86-145, 150-162

☐ a meaningful opinion could not be formed without the sequence listing; the applicant did not, within the prescribed time limit:

- ☐ furnish a sequence listing in the form of an Annex C/ST.25 text file, and such listing was not available to the International Searching Authority in the form and manner acceptable to it; or the sequence listing furnished did not comply with the standard provided for in Annex C of the Administrative Instructions.
- ☐ furnish a sequence listing on paper or in the form of an image file complying with the standard provided for in Annex C of the Administrative Instructions, and such listing was not available to the International Searching Authority in the form and manner acceptable to it; or the sequence listing furnished did not comply with the standard provided for in Annex C of the Administrative Instructions.
- ☐ pay the required late furnishing fee for the furnishing of a sequence listing in response to an invitation under Rule 13*ter*.1(a) or (b).

☐ See Supplemental Box for further details.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 556 of 3077

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US18/41579

Box No. V **Reasoned statement under Rule 43*bis*.1(a)(i) with regard to novelty, inventive step and industrial applicability; citations and explanations supporting such statement**

1. Statement

Novelty (N)	Claims	44-45, 46/44-45, 47/46/44-45, 48/47/46/44-45, 85, 149/146-148	YES
	Claims	1-4, 5/1-4, 43, 46/43, 47/46/43, 48/47/46/43, 83-84, 146-148	NO
Inventive step (IS)	Claims	NONE	YES
	Claims	-***-Please see below-***-	NO
Industrial applicability (IA)	Claims	-***-Please see below-***-	YES
	Claims	NONE	NO

2. Citations and explanations:

-***-Continued from Box No. V. 1. Statement-***-

Inventive step (IS): NO: Claims 1-4, 5/1-4, 43-45, 46/43-45, 47/46/43-45, 48/47/46/43-45, 83-85, 146-148, 149/146-148

Industrial Applicability (IA): YES: Claims Yes: 1-4, 5/1-4, 43-45, 46/43-45, 47/46/43-45, 48/47/46/43-45, 83-85, 146-148, 149/146-148

-***-Box No. V: 2. Citations and Explanations-***-

Claims 1-4, 5/1-4, and 83-84 lack novelty under PCT Article 33(2) as being anticipated by WO 2014/100504 A2 (VERDEZYNE, INC.) (hereinafter "Verdezyne").

As per claim 1, Verdezyne discloses a genetically modified microorganism (a host microorganism, engineered microorganism, genetically modified organism; page 10, 1st paragraph), comprising: one or more heterologous nucleic acids (microorganisms are engineered to contain at least one heterologous gene encoding an enzyme; page 9, 1st paragraph) encoding one or more terpene biosynthesis polypeptides (a cytochrome P450 reductase (a terpene biosynthesis polypeptide as defined by the instant application; page 530, lines 6-10) can be added to generate an engineered organism; page 27, 3rd paragraph), wherein expression of at least one of the heterologous nucleic acids is regulated by a nucleic acid that provides for fatty acid or alkane induction of expression of the terpene biosynthesis polypeptide (expressed under the control of the POX4 promoter (fatty acid inducible promoter as defined by the instant application (the nucleic acid that provides for fatty acid induction of expression of a terpene biosynthesis polypeptide comprises a promoter region chosen from promoter regions of genes encoding hydratase-dehydrogenaseepimerase, acyl co-A oxidase, POX; page 531, lines 1-4); page 41, 1st paragraph).

As per claim 2, Verdezyne discloses the microorganism of claim 1, and Verdezyne further discloses wherein the microorganism is a fungus (Candida viswanathii; page 41, 1st paragraph).

As per claim 3, Verdezyne discloses the microorganism of claim 2, and Verdezyne further discloses wherein the fungus is a yeast (Candida viswanathii; page 41, 1st paragraph).

As per claim 4, Verdezyne discloses the microorganism of claim 3, and Verdezyne further discloses wherein the yeast is chosen from Candida spp (Candida viswanathii; page 41, 1st paragraph), Yarrowia spp, Rhodotorula spp, Rhodosporidium spp, Cryptococcus spp, Trichosporon spp, Lipomyces spp, and Blastobotrys spp.

As per claims 5/1-4, Verdezyne discloses the microorganism of claims 1-4, and Verdezyne further discloses wherein the one or more heterologous nucleic acids encode one or more terpene biosynthesis polypeptides chosen from terpene synthase, phytoene synthase, geranylgeranyl diphosphate synthase, phytoene desaturase, lycopene cyclase, bifunctional lycopene cyclase/phytoene synthase, beta-carotene ketolase, beta-carotene hydroxylase, astaxanthin synthase, zeaxanthin glucosyltransferase, valencene synthase, and cytochrome p450 reductase (a cytochrome P450 reductase; page 143, 2nd paragraph).

As per claim 83, Verdezyne discloses a genetically modified Candida viswanathii yeast (a host microorganism, engineered microorganism, genetically modified organism, Candida viswanathii; page 10, 1st paragraph; page 41, 1st paragraph), comprising one or more heterologous nucleic acids encoding one or more terpene biosynthesis polypeptides (a cytochrome P450 reductase (a terpene biosynthesis polypeptide as defined by the instant application; page 530, lines 6-10) can be added to generate an engineered organism; page 27, 3rd paragraph).

As per claim 84, Verdezyne discloses the genetically modified Candida viswanathii yeast of claim 83, and Verdezyne further discloses wherein expression of at least one of the heterologous nucleic acids is regulated by a nucleic acid that provides for fatty acid or alkane induction of expression of the terpene biosynthesis polypeptide (expressed under the control of the POX4 promoter (fatty acid inducible promoter as defined by the instant application (the nucleic acid that provides for fatty acid induction of expression of a terpene biosynthesis polypeptide comprises a promoter region chosen from promoter regions of genes encoding hydratase-dehydrogenaseepimerase, acyl co-A oxidase, POX; page 531, lines 1-4); page 41, 1st paragraph).

-***-Continued Within the Next Supplemental Box-***-

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 557 of 3077

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US18/41579

Box No. VI Certain documents cited

1. Certain published documents (Rules 43*bis*.1 and 70.10)

Application No. Patent No.	Publication date *(day/month/year)*	Filing date *(day/month/year)*	Priority date (valid claim) *(day/month/year)*
WO 2018/094110 A2	24 May 2018 (24.05.2018)	16 November 2017 (16.11.2017)	16 November 2016 (16.11.2016)

2. Non-written disclosures (Rules 43*bis*.1 and 70.9)

Kind of non-written disclosure	Date of non-written disclosure *(day/month/year)*	Date of written disclosure referring to non-written disclosure *(day/month/year)*

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 558 of 3077

Supplemental Box

In case **the space in any of the preceding boxes is not sufficient.**
Continuation of:

-***-Continued from Box V: Citations and Explanations-***-

Claims 43, 46/43, 47/46/43, 48/47/46/43, and 146-148 lack novelty under PCT Article 33(2) as being anticipated by JP 2012-254044 A (KIRIN HOLDING CO LTD) (hereinafter "Kirin").

As per claim 43, Kirin discloses a genetically modified microorganism (modified artificially or native state yeast strains; page 5, 6th paragraph), comprising: one or more heterologous nucleic acids (heterologous gene; page 12, 1st paragraph) encoding one or more terpene biosynthesis polypeptides (Candida utilis that express GGPP synthase gene of Erwinia uredovora origin, phytoene synthase gene (a terpene biosynthesis polypeptide as defined by the instant application; page 530, lines 610); page 11, 3rd paragraph), and a genetic modification that alters the expression of a polypeptide providing for transport of acetyl-carnitine in the microorganism (carnitine transporter and of, CRC1, which preferably has increased expression in the yeast strain; page 6, 5th paragraph).

As per claim 46/43, Kirin discloses the microorganism of claim 43, and Kirin further discloses wherein the microorganism is a fungus (Candida utilis; claim 10).

As per claim 47/46/43, Kirin discloses the microorganism of claim 46/43, and Kirin further discloses wherein the fungus is a yeast (Candida utilis; claim 10).

As per claim 48/47/46/43, Kirin discloses the microorganism of claim 47/46/43, and Kirin further discloses wherein the yeast is chosen from Candida spp (Candida utilis; claim 10), Yarrowia spp, Rhodotorula spp, Rhodosporidium spp, Cryptococcus spp, Trichosporon spp, Lipomyces spp, and Blastobotrys spp.

As per claim 146, Kirin discloses a genetically modified yeast (modified artificially or native state yeast strains; page 5, 6th paragraph) comprising one or more heterologous nucleic acids (heterologous gene; page 12, 1st paragraph) encoding phytoene synthase (Candida utilis that express GGPP synthase gene of Erwinia uredovora origin, phytoene synthase gene (a terpene biosynthesis polypeptide as defined by the instant application; page 530, lines 610); page 11, 3rd paragraph), geranylgeranyl diphosphate synthase and phytoene desaturase.

As per claim 147, Kirin discloses a genetically modified yeast (modified artificially or native state yeast strains; page 5, 6th paragraph) comprising one or more heterologous nucleic acids (heterologous gene; page 12, 1st paragraph) encoding geranylgeranyl diphosphate synthase (Candida utilis that express GGPP synthase gene (geranylgeranyl diphosphate synthase) of Erwinia uredovora origin, phytoene synthase gene (a terpene biosynthesis polypeptide as defined by the instant application; page 530, lines 610); page 11, 3rd paragraph), phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase.

As per claim 148, Kirin discloses a genetically modified yeast (modified artificially or native state yeast strains; page 5, 6th paragraph) comprising one or more heterologous nucleic acids (heterologous gene; page 12, 1st paragraph) encoding geranylgeranyl diphosphate synthase (Candida utilis that express GGPP synthase gene (geranylgeranyl diphosphate synthase) of Erwinia uredovora origin, phytoene synthase gene (a terpene biosynthesis polypeptide as defined by the instant application; page 530, lines 610); page 11, 3rd paragraph), phytoene desaturase, bifunctional lycopene cyclase/phytoene synthase, beta-carotene ketolase, beta-carotene hydroxylase, astaxanthin synthase, and cytochrome p450 reductase.

Claims 44, 46/44, 47/46/44, and 48/47/46/44 lack(s) an inventive step under PCT Article 33(3) as being obvious over Kirin in view of Verdezyne.

As per claim 44, Kirin discloses the microorganism of claim 43, but Kirin does not disclose wherein expression of at least one of the heterologous nucleic acids is regulated by a nucleic acid that provides for fatty acid or alkane induction of expression of the terpene biosynthesis polypeptide. Verdezyne does disclose wherein expression of at least one of the heterologous nucleic acids is regulated by a nucleic acid that provides for fatty acid or alkane induction of expression of the terpene biosynthesis polypeptide (expressed under the control of the POX4 promoter (fatty acid inducible promoter as defined by the instant application (the nucleic acid that provides for fatty acid induction of expression of a terpene biosynthesis polypeptide comprises a promoter region chosen from promoter regions of genes encoding hydratase-dehydrogenaseepimerase, acyl co-A oxidase, POX; page 531, lines 1-4); page 41, 1st paragraph). It would have been obvious to one of ordinary skill in the art at the time of the invention to modify the Kirin invention to provide wherein expression of at least one of the heterologous nucleic acids is regulated by a nucleic acid that provides for fatty acid or alkane induction of expression of the terpene biosynthesis polypeptide, as taught by Verdezyne, in order to increases or decreases expression of the target sequence under certain conditions (page 62, 3rd paragraph).

As per claim 46/44, Kirin and Verdezyne, in combination, disclose the microorganism of claim 44, and Kirin further discloses wherein the microorganism is a fungus (Candida utilis; claim 10).

As per claim 47/46/44, Kirin and Verdezyne, in combination, disclose the microorganism of claim 46/44, and Kirin further discloses wherein the fungus is a yeast (Candida utilis; claim 10).

As per claim 48/47/46/44, Kirin and Verdezyne, in combination, disclose the microorganism of claim 47/46/44, and Kirin and Verdezyne, in combination, disclose wherein the yeast is chosen from Candida spp (Candida utilis; claim 10), Yarrowia spp, Rhodotorula spp, Rhodosporidium spp, Cryptococcus spp, Trichosporon spp, Lipomyces spp, and Blastobotrys spp.

-***-Continued Within the Next Supplemental Box-***-

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 559 of 3077

Supplemental Box

In case **the space in any of the preceding boxes is not sufficient.**
Continuation of:

-***-Continued from Previous Supplemental Box-***-

Claim 85 lacks an inventive step under PCT Article 33(3) as being obvious over Verdezyne in view of Daviet.

As per claim 85, Verdezyne discloses the genetically modified Candida viswanathii yeast of claim 83, but Verdezyne does not disclose wherein expression of at least one of the heterologous nucleic acids is regulated by a nucleic acid that provides for glucose induction of expression of the terpene biosynthesis polypeptide. Daviet does disclose wherein expression of at least one of the heterologous nucleic acids is regulated by a nucleic acid that provides for glucose induction of expression of the terpene biosynthesis polypeptide (glucose regulated promoter being induced by high glucose concentrations and repressed by low glucose concentrations; paragraph [0044]). It would have been obvious to one of ordinary skill in the art at the time of the invention to modify the Verdezyne invention to provide wherein expression of at least one of the heterologous nucleic acids is regulated by a nucleic acid that provides for glucose induction of expression of the terpene biosynthesis polypeptide, as taught by Daviet, in order to provide an increase in the amount of acyclic terpene precursor produced (paragraph [0038]).

Claims 149/146-148 lack an inventive step under PCT Article 33(3) as being obvious over Kirin in view of US 2003/0044946 A1 (LONGO, VD) (hereinafter "Longo").

As per claims 149/146-148, Kirin discloses the genetically modified yeast of claims 146-148, but Kirin does not disclose wherein the amount and/or activity of a Ras2 protein has been decreased. Longo does disclose wherein the amount and/or activity of a Ras2 protein has been decreased (deletion of RAS2; paragraph [0093]). It would have been obvious to one of ordinary skill in the art at the time of the invention to modify the Kirin invention to provide wherein the amount and/or activity of a Ras2 protein has been decreased, as taught by Longo, in order to increase the mean life span by over 100% (paragraph [0093]).

Claims 1-4, 5/1-4, 43-45, 46/43-45, 47/46/43-45, 48/47/46/43-45, 83-85, 146-148, and 149/146-148 have industrial applicability as defined by PCT Article 33(4) because the subject matter can be made or used in industry.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 560 of 3077

PCT Recordation of Search History

Case/PCT Application Number: PCT/US18/41579

CLIN Number/Technical Field of PCT Application: 0005

Dates During Which the Search was Conducted: 27-28 August 2018

Date of Completion of Recordation of Search History Form: 29 August 2018

Research Analyst Initials: SGH

Search Approval Official (SAO) Initials: DGF

Field of Search/Classification Information:

IPC(8) Classifications: C07H 21/04; C07K 14/415; C12N 1/16, 1/19, 1/21, 5/10, 9/00, 9/02, 9/88, 9/96, 15/00, 15/09, 15/52, 15/53, 15/60, 15/63, 15/70, 15/74, 15/82; C12P 5/00, 7/02, 7/26, 15/00; A01H 5/00, 7/00; C11B 9/00 (2018.01)

CPC Classifications: C12N 1/16, 5/10, 9/00, 9/88, 9/96, 15/00, 15/09, 15/52, 15/63, 15/70, 15/74, 15/82, 15/8249; C07H 21/04; C07K 14/415; C12P 5/00, 7/02, 7/26, 15/00; A01H 5/00, 7/00; C11B 9/00

Databases Searched (Patent and Non-Patent Literature (NPL), Including Sub-Databases and Files Searched) and Search Terms Used:

PatSeer (US, EP, WO, JP, DE, GB, CN, FR, KR, ES, AU, IN, CA, INPADOC Data); Google; Google Scholar; EBSCO

Terms: terpene, vector, biosynthesis, polypeptide, genetically, engineered, modified, microbe, microorganism, yeast, inducible, fatty acid, alkane, ras2, reduce, decrease, Candida viswanathii

Database Search String Recordation, Including Dates of Searches:

Patent Database Search Strategy/Results:

PatSeer

ID	Search String	Records	Created Date
L41	TAcd:(yeast) and (ras2 ws (reduc* or delet* or knock* or low* or decreas*))	277	28-Aug-18
L40	TAc:(yeast) and (ras2 ws (reduc* or delet* or knock* or low* or decreas*))	4	28-Aug-18
L39	TA:(yeast) and (ras2 ws (reduc* or delet* or knock* or low* or decreas*))	0	28-Aug-18
L38	TAcd:(yeast) and (ras2) and (geranyl_geranyl_diphosphate_synthase or phytoene_desaturase or lycopene_cyclase or phytoene_synthase or carotene_ketolase or carotene_hydroxylase or astaxanthin_synthase or p450_reductase)	58	28-Aug-18
L37	TAc:(yeast) and (ras2) and (geranyl_geranyl_diphosphate_synthase or phytoene_desaturase or lycopene_cyclase or phytoene_synthase or carotene_ketolase or carotene_hydroxylase or astaxanthin_synthase or p450_reductase)	0	28-Aug-18
L36	TA:(yeast) and (ras2) and (geranyl_geranyl_diphosphate_synthase or phytoene_desaturase or lycopene_cyclase or phytoene_synthase or carotene_ketolase or carotene_hydroxylase or astaxanthin_synthase or p450_reductase)	0	28-Aug-18
L35	PNC:(JP2012254044A)	1	28-Aug-18
L34	TAcd:((acetyl_carnitine) ws (transport)) and (terpen* or phyoten* or lycopen* or caroten* or astaxanthin or zeaxanthin or valencene or p450)	11	28-Aug-18
L33	TAcd:(acetyl_carnitine) ws (transport)	35	28-Aug-18
L32	TAc:(acetyl_carnitine) ws (transport)	2	28-Aug-18
L31	TA:(acetyl_carnitine) and (transport)	1	28-Aug-18
L30	TAcd:(acetyl_carnitine_translocase)	5	28-Aug-18
L29	TAc:(acetyl_carnitine_translocase)	5	28-Aug-18
L28	TA:(acetyl_carnitine_translocase)	1	28-Aug-18
L27	NLPM(TAC:(A genetically modified microorganism comprising one or more heterologous nucleic acids encoding one or more terpene biosynthesis polypeptides and a genetic modification that alters the expression of a polypeptide providing for transport of acetyl-carnitine in the microorganism))	86	28-Aug-18
L26	TAcd:((genetically_modified OR genet*_modifi* OR engin* OR heterolog* OR exogen*_dna* OR unnatur* OR alter*) ws (acetyl_carnitine_translocase)) and (yarrowia* OR yeast* OR microorgan* OR microb* OR bacteria* OR lipomyc* OR rhodotorula* OR candida* OR blastobotri* OR rhodosporidium* OR cryptococcus* OR trichosporon*) NOT L23	0	28-Aug-18
L25	TAcd:((genetically_modified OR genet*_modifi* OR engin* OR heterolog* OR exogen*_dna* OR unnatur* OR alter*) ws (acetyl_carnitine_translocase)) and (terpen*) and (yarrowia* OR yeast* OR microorgan* OR microb* OR bacteria* OR lipomyc* OR rhodotorula* OR candida* OR blastobotri* OR rhodosporidium* OR cryptococcus* OR trichosporon*) NOT L23	0	28-Aug-18
L24	TAcd:((genetically_modified OR genet*_modifi* OR engin* OR heterolog* OR exogen*_dna* OR unnatur* OR alter*) WS (yarrowia* OR yeast* OR microorgan* OR microb* OR bacteria* OR lipomyc* OR rhodotorula* OR candida* OR blastobotri* OR rhodosporidium* OR cryptococcus* OR trichosporon*)) AND (CRC* or acetyl_carnitine_translocase) and (terpen*) NOT L23	1186	28-Aug-18

L23	TAc:((genetically_modified OR genet*_modifi* OR engin* OR heterolog* OR exogen*_dna* OR unnatur* OR alter*) WS (yarrowia* OR yeast* OR microorgan* OR microb* OR bacteria* OR lipomyc* OR rhodotorula* OR candida* OR blastobotri* OR rhodosporidium* OR cryptococcus* OR trichosporon*)) AND (CRC* or acetyl_carnitine_translocase) and (terpen*)	5	28-Aug-18
L22	TAc:((genetically_modified OR genet*_modifi* OR engin* OR heterolog* OR exogen*_dna* OR unnatur* OR alter*) WS (yarrowia* OR yeast* OR microorgan* OR microb* OR bacteria* OR lipomyc* OR rhodotorula* OR candida* OR blastobotri* OR rhodosporidium* OR cryptococcus* OR trichosporon*)) AND (yeast*_artifici*_chromosom* OR vector* OR plasmid*) AND (CRC* or acetyl_carnitine_translocase)	87	28-Aug-18
L21	TA:((genetically_modified OR genet*_modifi* OR engin* OR heterolog* OR exogen*_dna* OR unnatur* OR alter*) WS (yarrowia* OR yeast* OR microorgan* OR microb* OR bacteria* OR lipomyc* OR rhodotorula* OR candida* OR blastobotri* OR rhodosporidium* OR cryptococcus* OR trichosporon*)) AND (yeast*_artifici*_chromosom* OR vector* OR plasmid*) AND (CRC* or acetyl_carnitine_translocase)	0	28-Aug-18
L20	TAc:(yeast) and (ras2 ws (decreas* or reduc*))	3	27-Aug-18
L19	TA:(yeast) and (ras2 ws (decreas* or reduc*))	0	27-Aug-18
L18	TA:(yeast) and (ras2 ws decreas*)	0	27-Aug-18
L17	TAcd:((genetically_modified OR genet*_modifi* OR engin* OR heterolog* OR exogen*_dna* OR unnatur* OR alter*) WS (yarrowia* OR yeast* OR microorgan* OR microb* OR bacteria* OR lipomyc* OR rhodotorula* OR candida* OR blastobotri* OR rhodosporidium* OR cryptococcus* OR trichosporon*)) AND (yeast*_artifici*_chromosom* OR vector* OR plasmid*) AND (terpene* OR tackifi* OR carvon* OR humulen* OR taxadien* OR squalen* OR pinene* OR carene* OR dipentene* OR limonene*) (CPC:(("C12N1/16" OR "C12N15/63" OR "C12N15/64" OR "C12N15/65" OR "C12N15/66" OR "C12N15/81" OR "C12N15/85" OR "C12N15/90" OR "C12N15/905" OR "C12N5/10" OR "C12P19/34" OR "C12Q1/68") AND ("A01H5/00" OR "A01H7/00" OR "C07H21/04" OR "C07K14/415" OR "C11B9/00" OR "C12N1/16" OR "C12N15/00" OR "C12N15/09" OR "C12N15/52" OR "C12N15/63" OR "C12N15/70" OR "C12N15/74" OR "C12N15/82" OR "C12N15/8249" OR "C12N5/10" OR "C12N9/00" OR "C12N9/88" OR "C12N9/96" OR "C12P15/00" OR "C12P5/00" OR "C12P7/02" OR "C12P7/26")) OR IC:(("C12N1/16" OR "C12N1/19" OR "C12N15/63" OR "C12N15/64" OR "C12N15/65" OR "C12N15/66" OR "C12N15/81" OR "C12N15/85" OR "C12N15/90" OR "C12N5/10" OR "C12P19/34" OR "C12Q1/68") AND ("A01H5/00" OR "A01H7/00" OR "C07H21/04" OR "C07K14/415" OR "C11B9/00" OR "C12N1/16" OR "C12N1/19" OR "C12N1/21" OR "C12N15/00" OR "C12N15/09" OR "C12N15/52" OR "C12N15/53" OR "C12N15/60" OR "C12N15/63" OR "C12N15/70" OR "C12N15/74" OR "C12N15/82" OR "C12N5/10" OR "C12N9/00" OR "C12N9/02" OR "C12N9/88" OR "C12N9/96" OR "C12P15/00" OR "C12P5/00" OR "C12P7/02" OR "C12P7/26")))	1666	27-Aug-18
L16	TAc:((genetically_modified OR genet*_modifi* OR engin* OR heterolog* OR exogen*_dna* OR unnatur* OR alter*) WS (yarrowia* OR yeast* OR microorgan* OR microb* OR bacteria* OR lipomyc* OR rhodotorula* OR candida* OR blastobotri* OR rhodosporidium* OR cryptococcus* OR trichosporon*)) AND (yeast*_artifici*_chromosom* OR vector* OR plasmid*) AND (terpene* OR tackifi* OR carvon* OR humulen* OR taxadien* OR squalen* OR pinene* OR carene* OR dipentene* OR limonene*) (CPC:(("C12N1/16" OR "C12N15/63" OR "C12N15/64" OR "C12N15/65" OR "C12N15/66" OR "C12N15/81" OR "C12N15/85" OR "C12N15/90" OR "C12N15/905" OR "C12N5/10" OR "C12P19/34" OR	43	27-Aug-18

	"C12Q1/68") AND ("A01H5/00" OR "A01H7/00" OR "C07H21/04" OR "C07K14/415" OR "C11B9/00" OR "C12N1/16" OR "C12N15/00" OR "C12N15/09" OR "C12N15/52" OR "C12N15/63" OR "C12N15/70" OR "C12N15/74" OR "C12N15/82" OR "C12N15/8249" OR "C12N5/10" OR "C12N9/00" OR "C12N9/88" OR "C12N9/96" OR "C12P15/00" OR "C12P5/00" OR "C12P7/02" OR "C12P7/26")) OR IC:(("C12N1/16" OR "C12N1/19" OR "C12N15/63" OR "C12N15/64" OR "C12N15/65" OR "C12N15/66" OR "C12N15/81" OR "C12N15/85" OR "C12N15/90" OR "C12N5/10" OR "C12P19/34" OR "C12Q1/68") AND ("A01H5/00" OR "A01H7/00" OR "C07H21/04" OR "C07K14/415" OR "C11B9/00" OR "C12N1/16" OR "C12N1/19" OR "C12N1/21" OR "C12N15/00" OR "C12N15/09" OR "C12N15/52" OR "C12N15/53" OR "C12N15/60" OR "C12N15/63" OR "C12N15/70" OR "C12N15/74" OR "C12N15/82" OR "C12N5/10" OR "C12N9/00" OR "C12N9/02" OR "C12N9/88" OR "C12N9/96" OR "C12P15/00" OR "C12P5/00" OR "C12P7/02" OR "C12P7/26")))		
L15	TAcd:((genetically_modified OR genet*_modifi* OR engin* OR heterolog* OR exogen*_dna* OR unnatur* OR alter*) WS (yarrowia* OR yeast* OR microorgan* OR microb* OR bacteria* OR lipomyc* OR rhodotorula* OR candida* OR blastobotri* OR rhodosporidium* OR cryptococcus* OR trichosporon*)) AND (yeast*_artifici*_chromosom* OR vector* OR plasmid*) AND (terpene* OR tackifi* OR carvon* OR humulen* OR taxadien* OR squalen* OR pinene* OR carene* OR dipentene* OR limonene*)	6686	27-Aug-18
L14	TAc:((genetically_modified OR genet*_modifi* OR engin* OR heterolog* OR exogen*_dna* OR unnatur* OR alter*) WS (yarrowia* OR yeast* OR microorgan* OR microb* OR bacteria* OR lipomyc* OR rhodotorula* OR candida* OR blastobotri* OR rhodosporidium* OR cryptococcus* OR trichosporon*)) AND (yeast*_artifici*_chromosom* OR vector* OR plasmid*) AND (terpene* OR tackifi* OR carvon* OR humulen* OR taxadien* OR squalen* OR pinene* OR carene* OR dipentene* OR limonene*)	105	27-Aug-18
L13	TA:((genetically_modified OR genet*_modifi* OR engin* OR heterolog* OR exogen*_dna* OR unnatur* OR alter*) WS (yarrowia* OR yeast* OR microorgan* OR microb* OR bacteria* OR lipomyc* OR rhodotorula* OR candida* OR blastobotri* OR rhodosporidium* OR cryptococcus* OR trichosporon*)) AND (yeast*_artifici*_chromosom* OR vector* OR plasmid*) AND (terpene* OR tackifi* OR carvon* OR humulen* OR taxadien* OR squalen* OR pinene* OR carene* OR dipentene* OR limonene*)	3	27-Aug-18
L12	TAc:(yeast) and (heterolog*) and (terpen*) and (vector or nucleic*) (CPC:(("C12N1/16" OR "C12N15/63" OR "C12N15/64" OR "C12N15/65" OR "C12N15/66" OR "C12N15/81" OR "C12N15/85" OR "C12N15/90" OR "C12N15/905" OR "C12N5/10" OR "C12P19/34" OR "C12Q1/68") AND ("A01H5/00" OR "A01H7/00" OR "C07H21/04" OR "C07K14/415" OR "C11B9/00" OR "C12N1/16" OR "C12N15/00" OR "C12N15/09" OR "C12N15/52" OR "C12N15/63" OR "C12N15/70" OR "C12N15/74" OR "C12N15/82" OR "C12N15/8249" OR "C12N5/10" OR "C12N9/00" OR "C12N9/88" OR "C12N9/96" OR "C12P15/00" OR "C12P5/00" OR "C12P7/02" OR "C12P7/26")) OR IC:(("C12N1/16" OR "C12N1/19" OR "C12N15/63" OR "C12N15/64" OR "C12N15/65" OR "C12N15/66" OR "C12N15/81" OR "C12N15/85" OR "C12N15/90" OR "C12N5/10" OR "C12P19/34" OR "C12Q1/68") AND ("A01H5/00" OR "A01H7/00" OR "C07H21/04" OR "C07K14/415" OR "C11B9/00" OR "C12N1/16" OR "C12N1/19" OR "C12N1/21" OR "C12N15/00" OR "C12N15/09" OR "C12N15/52" OR "C12N15/53" OR "C12N15/60" OR "C12N15/63" OR	40	27-Aug-18

	"C12N15/70" OR "C12N15/74" OR "C12N15/82" OR "C12N5/10" OR "C12N9/00" OR "C12N9/02" OR "C12N9/88" OR "C12N9/96" OR "C12P15/00" OR "C12P5/00" OR "C12P7/02" OR "C12P7/26")))		
L11	TAcd:(((genetically_modified OR genet*_modifi* OR engin* OR heterolog* OR exogen*_dna* OR unnatur* OR alter*) WS (yarrowia* OR yeast* OR microorgan* OR microb* OR bacteria* OR lipomyc* OR rhodotorula* OR candida* OR blastobotri* OR rhodosporidium* OR cryptococcus* OR trichosporon*)) AND (yeast*_artifici*_chromosom* OR vector* OR plasmid*) AND ((terpene* OR tackifi* OR carvon* OR humulen* OR taxadien* OR squalen* OR pinene* OR carene* OR dipentene* OR limonene*) WS (biosynthesi*_polypeptid* OR terpen*_synthas* OR valencen*_synthas* OR astaxanthin*_synthas* OR phytoen*_synthas* OR beta*_caroten*_hydroxylas* OR beta**caroten*ketolas* OR cytochrom*_p450*_reductas* OR geranylgeranyl*_diphosph*_synthas* OR zeaxanthin*_glucosyltransferas* OR phytoen*_desaturas* OR bifunct*_lycopen*_cyclase/phytoen*_synthas* OR lycopen*_cyclas*)) AND (((induc*) WP (fatty_acid OR fatti*acid* OR linolen*_acid* OR arachidon*_acid* OR oleic*_acid* OR linol*_acid*)) or (alkane* OR methan* OR ethan* OR propan* OR paraffin*))) (CPC:(("C12N1/16" OR "C12N15/63" OR "C12N15/64" OR "C12N15/65" OR "C12N15/66" OR "C12N15/81" OR "C12N15/85" OR "C12N15/90" OR "C12N15/905" OR "C12N5/10" OR "C12P19/34" OR "C12Q1/68") AND ("A01H5/00" OR "A01H7/00" OR "C07H21/04" OR "C07K14/415" OR "C11B9/00" OR "C12N1/16" OR "C12N15/00" OR "C12N15/09" OR "C12N15/52" OR "C12N15/63" OR "C12N15/70" OR "C12N15/74" OR "C12N15/82" OR "C12N15/8249" OR "C12N5/10" OR "C12N9/00" OR "C12N9/88" OR "C12N9/96" OR "C12P15/00" OR "C12P5/00" OR "C12P7/02" OR "C12P7/26")) OR IC:(("C12N1/16" OR "C12N1/19" OR "C12N15/63" OR "C12N15/64" OR "C12N15/65" OR "C12N15/66" OR "C12N15/81" OR "C12N15/85" OR "C12N15/90" OR "C12N5/10" OR "C12P19/34" OR "C12Q1/68") AND ("A01H5/00" OR "A01H7/00" OR "C07H21/04" OR "C07K14/415" OR "C11B9/00" OR "C12N1/16" OR "C12N1/19" OR "C12N1/21" OR "C12N15/00" OR "C12N15/09" OR "C12N15/52" OR "C12N15/53" OR "C12N15/60" OR "C12N15/63" OR "C12N15/70" OR "C12N15/74" OR "C12N15/82" OR "C12N5/10" OR "C12N9/00" OR "C12N9/02" OR "C12N9/88" OR "C12N9/96" OR "C12P15/00" OR "C12P5/00" OR "C12P7/02" OR "C12P7/26")))	136	27-Aug-18
L10	TAcd:(((genetically_modified OR genet*_modifi* OR engin* OR heterolog* OR exogen*_dna* OR unnatur* OR alter*) WS (yarrowia* OR yeast* OR microorgan* OR microb* OR bacteria* OR lipomyc* OR rhodotorula* OR candida* OR blastobotri* OR rhodosporidium* OR cryptococcus* OR trichosporon*)) AND (yeast*_artifici*_chromosom* OR vector* OR plasmid*) AND ((terpene* OR tackifi* OR carvon* OR humulen* OR taxadien* OR squalen* OR pinene* OR carene* OR dipentene* OR limonene*) WS (biosynthesi*_polypeptid* OR terpen*_synthas* OR valencen*_synthas* OR astaxanthin*_synthas* OR phytoen*_synthas* OR beta*_caroten*_hydroxylas* OR beta**caroten*ketolas* OR cytochrom*_p450*_reductas* OR geranylgeranyl*_diphosph*_synthas* OR zeaxanthin*_glucosyltransferas* OR phytoen*_desaturas* OR bifunct*_lycopen*_cyclase/phytoen*_synthas* OR lycopen*_cyclas*)) AND (((induc*) WP (fatty_acid OR fatti*acid* OR linolen*_acid* OR arachidon*_acid* OR oleic*_acid* OR linol*_acid*)) or (alkane* OR methan* OR ethan* OR propan* OR paraffin*)))	322	27-Aug-18
L9	TAcd:(((genetically_modified OR genet*_modifi* OR engin* OR heterolog* OR exogen*_dna* OR unnatur* OR alter*) WS (yarrowia* OR yeast* OR microorgan* OR microb* OR bacteria* OR lipomyc* OR rhodotorula* OR	319	27-Aug-18

	candida* OR blastobotri* OR rhodosporidium* OR cryptococcus* OR trichosporon*)) AND (yeast*_artifici*_chromosom* OR vector* OR plasmid*) AND ((terpene* OR tackifi* OR carvon* OR humulen* OR taxadien* OR squalen* OR pinene* OR carene* OR dipentene* OR limonene*) WS (biosynthesi*_polypeptid* OR terpen*_synthas* OR valencen*_synthas* OR astaxanthin*_synthas* OR phytoen*_synthas* OR beta*_caroten*_hydroxylas* OR beta**caroten*ketolas* OR cytochrom*_p450*_reductas* OR geranylgeranyl*_diphosph*_synthas* OR zeaxanthin*_glucosyltransferas* OR phytoen*_desaturas* OR bifunct*_lycopen*_cyclase/phytoen*_synthas* OR lycopen*_cyclas*)) AND (((induc*) WP (fatty_acid OR fatti*acid* OR linolen*_acid* OR arachidon*_acid* OR oleic*_acid* OR linol*_acid*)) or (alkane* OR methan* OR ethan* OR propan* OR paraffin*))) NOT L8		
L8	TAc:(((genetically_modified OR genet*_modifi* OR engin* OR heterolog* OR exogen*_dna* OR unnatur* OR alter*) WS (yarrowia* OR yeast* OR microorgan* OR microb* OR bacteria* OR lipomyc* OR rhodotorula* OR candida* OR blastobotri* OR rhodosporidium* OR cryptococcus* OR trichosporon*)) AND (yeast*_artifici*_chromosom* OR vector* OR plasmid*) AND ((terpene* OR tackifi* OR carvon* OR humulen* OR taxadien* OR squalen* OR pinene* OR carene* OR dipentene* OR limonene*) WS (biosynthesi*_polypeptid* OR terpen*_synthas* OR valencen*_synthas* OR astaxanthin*_synthas* OR phytoen*_synthas* OR beta*_caroten*_hydroxylas* OR beta**caroten*ketolas* OR cytochrom*_p450*_reductas* OR geranylgeranyl*_diphosph*_synthas* OR zeaxanthin*_glucosyltransferas* OR phytoen*_desaturas* OR bifunct*_lycopen*_cyclase/phytoen*_synthas* OR lycopen*_cyclas*)) AND (((induc*) WP (fatty_acid OR fatti*acid* OR linolen*_acid* OR arachidon*_acid* OR oleic*_acid* OR linol*_acid*)) or (alkane* OR methan* OR ethan* OR propan* OR paraffin*)))	6	27-Aug-18
L7	TA:(((genetically_modified OR genet*_modifi* OR engin* OR heterolog* OR exogen*_dna* OR unnatur* OR alter*) WS (yarrowia* OR yeast* OR microorgan* OR microb* OR bacteria* OR lipomyc* OR rhodotorula* OR candida* OR blastobotri* OR rhodosporidium* OR cryptococcus* OR trichosporon*)) AND (yeast*_artifici*_chromosom* OR vector* OR plasmid*) AND ((terpene* OR tackifi* OR carvon* OR humulen* OR taxadien* OR squalen* OR pinene* OR carene* OR dipentene* OR limonene*) WS (biosynthesi*_polypeptid* OR terpen*_synthas* OR valencen*_synthas* OR astaxanthin*_synthas* OR phytoen*_synthas* OR beta*_caroten*_hydroxylas* OR beta**caroten*ketolas* OR cytochrom*_p450*_reductas* OR geranylgeranyl*_diphosph*_synthas* OR zeaxanthin*_glucosyltransferas* OR phytoen*_desaturas* OR bifunct*_lycopen*_cyclase/phytoen*_synthas* OR lycopen*_cyclas*)) AND (((induc*) WP (fatty_acid OR fatti*acid* OR linolen*_acid* OR arachidon*_acid* OR oleic*_acid* OR linol*_acid*)) or (alkane* OR methan* OR ethan* OR propan* OR paraffin*)))	0	27-Aug-18
L6	PNC:(US9745602B2)	1	27-Aug-18
L5	TAc:(yeast) and (heterolog*) and (terpen*) and (vector or nucleic*)	82	27-Aug-18
L4	TA:(yeast) and (heterolog*) and (terpen*) and (vector or nucleic*)	0	27-Aug-18
L3	INV:(aeling w1 kimberly)	0	27-Aug-18
L2	((CASN:(verdezyne) OR ASN:(verdezyne)) AND TACD:(terpene))	0	27-Aug-18
L1	(CASN:(verdezyne) OR ASN:(verdezyne))	14	27-Aug-18

Non-Patent Literature (NPL) Search Strategy/Results:

Google: 27-28 August 2018

Query	Result
biological methods for preparing terpenes	5860000
A genetically modified yeast with a vector encoding a terpene biosynthesis polypeptide inducible by fatty acid or alkane	15100
Candida viswanathii ras2	20
(("genetically modified" OR "Genetically modified") * (Yarrowia OR yeast OR microorganism OR microbe)) AND ("yeast artificial chromosome" OR vector) AND ((terpene OR tackifier OR carvone OR humulene) * ("biosynthesis polypeptide" OR "terpene synthase")) AND ((inducible) * ("fatty acid" OR "Fatty acid")) AND (alkane OR methane OR ethane OR propane)	0
Decreased ras2 activity in yeast	60500
Genetically modified acetyl-carnitine translocase	91200
yeast expression of a Genetically modified acetyl-carnitine translocase	31500
a genetically modified yeast strain with reduced expression of ras2 that expresses terpene enzymes	20800

Google Scholar: 27-28 August 2018

Query	Result
biological methods for preparing terpenes	70000
A genetically modified yeast with a vector encoding a terpene biosynthesis polypeptide inducible by fatty acid or alkane	591
Candida viswanathii ras2	1
(("genetically modified" OR "Genetically modified") * (Yarrowia OR yeast OR microorganism OR microbe)) AND ("yeast artificial chromosome" OR vector) AND ((terpene OR tackifier OR carvone OR humulene) * ("biosynthesis polypeptide" OR "terpene synthase")) AND (((inducible) * ("fatty acid" OR "Fatty acid")) or (alkane OR methane OR ethane OR propane)	362
Decreased ras2 activity in yeast	5230
yeast expression of a Genetically modified acetyl-carnitine translocase	250
a genetically modified yeast strain with reduced expression of ras2 that expresses terpene enzymes	116

EBSCO: 27 August 2018

Query	Result
TI ((("genetically modified" OR "genet* modifi*" OR engin* OR heterolog* OR "exogen* dna*" OR unnatur* OR alter*) N15 (yarrowia* OR yeast* OR microorgan* OR microb* OR bacteria* OR lipomyc* OR rhodotorula* OR candida* OR blastobotri* OR rhodosporidium* OR cryptococcus* OR trichosporon*)) AND ("yeast* artifici* chromosom*" OR vector* OR plasmid*) AND ((terpene* OR tackifi* OR carvon* OR humulen* OR taxadien* OR squalen* OR pinene* OR carene* OR dipentene* OR limonene*) N15 ("biosynthesi* polypeptid*" OR "terpen* synthas*" OR "valencen* synthas*" OR "astaxanthin* synthas*" OR "phytoen* synthas*" OR "beta* caroten* hydroxylas*" OR beta*caroten*ketolas* OR "cytochrom* p450* reductas*" OR "geranylgeranyl* diphosph* synthas*" OR "zeaxanthin* glucosyltransferas*" OR "phytoen* desaturas*" OR "bifunct* lycopen* cyclase/phytoen* synthas*" OR "lycopen* cyclas*")) AND ((induc*) N50 ("fatty acid" OR fatti*acid* OR "linolen* acid*" OR "arachidon* acid*" OR "oleic* acid*" OR "linol* acid*")) AND (alkane* OR methan* OR ethan* OR propan* OR paraffin*))	0
genetically modified organisms AND yeast AND terpene AND viswanathii	15
genetically modified organisms AND yeast AND terpene AND Ras2	52

(19)日本国特許庁(JP)　(12)公開特許公報(A)　(11)特許出願公開番号
特開2012-254044
(P2012-254044A)
(43)公開日　平成24年12月27日(2012.12.27)

(51)Int.Cl.			F I			テーマコード（参考）
C12N	***1/19***	***(2006.01)***	C12N	1/19	ZNA	4B024
C12P	***7/04***	***(2006.01)***	C12P	7/04		4B050
C12N	*15/09*	*(2006.01)*	C12N	15/00	A	4B064
C12N	*9/10*	*(2006.01)*	C12N	9/10		4B065

審査請求 未請求　請求項の数 15　ＯＬ　（全 27 頁）

(21)出願番号　特願2011-129527 (P2011-129527)
(22)出願日　平成23年6月9日 (2011.6.9)

（出願人による申告）平成２２年度、独立行政法人新エネルギー・産業技術総合開発機構・グリーン・サステイナブルケミカルプロセス基盤技術開発委託研究、産業技術力強化法第１９条の適用を受ける特許出願

(71)出願人　000253503
キリンホールディングス株式会社
東京都中央区新川二丁目１０番１号
(74)代理人　100117787
弁理士　勝沼　宏仁
(74)代理人　100091487
弁理士　中村　行孝
(74)代理人　100107342
弁理士　横田　修孝
(74)代理人　100111730
弁理士　伊藤　武泰
(72)発明者　玉　川　英　幸
東京都中央区新川二丁目１０番１号　キリンホールディングス株式会社内

最終頁に続く

(54)【発明の名称】酵母におけるアセチルＣｏＡを経由する代謝経路を利用した物質の製造法

(57)【要約】
【課題】アセチルＣｏＡを経由する代謝産物を高生産し得る酵母菌株の提供。
【解決手段】以下の酵素：（ａ）アセチルＣｏＡ合成酵素（ＡＣＳ）；および（ｂ）アセチルＣｏＡアセチル基転移酵素（ＥＲＧ１０）の活性が増強されている、酵母菌株。
【選択図】なし

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 568 of 3077

【特許請求の範囲】
【請求項１】
以下の酵素：
（ａ）アセチルＣｏＡ合成酵素（ＡＣＳ）；および
（ｂ）アセチルＣｏＡアセチル基転移酵素（ＥＲＧ１０）
の活性が増強されている、酵母菌株。
【請求項２】
以下の遺伝子：
（ａ）アセチルＣｏＡ合成酵素（ＡＣＳ）の活性を有するポリペプチドをコードする遺伝子；および
（ｂ）アセチルＣｏＡアセチル基転移酵素（ＥＲＧ１０）の活性を有するポリペプチドをコードする遺伝子
の発現が増強されている、請求項１に記載の酵母菌株。
【請求項３】
ＡＣＳ活性を有するポリペプチドが、
（ｉ）配列番号１２または配列番号１４で表されるアミノ酸配列を含むポリペプチド、または
（ｉｉ）配列番号１２または配列番号１４で表されるアミノ酸配列において、１もしくは数個のアミノ酸が欠失、置換、付加もしくは挿入されたアミノ酸配列を含み、かつＡＣＳ活性を有するポリペプチド
である、請求項２に記載の酵母菌株。
【請求項４】
ＥＲＧ１０活性を有するポリペプチドが、
（ｉ）配列番号６２で表されるアミノ酸配列を含むポリペプチド、または
（ｉｉ）配列番号６２で表されるアミノ酸配列において、１もしくは数個のアミノ酸が欠失、置換、付加もしくは挿入されたアミノ酸配列を含み、かつＥＲＧ１０活性を有するポリペプチド
である、請求項２または３に記載の酵母菌株。
【請求項５】
さらに、以下の遺伝子：
（ｃ）ＡＴＰ－クエン酸除去付加酵素（ＡＣＬ）の活性を有するポリペプチドをコードする遺伝子；
（ｄ）ミトコンドリア内膜カルニチン輸送体（ＣＲＣ１）の活性を有するポリペプチドをコードする遺伝子；および
（ｅ）カルニチンアセチル転移酵素遺伝子（ＹＡＴ１）の活性を有するポリペプチドをコードする遺伝子
からなる群から選択される少なくとも一種の遺伝子の発現が増強されている、請求項１～４のいずれか一項に記載の酵母菌株。
【請求項６】
ＡＣＬ活性を有するポリペプチドが、ＡＣＬ１活性を有するポリペプチドおよびＡＣＬＹ活性を有するポリペプチドからなるものであり、
ＡＣＬ１活性を有するポリペプチドが、
（ｉ）配列番号３２で表されるアミノ酸配列を含むポリペプチド、または
（ｉｉ）配列番号３２で表されるアミノ酸配列において、１もしくは数個のアミノ酸が欠失、置換、付加もしくは挿入されたアミノ酸配列を含み、かつＡＣＬ１活性を有するポリペプチド
であり、
ＡＣＬＹ活性を有するポリペプチドが、
（ｉ）配列番号３４で表されるアミノ酸配列を含むポリペプチド、または
（ｉｉ）配列番号３４で表されるアミノ酸配列において、１もしくは数個のアミノ酸が欠失

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 569 of 3077

、置換、付加もしくは挿入されたアミノ酸配列を含み、かつＡＣＬＹ活性を有するポリペプチド
である、請求項５に記載の酵母菌株。
【請求項７】
ＣＲＣ１活性を有するポリペプチドが、
（ｉ）配列番号２４で表されるアミノ酸配列を含むポリペプチド、または
（ｉｉ）配列番号２４で表されるアミノ酸配列において、１もしくは数個のアミノ酸が欠失、置換、付加もしくは挿入されたアミノ酸配列を含み、かつＣＲＣ１活性を有するポリペプチド
である、請求項５または６に記載の酵母菌株。
【請求項８】
ＹＡＴ１活性を有するポリペプチドが、
（ｉ）配列番号２８で表されるアミノ酸配列を含むポリペプチド、または
（ｉｉ）配列番号２８で表されるアミノ酸配列において、１もしくは数個のアミノ酸が欠失、置換、付加もしくは挿入されたアミノ酸配列を含み、かつＹＡＴ１活性を有するポリペプチド
である、請求項５～７のいずれか一項に記載の酵母菌株。
【請求項９】
クラブツリー効果陰性酵母である、請求項１～８のいずれか一項に記載の酵母菌株。
【請求項１０】
キャンディダ・ユティリスである、請求項１～９のいずれか一項に記載の酵母菌株。
【請求項１１】
アセチルＣｏＡを経由する代謝産物を合成するための遺伝子により形質転換されているものである、請求項１～１０のいずれか一項に記載の酵母菌株。
【請求項１２】
ＣｏＡ転移酵素サブユニットＡ（ｃｔｆＡ）、ＣｏＡ転移酵素サブユニットＢ（ｃｔｆＢ）、アセト酢酸脱炭酸酵素（ａｄｃ）および二級アルコール脱水素酵素（ｓａｄｈ）の活性を有するポリペプチドをコードしている４種の遺伝子により形質転換されてなる、請求項１１に記載の酵母菌株。
【請求項１３】
請求項１～１２のいずれか一項に記載の酵母菌株を培養することを含んでなる、物質を製造する方法。
【請求項１４】
前記物質が、アセチルＣｏＡを経由する代謝産物である、請求項１３に記載の方法。
【請求項１５】
前記物質がイソプロパノールである、請求項１３または１４に記載の方法。
【発明の詳細な説明】
【発明の背景】
【０００１】
技術分野
本発明は、酵母において、アセチルＣｏＡを経由する代謝経路を利用した物質の製造方法に関する。
【０００２】
背景技術
深刻化する環境問題を背景に、バイオ燃料が石油の代替燃料として注目を集めている。エタノールはバイオ燃料としてよく知られているが、ガソリンより低いエネルギー密度しか有さないこと、輸送燃料として機能させるためには限られた濃度範囲でガソリンと混合する必要があることから、燃料としての欠点も多い。イソプロパノールは、１－ブタノールよりも高いオクタン価を有し、また、工業原料としても利用される汎用性の高い溶剤である。イソプロパノールは、ナフサ由来のプロピレンの水和によって製造されており、再

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 570 of 3077

生可能な供給源から商業的に生産されたことはない。

【０００３】

微生物によるアセトンやブタノールの発酵生産は、２０世紀初頭に産業的に行われていた。絶対嫌気性菌のある種のクロストリジウム属細菌はアセトンとブタノールを同時に生産することができる（アセトン・ブタノール発酵）。アセトンは細菌細胞内に存在するアセチルＣｏＡを原料として生産される。アセチルＣｏＡは、アセチルＣｏＡアセチル基転移酵素（チオラーゼ）によってアセトアセチルＣｏＡに変換され、さらにＣｏＡ転移酵素によってアセト酢酸に変換され、さらにアセト酢酸脱炭酸酵素によってアセトンへと変換される。ある種のクロストリジウム属細菌はさらに二級アルコールの脱水素酵素を有し、アセトンからイソプロパノールを生産する。Ｃｈｅｎらは、クロストリジウム・ベイジェリンキを含む５２株のクロストリジウム属細菌の中で、最大で３０ｍＭ（１．８ｇ／Ｌ）のイソプロパノールを生産する株が存在したことを報告している（非特許文献１：Chen, J.-S., and S. F. Hiu. Acetone-butanol-isopropanol production by Clostridium beijerinckii (synonym, Clostridium butylicum). Biotechnol. Lett. 1986. 8:371-376）。また、最近になってＨａｎａｉらおよびＪｏｊｉｍａらは、大腸菌にクロストリジウム属由来のＣｏＡ転移酵素、アセト酢酸脱炭酸酵素、二級アルコール脱水素酵素を導入し、内在性のアセチルＣｏＡアセチル基転移酵素を過剰発現させることによってイソプロパノールを効率的に生産できる大腸菌株を報告している（非特許文献２：T. Hanai, S. Atsumi, and J. C. Liao. Engineered Synthetic Pathway for Isopropanol Production in Escherichia coli. Appl Environ Microbiol, Dec. 2007, p. 7814-7818；非特許文献３：Toru Jojima & Masayuki Inui & Hideaki Yukawa. Production of isopropanol by metabolically engineered Escherichia coli. Appl Microbiol Biotechnol 2008 77:1219-1224）。しかしながら、大腸菌は、糖耐性、浸透圧耐性、塩耐性、有機溶剤耐性など各種ストレス耐性はあまり高くないことが多く、ファージ汚染によるリスクもある。また、イソプロパノール耐性が低いことは生産物による発酵の阻害が起こることを意味する。発酵阻害を回避するためにガスストリッピング法などを用いて生産物を培地から除去しながら発酵を続ける方法が知られているが、該方法の使用は発酵生産の工程を煩雑化させるだけでなく、コスト増加につながる。これらの問題を解決するための一つの手段として、高い有機溶剤耐性を有する微生物を用いてイソプロパノール生産をすることが考えられる。

【０００４】

単細胞真核微生物である酵母は高い有機溶剤耐性を示し、また、大腸菌などの微生物と同様に、培養設備、培養時間および培養コストの面で高等生物の細胞よりも安価に培養することができるという特徴を有する。特に、出芽酵母はビールやワインなどの酒類やパンの製造に用いられ、その食品としての安全性が確認されている。

【０００５】

酵母でイソプロパノールを生産するためには、その酵母菌株に、外因性のＣｏＡ転移酵素、アセト酢酸脱炭酸酵素、および二級アルコール脱水素酵素を導入する必要がある。これまで、酵母によるイソプロパノールの製造例としては、サッカロマイセス・セレビシエ（Ｓａｃｃｈａｒｏｍｙｃｅｓ　ｃｅｒｅｖｉｓｉａｅ）にイソプロピルアルコール生産酵素群の遺伝子を導入し、この形質転換体を培養することによりイソプロパノールを製造したことが報告されている（特許文献１：特開２０１１－９７９２９号公報）。しかし、この方法によるイソプロパノールの生産量は、培養液中の濃度として１００ｍｇ／Ｌ前後という少量に過ぎない。酵母は真核生物であり、オルガネラが存在することから、イソプロパノール合成の基質であるアセチルＣｏＡを細胞質に供給する経路が複数存在している。どの経路からアセチルＣｏＡを細胞質に供給することがイソプロパノール合成にとって有利であるのか、あるいはどのような酵母がそういった能力に優れているか、これまでに調査された例はない。

【０００６】

キャンディダ・ユティリス（Ｃａｎｄｉｄａ　ｕｔｉｌｉｓ）は、サッカロマイセス・セレビシエ（Ｓａｃｃｈａｒｏｍｙｃｅｓ　ｃｅｒｅｖｉｓｉａｅ）およびサッカロマイ

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 571 of 3077

セス・フラジリス（Ｓａｃｃｈａｒｏｍｙｃｅｓ　ｆｒａｇｉｌｉｓ）と同様に、ＦＤＡ（米国連邦食品医薬品局）によって食用酵母として認定されており、食品添加物として安全に使用できることが認められている食用酵母である。実際に、現在でも、キャンディダ・ユティリスは、日本をはじめとして、ドイツやアメリカなど世界各国で生産され、食飼料として利用されている。また、キャンディダ・ユティリスは、ペントースの資化による菌体生産、エチルアセテート、Ｌ－グルタミン酸、グルタチオン、インベルターゼなどの生産株として広く産業界で利用されてきた。近年、キャンディダ・ユティリスにおいても形質転換法が開発され、様々な化合物が代謝工学的に生産される技術が開発されている（非特許文献４：Hiroshi Shimada, Keiji Kondo, Paul D. Fraser, Yutaka Miura, Toshiko Saito, and Norihiko Misawa., Increased Carotenoid Production by the Food Yeast Candida utilis through Metabolic Engineering of the Isoprenoid Pathway., Appl Environ Microbiol, July 1998, p. 2676-2680, Vol. 64, No. 7；非特許文献５：Yutaka Miura, Keiji Kondo, Toshiko Saito, Hiroshi Shimada, Paul D. Fraser, and Norihiko MisawaProduction of the Carotenoids Lycopene, β-Carotene, and Astaxanthin in the Food Yeast Candida utilis., Appl Environ Microbiol, April 1998, p. 1226-1229, Vol. 64, No. 4；非特許文献６：Keiji Kondo', Yutaka Miura, Hidetaka Sone, Kazuo Kobayashi & Hiroshi lijima., High-level expression of a sweet protein, monellin, in the food yeast Candida utilis., Nature Biotechnology 15, 453 - 457 (1997)；非特許文献７：Yutaka Miura, Keiji Kondo, Hiroshi Shimada, Toshiko Saito, Katsumi Nakamura, Norihiko Misawa., Production of lycopene by the food yeast, Candida utilis that does not naturally synthesize carotenoid., Biotechnology and Bioengineering Volume 58 Issue 2-3, Pages 306-308；非特許文献８：Shigehito IKUSHIMA, Toshio FUJII, Osamu KOBAYASHI, Satoshi YOSHIDA and Aruto YOSHIDA., Genetic Engineering of Candida utilis Yeast for Efficient Production of L-Lactic Acid., Bioscience, Biotechnology, and Biochemistry Vol. 73 (2009) , No. 8 pp.1818-1824；非特許文献９：Yi-Ren Hong, Ya-Lei Chen, Lynn Farh, Wen-Jen Yang, Chen-Hua Liao and David Shiuan.,Recombinant Candida utilis for the production of biotin., Applied Microbiology and Biotechnology.,Volume 71(2006), Number 2, 211-221）。また、キャンディダ・ユティリスは、サッカロマイセス属とは異なり、クラブツリー（Ｃｒａｂｔｒｅｅ）効果陰性酵母であることから、グルコース存在下でも呼吸活性が低下しない。これにより、十分に通気を行う条件を採用することにより、エタノールの生成を抑えることができる。さらに、キャンディダ・ユティリスは呼吸活性も高く、一旦生成したエタノールも速やかに再同化されることから、効率的な菌体や代謝産物の生産を可能とする。

【０００７】

酵母でイソプロパノールのような物質を生産する場合、一般に、副産物としてのエタノールの生成が問題となる。このエタノール副生を抑制するために、エタノール生産の重要酵素であるピルビン酸脱炭酸酵素遺伝子（ＰＤＣ）を欠損させる取り組みがなされている（特許文献２：国際公開第２０１０／０９５７５１号パンフレット；特許文献３：国際公開第２０１０／０９５７５０号パンフレット；非特許文献１０：Flikweert MT, Van Der Zanden L, Janssen WM, et al. Pyruvate decarboxylase: an indispensable enzyme for growth of Saccharomyces cerevisiae on glucose. Yeast 12(1996): 247-257；非特許文献１１：van Maris AJA, Winkler AA, Porro D, et al. Homofermentative lactate production cannot sustain anaerobic growth of engineered Saccharomyces cerevisiae: Possible consequence of energy-dependent lactate export. Appl Environ Microbiol 70(2004): 2898-2905）。

【０００８】

【非特許文献１】Chen, J.-S., and S. F. Hiu. Acetone-butanol-isopropanol production by Clostridium beijerinckii (synonym, Clostridium butylicum). Biotechnol. Lett. 1986. 8:371-376

【非特許文献２】T. Hanai, S. Atsumi, and J. C. Liao. Engineered Synthetic Pathwa

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 572 of 3077

y for Isopropanol Production in Escherichia coli. Appl Environ Microbiol, Dec. 2007, p. 7814-7818
【非特許文献３】Toru Jojima & Masayuki Inui & Hideaki Yukawa. Production of isopropanol by metabolically engineered Escherichia coli. Appl Microbiol Biotechnol 2008 77:1219-1224
【非特許文献４】Hiroshi Shimada, Keiji Kondo, Paul D. Fraser, Yutaka Miura, Toshiko Saito, and Norihiko Misawa., Increased Carotenoid Production by the Food Yeast Candida utilis through Metabolic Engineering of the Isoprenoid Pathway., Appl Environ Microbiol, July 1998, p. 2676-2680, Vol. 64, No. 7
【非特許文献５】Yutaka Miura, Keiji Kondo, Toshiko Saito, Hiroshi Shimada, Paul D. Fraser, and Norihiko MisawaProduction of the Carotenoids Lycopene, β-Carotene, and Astaxanthin in the Food Yeast Candida utilis., Appl Environ Microbiol, April 1998, p. 1226-1229, Vol. 64, No. 4
【非特許文献６】Keiji Kondo,, Yutaka Miura, Hidetaka Sone, Kazuo Kobayashi & Hiroshi lijima., High-level expression of a sweet protein, monellin, in the food yeast Candida utilis., Nature Biotechnology 15, 453 - 457 (1997)
【非特許文献７】Yutaka Miura, Keiji Kondo, Hiroshi Shimada, Toshiko Saito, Katsumi Nakamura, Norihiko Misawa., Production of lycopene by the food yeast, Candida utilis that does not naturally synthesize carotenoid., Biotechnology and Bioengineering Volume 58 Issue 2-3, Pages 306-308
【非特許文献８】Shigehito IKUSHIMA, Toshio FUJII, Osamu KOBAYASHI, Satoshi YOSHIDA and Aruto YOSHIDA., Genetic Engineering of Candida utilis Yeast for Efficient Production of L-Lactic Acid., Bioscience, Biotechnology, and Biochemistry Vol. 73 (2009) , No. 8 pp.1818-1824
【非特許文献９】Yi-Ren Hong, Ya-Lei Chen, Lynn Farh, Wen-Jen Yang, Chen-Hua Liao and David Shiuan.,Recombinant Candida utilis for the production of biotin., Applied Microbiology and Biotechnology.,Volume 71(2006), Number 2, 211-221
【非特許文献１０】Flikweert MT, Van Der Zanden L, Janssen WM, et al. Pyruvate decarboxylase: an indispensable enzyme for growth of Saccharomyces cerevisiae on glucose. Yeast 12(1996): 247-257
【非特許文献１１】van Maris AJA, Winkler AA, Porro D, et al. Homofermentative lactate production cannot sustain anaerobic growth of engineered Saccharomyces cerevisiae: Possible consequence of energy-dependent lactate export. Appl Environ Microbiol 70(2004): 2898-2905
【特許文献１】特開２０１１－９７９２９号公報
【特許文献２】国際公開第２０１０／０９５７５１号パンフレット
【特許文献３】国際公開第２０１０／０９５７５０号パンフレット

【発明の概要】

【０００９】

本発明者らは、酵母におけるアセチルＣｏＡ合成酵素（ＡＣＳ）およびアセチルＣｏＡアセチル基転移酵素（ＥＲＧ１０）の活性を増強することにより、酵母内におけるアセチルＣｏＡを経由する物質の生産量が顕著に増加することを見出した。本発明はこの知見に基づくものである。

【００１０】

従って、本発明の目的は、様々な物質の発酵生産において中間物質として関与するアセチルＣｏＡならびにアセトアセチルＣｏＡの合成・代謝に係る酵素活性が増強された酵母菌株、ならびに該酵母菌株を用いた物質の製造法を提供することにある。

【００１１】

そして、本発明による酵母菌株は、以下の酵素：

（ａ）アセチルＣｏＡ合成酵素（ＡＣＳ）；および

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 573 of 3077

（ｂ）アセチルＣｏＡアセチル基転移酵素（ＥＲＧ１０）
の活性が増強されている酵母菌株である。
【００１２】
さらに、本発明による製造法は、本発明による酵母菌株を培養することを含んでなる、物質を製造する方法である。
【００１３】
本発明によれば、アセチルＣｏＡやアセトアセチルＣｏＡが関与する酵母の代謝経路が活性化される。アセチルＣｏＡやアセトアセチルＣｏＡが関与する代謝経路は多岐にわたるため、生成する代謝産物も多種多様である。従って、本発明によれば、酵母菌株の培養によって様々な物質を高効率で製造することが可能となる。また、製造しようとする物質に応じて、必要な酵素の遺伝子を発現可能な形で酵母に組み込み、得られた酵母菌株を培養することによりその物質を製造することも可能である。
【発明の具体的説明】
【００１４】
本発明による酵母菌株では、以下の遺伝子：（ａ）アセチルＣｏＡ合成酵素（ＡＣＳ）の活性を有するポリペプチドをコードする遺伝子；および（ｂ）アセチルＣｏＡアセチル基転移酵素（ＥＲＧ１０）の活性を有するポリペプチドをコードする遺伝子の発現が増強されている。
【００１５】
本明細書では、上記のアセチルＣｏＡ合成酵素（ＡＣＳ）遺伝子およびアセチルＣｏＡアセチル基転移酵素（ＥＲＧ１０）遺伝子とともに、イソプロパノールを合成するための４つの遺伝子を過剰発現する酵母菌株を作製し、これを培養することにより、極めて高濃度のイソプロパノールが生産されることが実証されている。この酵母菌株におけるイソプロパノール生産の経路は図１に示す通りである。図１に示される経路から、イソプロパノールの高い生産量が、アセチルＣｏＡおよびアセトアセチルＣｏＡの合成・代謝に係る酵素活性の増強に起因することは明らかである。従って、本発明によって、イソプロパノールのみならず、アセチルＣｏＡおよびアセトアセチルＣｏＡを経由する様々な物質を高生産することができる。また、図１に示されるように、ＡＣＳおよびＥＲＧ１０は細胞質において機能する酵素である。
【００１６】
細胞質に供給されるアセチルＣｏＡは、フラボノイド骨格を有する化合物、テルペノイド骨格を有する化合物、ステロール骨格を有する化合物、脂肪酸など、様々な有用な二次代謝物質の基幹物質である。よって、酵母菌株が天然の状態もしくは人為的に遺伝子を改変された状態でこれらの二次代謝産物の合成経路を有していれば、本発明の方法によりこれらの物質を高生産させることができる。
【００１７】
一般的に、微生物による物質生産を行う際には、副産物生産を抑制する取り組みがなされる場合が多い。例えば、酵母において物質を生産する場合、一般に、副産物としてのエタノールの生成を抑制するために、エタノール生産に関与する酵素遺伝子、例えばピルビン酸脱炭酸酵素遺伝子（ＰＤＣ）を欠損させる取り組みがなされている（国際公開第２０１０／０９５７５１号パンフレット；国際公開第２０１０／０９５７５０号パンフレット；Flikweert MT, Van Der Zanden L, Janssen WM, et al. Pyruvate decarboxylase: an indispensable enzyme for growth of Saccharomyces cerevisiae on glucose. Yeast 12 (1996): 247-257；van Maris AJA, Winkler AA, Porro D, et al. Homofermentative lactate production cannot sustain anaerobic growth of engineered Saccharomyces cerevisiae: Possible consequence of energy-dependent lactate export. Appl Environ Microbiol 70(2004): 2898-2905）。しかし、本明細書では、上記のＰＤＣを欠損させた酵母菌株よりも、ＰＤＣを欠損させなかった酵母菌株の方が顕著に高いイソプロパノール生産量を示すことが実証されている。よって、本発明による酵母菌株は、ＰＤＣなどのエタノール生産に関与する酵素遺伝子を欠失させずにそのまま維持していることも特徴の一つ

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 574 of 3077

であり、これによりイソプロパノール生産量、ひいてはアセチルＣｏＡやその代謝産物であるアセトアセチルＣｏＡの生産量が顕著に向上することは、驚くべきことである。

【００１８】

本明細書において酵素活性が「増強されている」とは、培養過程の少なくとも一時期において、活性増強のための手段を施されていない酵母と比較してより強い酵素活性が観察されることを意味する。酵素活性の増強は、当該酵素遺伝子の発現を増強すること、あるいは比活性のより高い酵素遺伝子への置き換えや、タンパク質工学的な方法による酵素遺伝子の改変によって実現することができる。本明細書において遺伝子の発現が「増強されている」とは、酵母の遺伝的改変により該酵母におけるその遺伝子の発現量が増加すること、つまり、その遺伝子を過剰発現させることを意味する。このような遺伝的改変の方法としては、例えば、その遺伝子により酵母を形質転換する方法（つまり、その遺伝子を発現し得る形で酵母にトランスフェクトする方法）、その酵母のゲノム上に存在する遺伝子のプロモーター活性を強化する方法、その酵母のゲノム上に存在する遺伝子のプロモーターを他のプロモーターと置換することにより該遺伝子の発現を増強する方法など、様々な方法が知られているが、好ましくはその遺伝子により酵母を形質転換する方法とされる。

【００１９】

酵母

本発明による酵母菌株は、特定の遺伝子が過剰発現するように遺伝的に改変された酵母である。遺伝的改変に用いる酵母は、アセチルＣｏＡまたはアセトアセチルＣｏＡを経由する代謝経路を有する酵母であればよく、特に制限されないが、好ましくは呼吸活性の高いクラブツリー効果陰性酵母とされ、より好ましくはクラブトゥリー陰性酵母であるキャンディダ・ユティリスとされる。キャンディダ・ユティリスの菌株は当技術分野において公知の様々な株、例えば、ＮＢＲＣ０３９６株、ＮＢＲＣ０６１９株、ＮＢＲＣ０９８８株、ＮＢＲＣ０６３９株、ＮＢＲＣ０６２６株、ＮＢＲＣ１０８６株、ＮＢＲＣ１０７０７株、ＡＴＣＣ９２５５株、ＡＴＣＣ２８９５５株、ＡＴＣＣ４４６３８株、ＡＴＣＣ９２０６株、ＡＴＣＣ９６６２１株、ＡＴＣＣ１８２０１株、およびＡＴＣＣ１５２３９株等であってよいが、好ましくはＮＢＲＣ１０８６株、ＡＴＣＣ９２５５株またはＡＴＣＣ１５２３９株とされる。

【００２０】

活性が増強される酵素およびその遺伝子

本発明による酵母菌株において活性が増強される酵素は、以下の酵素：（ａ）アセチルＣｏＡ合成酵素（ＡＣＳ）；および（ｂ）アセチルＣｏＡアセチル基転移酵素（ＥＲＧ１０）である。本発明の好ましい実施態様によれば、本発明による酵母菌株においてはこれら２種類の酵素の遺伝子が過剰発現され、すなわち、以下の遺伝子：（ａ）アセチルＣｏＡ合成酵素（ＡＣＳ）の活性を有するポリペプチドをコードする遺伝子；および（ｂ）アセチルＣｏＡアセチル基転移酵素（ＥＲＧ１０）の活性を有するポリペプチドをコードする遺伝子の発現が増強されている。

【００２１】

本発明による酵母菌株では、さらに、以下の遺伝子：（ｃ）ＡＴＰ－クエン酸除去付加酵素（ＡＣＬ）の活性を有するポリペプチドをコードする遺伝子；（ｄ）ミトコンドリア内膜カルニチン輸送体（ＣＲＣ１）の活性を有するポリペプチドをコードする遺伝子；および（ｅ）カルニチンアセチル転移酵素遺伝子（ＹＡＴ１）の活性を有するポリペプチドをコードする遺伝子からなる群から選択される少なくとも一種の遺伝子の発現が増強されていてもよく、好ましくはこれら（ｃ）、（ｄ）および（ｅ）の全ての遺伝子の発現が増強される。

【００２２】

これらの酵素および遺伝子は様々な酵母に由来するものが知られており、その起源は特に制限されない。これらの酵素および遺伝子は、育種に用いられる酵母の内在性遺伝子およびその発現産物であってよく、あるいは、育種に用いられる酵母以外の他の酵母に由来するものであってもよい。例えば、これらの酵素および遺伝子は、それぞれ独立に、キャ

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 575 of 3077

ンディダ・ユティリスまたはヤロウィア・リポリティカに由来するものとすることができる。それぞれの酵素の起源としては、例えば、アセチルＣｏＡ合成酵素（ＡＣＳ）はキャンディダ・ユティリス酵母、アセチルＣｏＡアセチル基転移酵素（ＥＲＧ１０）はキャンディダ・ユティリス酵母、ミトコンドリア内膜カルニチン輸送体（ＣＲＣ１）はキャンディダ・ユティリス酵母、カルニチンアセチル転移酵素遺伝子（ＹＡＴ１）はキャンディダ・ユティリス酵母、ＡＴＰ－クエン酸除去付加酵素（ＡＣＬ）はヤロウィア・リポリティカ酵母（ヤロウィア・リポリティカ由来のＡＣＬ１とＡＣＬＹがサブユニットを構成する遺伝子）が好ましい。

【００２３】

キャンディダ・ユティリス由来のアセチルＣｏＡ合成酵素（ＡＣＳ）遺伝子には、ＡＣＳ１遺伝子とＡＣＳ２遺伝子がある。ＡＣＳ１遺伝子のコード配列は配列番号１１で表され、コードされるアミノ酸配列は配列番号１２で表される。ＡＣＳ２遺伝子のコード配列は配列番号１３で表され、コードされるアミノ酸配列は配列番号１４で表される。アセチルＣｏＡ合成酵素（ＡＣＳ）は、酢酸をアセチルＣｏＡに変換する酵素である。本発明に用いられるＡＣＳ遺伝子はＡＣＳ１およびＡＣＳ２のいずれであってもよいが、好ましくはＡＣＳ２とされる。また、アセチルＣｏＡ合成酵素（ＡＣＳ）の活性を有するポリペプチドは、配列番号１２または配列番号１４で表されるアミノ酸配列において、１もしくは数個のアミノ酸が欠失、置換、付加もしくは挿入されたアミノ酸配列を含み、かつＡＣＳ活性を有するポリペプチドであってもよい。

【００２４】

キャンディダ・ユティリス由来のアセチルＣｏＡアセチル基転移酵素（ＥＲＧ１０）遺伝子のコード配列は配列番号６１で表され、コードされるアミノ酸配列は配列番号６２で表される。アセチルＣｏＡアセチル基転移酵素（ＥＲＧ１０）は、アセチルＣｏＡをアセトアセチルＣｏＡに変換する酵素である。また、アセチルＣｏＡアセチル基転移酵素（ＥＲＧ１０）の活性を有するポリペプチドは、配列番号６２で表されるアミノ酸配列において、１もしくは数個のアミノ酸が欠失、置換、付加もしくは挿入されたアミノ酸配列を含み、かつＥＲＧ１０活性を有するポリペプチドであってもよい。

【００２５】

ヤロウィア・リポリティカ由来のＡＣＬ１遺伝子のコード配列は配列番号３１で表され、コードされるアミノ酸配列は配列番号３２で表される。また、ＡＣＬ１の活性を有するポリペプチドは、配列番号３２で表されるアミノ酸配列において、１もしくは数個のアミノ酸が欠失、置換、付加もしくは挿入されたアミノ酸配列を含み、かつＡＣＬ１活性を有するポリペプチドであってもよい。さらに、ヤロウィア・リポリティカ由来のＡＣＬＹ遺伝子のコード配列は配列番号３３で表され、コードされるアミノ酸配列は配列番号３４で表される。また、ＡＣＬＹの活性を有するポリペプチドは、配列番号３４で表されるアミノ酸配列において、１もしくは数個のアミノ酸が欠失、置換、付加もしくは挿入されたアミノ酸配列を含み、かつＡＣＬＹ活性を有するポリペプチドであってもよい。これらのＡＣＬ１とＡＣＬＹをサブユニットとするＡＴＰ－クエン酸除去付加酵素（ＡＣＬ）は、クエン酸をアセチルＣｏＡに変換する酵素である。

【００２６】

キャンディダ・ユティリス由来のミトコンドリア内膜カルニチン輸送体（ＣＲＣ１）遺伝子のコード配列は配列番号２３で表され、コードされるアミノ酸配列は配列番号２４で表される。ミトコンドリア内膜カルニチン輸送体（ＣＲＣ１）は、アセチルカルニチンをミトコンドリアから細胞質へ輸送する酵素である。また、ミトコンドリア内膜カルニチン輸送体（ＣＲＣ１）の活性を有するポリペプチドは、配列番号２４で表されるアミノ酸配列において、１もしくは数個のアミノ酸が欠失、置換、付加もしくは挿入されたアミノ酸配列を含み、かつＣＲＣ１活性を有するポリペプチドであってもよい。

【００２７】

キャンディダ・ユティリス由来のカルニチンアセチル転移酵素遺伝子（ＹＡＴ１）遺伝子のコード配列は配列番号２７で表され、コードされるアミノ酸配列は配列番号２８で表

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 576 of 3077

される。カルニチンアセチル転移酵素遺伝子（ＹＡＴ１）は、アセチルカルニチンをアセチルＣｏＡに変換する酵素である。また、カルニチンアセチル転移酵素遺伝子（ＹＡＴ１）の活性を有するポリペプチドは、配列番号２８で表されるアミノ酸配列において、１もしくは数個のアミノ酸が欠失、置換、付加もしくは挿入されたアミノ酸配列を含み、かつＹＡＴ１活性を有するポリペプチドであってもよい。

【００２８】

アミノ酸の欠失、置換、付加、又は挿入は、上記ポリペプチドをコードする遺伝子を、当技術分野で公知の手法によって改変することによって行うことができる。遺伝子に変異を導入するには、Ｋｕｎｋｅｌ法又はＧａｐｐｅｄ ｄｕｐｌｅｘ法等の公知手法又はこれに準ずる方法により行うことができ、例えば部位特異的突然変異誘発法を利用した変異導入用キット、例えばＭｕｔａｎｔ－Ｋ（タカラバイオ社）やＭｕｔａｎｔ－Ｇ（タカラバイオ社）などを用いて、あるいは、タカラバイオ社のＬＡ ＰＣＲ ｉｎ ｖｉｔｒｏ Ｍｕｔａｇｅｎｅｓｉｓシリーズキット、ＫＯＤ－Ｐｌｕｓ－Ｍｕｔａｇｅｎｅｓｉｓ Ｋｉｔ（ＴＯＹＯＢＯ）などを用いて変異を導入することができる。上記の各酵素の活性は、当技術分野において公知の手法により確認することができる。

【００２９】

上記の遺伝子に代えて、それぞれの同等物を用いることもできる。この同等物は、それぞれの遺伝子と同等の機能を有することを条件に、一部のヌクレオチド残基が異なる遺伝子を意味する。このような同等物としては、それぞれのヌクレオチド配列と７０％以上、好ましくは８０％以上、より好ましくは８５％以上、さらに好ましくは９０％以上、最も好ましくは９５％以上の相同性（配列同一性）があり、かつ各酵素の活性を有するポリペプチドをコードするヌクレオチド配列を含む遺伝子が挙げられる。前記同等物としてはさらに、それぞれのヌクレオチド配列もしくはその相補配列とストリンジェントな条件下でハイブリダイズし、かつ各酵素の活性を有するポリペプチドをコードするヌクレオチド配列を含む遺伝子が挙げられる。前記同等物としてはさらに、それぞれのヌクレオチド配列において１もしくは数個のヌクレオチド残基が欠失、置換、付加、または挿入された配列を含み、かつ各酵素の活性を有するポリペプチドをコードするヌクレオチド配列を含む遺伝子が挙げられる。

【００３０】

ここで、ヌクレオチド残基の欠失、置換、付加、又は挿入は、上記配列を含む遺伝子を、当技術分野で公知の手法によって改変することによって行うことができる。遺伝子に変異を導入するには、Ｋｕｎｋｅｌ法又はＧａｐｐｅｄ ｄｕｐｌｅｘ法等の公知手法又はこれに準ずる方法により行うことができ、例えば部位特異的突然変異誘発法を利用した変異導入用キットを用いればよい。例えばＭｕｔａｎｔ－Ｋ（タカラバイオ社）やＭｕｔａｎｔ－Ｇ（タカラバイオ社）などを用いて、あるいは、タカラバイオ社のＬＡ ＰＣＲ ｉｎ ｖｉｔｒｏ Ｍｕｔａｇｅｎｅｓｉｓシリーズキット、ＫＯＤ－Ｐｌｕｓ－Ｍｕｔａｇｅｎｅｓｉｓ Ｋｉｔ（ＴＯＹＯＢＯ）などを用いて変異を導入することができる。各酵素の活性は、当技術分野において公知の手法により確認することができる。

【００３１】

相同性（配列同一性）を示す数値（％）は、塩基配列比較用プログラム：例えばＧＥＮＥＴＹＸ－ＷＩＮ７．０．０を用いて、デフォルト（初期設定）のパラメーターにより算出されるものである。すなわち、酵母染色体上の各遺伝子が、同一ではないが同等の機能、すなわち各活性を有するポリペプチドをコードする遺伝子によって相同組換え等を介して置換されていてもよい。各酵素の活性は、当技術分野において公知の手法により確認することができる。

【００３２】

ストリンジェントな条件とは、例えば、Ｒａｐｉｄ－Ｈｙｂ Ｂｕｆｆｅｒ（ＧＥヘルスケアバイオサイエンス社製）を用い、温度条件を好ましくは４０～７０℃、より好ましくは６０℃として、その他は添付のプロトコールに従って行うハイブリダイゼーション条件である。その後、例えば当業者に公知の一般的な方法を用い、２×ＳＳＣと０．１％（

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 577 of 3077

ｗ／ｖ）ＳＤＳから成り立つ溶液での５分間の洗浄、続いて１×ＳＳＣと０．１％（ｗ／ｖ）ＳＤＳから成り立つ溶液での１０分間の洗浄、さらに０．１×ＳＳＣと０．１％（ｗ／ｖ）ＳＤＳから成り立つ溶液での１０分間の洗浄を行うことを指す。ただしハイブリダイゼーション時の温度条件や、その後のメンブレンの洗浄に用いる溶液の塩濃度等の条件を適宜設定することにより、ある一定（７０％、８０％、８５％、９０％、９５％のいずれか）以上の相同性を有する塩基配列を含むＤＮＡをクローニングできる。そのようにして得られる遺伝子が、配列上は同一ではないが同等の機能、すなわち各活性を有するポリペプチドをコードする遺伝子によって相同組換え等を介して置換されていてもよい。各酵素の活性は、当技術分野において公知の手法により確認することができる。

【００３３】

有用物質を合成するための遺伝子

本発明による酵母菌株は、所望の有用物質の製造に用いるために、該物質の生産に必要となる遺伝子で形質転換されていてもよい。例えば、本発明の方法によってアセチルＣｏＡを経由する代謝産物を製造するために、本発明による酵母菌株は、アセチルＣｏＡを経由する代謝産物を合成するための遺伝子により形質転換されていてもよい。必要な遺伝子は、製造しようとする物質に応じて、当業者であれば適切に選択することができる。

【００３４】

例えば、イソプロパノールを製造しようとする場合には、さらなる遺伝的改変として、イソプロパノール合成遺伝子で酵母菌株を形質転換することができる。イソプロパノール合成遺伝子としては、特開２０１１－９７９２９号公報に記載の酵素遺伝子群など、当業者に公知の遺伝子を用いることができ、例えば、ＣｏＡ基転移酵素サブユニットＡ（ｃｔｆＡ）、同サブユニットＢ（ｃｔｆＢ）、アセト酢酸脱炭酸酵素（ａｄｃ）および二級アルコール脱水素酵素（ｓａｄｈ）の４つの遺伝子の組み合わせを用いることができる。

【００３５】

本発明の一つの実施態様によれば、本発明による酵母菌株は、さらなる遺伝的改変として、ＣｏＡ基転移酵素サブユニットＡ（ｃｔｆＡ）、ＣｏＡ転移酵素サブユニットＢ（ｃｔｆＢ）、アセト酢酸脱炭酸酵素（ａｄｃ）および二級アルコール脱水素酵素（ｓａｄｈ）の活性を有するポリペプチドをコードしている４種の遺伝子により形質転換されている。このような酵母菌株はイソプロパノールの製造に用いることができ、この製造法は、本実施態様による酵母菌株を培養することを含む。

【００３６】

これらの遺伝子は様々な生物に由来するものが知られており、その起源は特に制限されない。例えば、これらの遺伝子は、それぞれ独立に、クロストリジウム・アセトブチリカムまたはクロストリジウム・ベイジェリンキに由来するものとすることができる。それぞれの酵素の起源としては、例えば、ＣｏＡ転移酵素遺伝子サブユニットＡ（ｃｔｆＡ）および同サブユニットＢ（ｃｔｆＢ）はクロストリジウム・アセトブチリカムが好ましく、アセト酢酸脱炭酸酵素（ａｄｃ）および二級アルコール脱水素酵素（ｓａｄｈ）はクロストリジウム・ベイジェリンキが好ましい。

【００３７】

クロストリジウム・アセトブチリカム由来のＣｏＡ転移酵素サブユニットＡ（ｃｔｆＡ）のアミノ酸配列は配列番号２で表される。ＣｏＡ転移酵素サブユニットＡ（ｃｔｆＡ）は、同サブユニットＢ（ｃｔｆＢ）と一緒になって、アセトアセチルＣｏＡをアセト酢酸に変換する酵素である。配列番号２で表されるアミノ酸配列をコードする遺伝子配列は、クロストリジウム・アセトブチリカム由来の同遺伝子の配列など、当業者であれば適宜選択することができるため、特に制限されないが、好ましくは本発明による酵母菌株の種において頻度の高いコドンを使用した配列とされる。例えば、キャンディダ・ユティリスにおけるコドン頻度に適合した遺伝子配列としては、配列番号１で表されるヌクレオチド配列が挙げられる。ＣｏＡ転移酵素サブユニットＡ（ｃｔｆＡ）の活性を有するポリペプチドは、配列番号２で表されるアミノ酸配列において、１もしくは数個のアミノ酸が欠失、置換、付加もしくは挿入されたアミノ酸配列を含み、かつｃｔｆＡ活性を有するポリペプ

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 578 of 3077

チドであってもよい。

【００３８】

クロストリジウム・アセトブチリカム由来のＣｏＡ転移酵素サブユニットＢ（ｃｔｆＢ）のアミノ酸配列は配列番号４で表される。ＣｏＡ転移酵素サブユニットＢ（ｃｔｆＢ）は、同サブユニットＡ（ｃｔｆＡ）と一緒になって、アセトアセチルＣｏＡをアセト酢酸に変換する酵素である。配列番号４で表されるアミノ酸配列をコードする遺伝子配列は、クロストリジウム・アセトブチリカム由来の同遺伝子の配列など、当業者であれば適宜選択することができるため、特に制限されないが、好ましくは本発明による酵母菌株の種において頻度の高いコドンを使用した配列とされる。例えば、キャンディダ・ユティリスにおけるコドン頻度に適合した遺伝子配列としては、配列番号３で表されるヌクレオチド配列が挙げられる。ＣｏＡ転移酵素サブユニットＢ（ｃｔｆＢ）の活性を有するポリペプチドは、配列番号４で表されるアミノ酸配列において、１もしくは数個のアミノ酸が欠失、置換、付加もしくは挿入されたアミノ酸配列を含み、かつｃｔｆＢ活性を有するポリペプチドであってもよい。

【００３９】

クロストリジウム・ベイジェリンキ由来のアセト酢酸脱炭酸酵素（ａｄｃ）のアミノ酸配列は配列番号６で表される。アセト酢酸脱炭酸酵素（ａｄｃ）は、アセト酢酸をアセトンに変換する酵素である。配列番号６で表されるアミノ酸配列をコードする遺伝子配列は、クロストリジウム・ベイジェリンキ由来の同遺伝子の配列など、当業者であれば適宜選択することができるため、特に制限されないが、好ましくは本発明による酵母菌株の種において頻度の高いコドンを使用した配列とされる。例えば、キャンディダ・ユティリスにおけるコドン頻度に適合した遺伝子配列としては、配列番号５で表されるヌクレオチド配列が挙げられる。アセト酢酸脱炭酸酵素（ａｄｃ）の活性を有するポリペプチドは、配列番号６で表されるアミノ酸配列において、１もしくは数個のアミノ酸が欠失、置換、付加もしくは挿入されたアミノ酸配列を含み、かつａｄｃ活性を有するポリペプチドであってもよい。

【００４０】

クロストリジウム・ベイジェリンキ由来の二級アルコール脱水素酵素（ｓａｄｈ）のアミノ酸配列は配列番号８で表される。二級アルコール脱水素酵素（ｓａｄｈ）は、アセトンをイソプロパノールに変換する酵素である。配列番号８で表されるアミノ酸配列をコードする遺伝子配列は、クロストリジウム・ベイジェリンキ由来の同遺伝子の配列など、当業者であれば適宜選択することができるため、特に制限されないが、好ましくは本発明による酵母菌株の種において頻度の高いコドンを使用した配列とされる。例えば、キャンディダ・ユティリスにおけるコドン頻度に適合した遺伝子配列としては、配列番号７で表されるヌクレオチド配列が挙げられる。二級アルコール脱水素酵素（ｓａｄｈ）の活性を有するポリペプチドは、配列番号８で表されるアミノ酸配列において、１もしくは数個のアミノ酸が欠失、置換、付加もしくは挿入されたアミノ酸配列を含み、かつｓａｄｈ活性を有するポリペプチドであってもよい。

【００４１】

アミノ酸の欠失、置換、付加、又は挿入は、上記ポリペプチドをコードする遺伝子を、当技術分野で公知の手法によって改変することによって行うことができる。遺伝子に変異を導入するには、Ｋｕｎｋｅｌ法又はＧａｐｐｅｄ ｄｕｐｌｅｘ法等の公知手法又はこれに準ずる方法により行うことができ、例えば部位特異的突然変異誘発法を利用した変異導入用キット、例えばＭｕｔａｎｔ－Ｋ（タカラバイオ社）やＭｕｔａｎｔ－Ｇ（タカラバイオ社）などを用いて、あるいは、タカラバイオ社のＬＡ ＰＣＲ ｉｎ ｖｉｔｒｏ Ｍｕｔａｇｅｎｅｓｉｓシリーズキット、ＫＯＤ－Ｐｌｕｓ－Ｍｕｔａｇｅｎｅｓｉｓ Ｋｉｔ（ＴＯＹＯＢＯ）などを用いて変異を導入することができる。上記の各酵素の活性は、当技術分野において公知の手法により確認することができる。

【００４２】

上記の遺伝子に代えて、それぞれの同等物を用いることもできる。この同等物は、それ

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 579 of 3077

ぞれの遺伝子と同等の機能を有することを条件に、一部のヌクレオチド残基が異なる遺伝子を意味する。このような同等物としては、それぞれのヌクレオチド配列と７０％以上、好ましくは８０％以上、より好ましくは８５％以上、さらに好ましくは９０％以上、最も好ましくは９５％以上の相同性（配列同一性）があり、かつ各酵素の活性を有するポリペプチドをコードするヌクレオチド配列を含む遺伝子が挙げられる。前記同等物としてはさらに、それぞれのヌクレオチド配列もしくはその相補配列とストリンジェントな条件下でハイブリダイズし、かつ各酵素の活性を有するポリペプチドをコードするヌクレオチド配列を含む遺伝子が挙げられる。前記同等物としてはさらに、それぞれのヌクレオチド配列において１もしくは数個のヌクレオチド残基が欠失、置換、付加、または挿入された配列を含み、かつ各酵素の活性を有するポリペプチドをコードするヌクレオチド配列を含む遺伝子が挙げられる。

【００４３】

ここで、ヌクレオチド残基の欠失、置換、付加、又は挿入は、上記配列を含む遺伝子を、当技術分野で公知の手法によって改変することによって行うことができる。遺伝子に変異を導入するには、Ｋｕｎｋｅｌ法又はＧａｐｐｅｄ ｄｕｐｌｅｘ法等の公知手法又はこれに準ずる方法により行うことができ、例えば部位特異的突然変異誘発法を利用した変異導入用キットを用いればよい。例えばＭｕｔａｎｔ－Ｋ（タカラバイオ社）やＭｕｔａｎｔ－Ｇ（タカラバイオ社）などを用いて、あるいは、タカラバイオ社のＬＡ ＰＣＲ ｉｎ ｖｉｔｒｏ Ｍｕｔａｇｅｎｅｓｉｓシリーズキット、ＫＯＤ－Ｐｌｕｓ－Ｍｕｔａｇｅｎｅｓｉｓ Ｋｉｔ（ＴＯＹＯＢＯ）などを用いて変異を導入することができる。各酵素の活性は、当技術分野において公知の手法により確認することができる。

【００４４】

相同性（配列同一性）を示す数値（％）は、塩基配列比較用プログラム：例えばＧＥＮＥＴＹＸ－ＷＩＮ７．０．０を用いて、デフォルト（初期設定）のパラメーターにより算出されるものである。すなわち、酵母染色体上の各遺伝子が、同一ではないが同等の機能、すなわち各活性を有するポリペプチドをコードする遺伝子によって相同組換え等を介して置換されていてもよい。各酵素の活性は、当技術分野において公知の手法により確認することができる。

【００４５】

ストリンジェントな条件とは、例えば、Ｒａｐｉｄ－Ｈｙｂ Ｂｕｆｆｅｒ（ＧＥヘルスケアバイオサイエンス社製）を用い、温度条件を好ましくは４０～７０℃、より好ましくは６０℃として、その他は添付のプロトコールに従って行うハイブリダイゼーション条件である。その後、例えば当業者に公知の一般的な方法を用い、２×ＳＳＣと０．１％（ｗ／ｖ）ＳＤＳから成り立つ溶液での５分間の洗浄、続いて１×ＳＳＣと０．１％（ｗ／ｖ）ＳＤＳから成り立つ溶液での１０分間の洗浄、さらに０．１×ＳＳＣと０．１％（ｗ／ｖ）ＳＤＳから成り立つ溶液での１０分間の洗浄を行うことを指す。ただしハイブリダイゼーション時の温度条件や、その後のメンブレンの洗浄に用いる溶液の塩濃度等の条件を適宜設定することにより、ある一定（７０％、８０％、８５％、９０％、９５％のいずれか）以上の相同性を有する塩基配列を含むＤＮＡをクローニングできる。そのようにして得られる遺伝子が、配列上は同一ではないが同等の機能、すなわち各活性を有するポリペプチドをコードする遺伝子によって相同組換え等を介して置換されていてもよい。各酵素の活性は、当技術分野において公知の手法により確認することができる。

【００４６】

イソプロパノール以外にも、リコペンなどの特定のカロテノイド、スクアレンあるいはスクアレンから生産される特定のステロールなどの物質を生産する酵母菌株を構築することが可能である。例えば、Ｅｒｗｉｎｉａ ｕｒｅｄｏｖｏｒａ由来のＧＧＰＰ合成酵素遺伝子、フィトエン合成酵素遺伝子、フィトエン不飽和化酵素遺伝子を発現し、内在性のＨＭＧ－ＣｏＡ還元酵素遺伝子（ＨＭＧ１）を過剰発現するキャンディダ・ユティリスはリコペンを高蓄積することが報告されている（Shimada H, Kondo K, Fraser PD, Miura Y, Saito T, Misawa N. Increased carotenoid production by the food yeast Candida u

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 580 of 3077

tilis through metabolic engineering of the isoprenoid pathway. Appl Environ Microbiol. 1998 Jul;64(7):2676-80）。また、ザイモステロール-２４-メチルトランスフェラーゼ遺伝子およびエルゴスタ５，７，２４，（２８）－トリエノール－２２－脱水素酵素遺伝子の発現に欠損を持ち、ＨＭＧ１が過剰発現するように操作されたサッカロマイセス・セレビシエ組み換え株は、高濃度のスクアレンを蓄積することが知られている（特開平５－１９２１８４号公報）。これらの手法を用いることにより、リコペン、スクアレンを生産する酵母菌株を構築することが可能である。

【００４７】

酵母菌株の分子育種

本発明による酵母菌株の分子育種は、上記の遺伝子群を過剰発現させるための遺伝的改変を酵母菌株に施すこと、例えば、上記の遺伝子群を発現可能な状態で酵母菌株に導入することによって行うことができる。これら遺伝子群は、相同組換え等の技術により酵母の染色体上に組み込まれることが好ましい。このような染色体上への組込みを実現するための相同組換え用遺伝子配列の選択は、当業者において周知であり、当業者であれば必要に応じて適切な相同組換え用遺伝子配列を選択して相同組換え用ＤＮＡ断片を構成することができる。

【００４８】

上記の遺伝子群が酵母ゲノムに組込まれる染色体上の位置は特に制限されるものではないが、例えば、オロチジン５’リン酸デカルボキシラーゼ遺伝子（ＵＲＡ３）座、アルファアミノアジピン酸還元酵素遺伝子（ＬＹＳ２）座などを選択することができる。また、これらの遺伝子は、組み込まれた遺伝子座に当初から存在するプロモーター、例えばＵＲＡ３遺伝子プロモーター、ＬＹＳ２遺伝子プロモーターなどの制御下に置いてもよいが、グリセルアルデヒド－３－リン酸脱水素酵素（ＧＡＰ）遺伝子プロモーターなどの他のプロモーターの制御下に置いてもよい。

【００４９】

本発明の好ましい実施態様によれば、本発明による酵母菌株の育種に用いられる酵母はキャンディダ・ユティリスとされる。キャンディダ・ユティリスは、その倍数性が高く、胞子を形成しない。倍数性の高い株の遺伝子に変異を導入しようとする場合、１倍体の株に比べて、その変異を重度に加える必要があるが、その際に、変異を与えたい遺伝子ではない遺伝子にも変異が加えられる可能性が高まると考えられる。従って、キャンディダ・ユティリスの遺伝子に変異を導入する場合には、標的とする遺伝子のみに効率よく多重に変異を加えることができる技術を用いることが好ましい。キャンディダ・ユティリスの形質転換法としては、特開２００３－１４４１８５号公報に記載の技術が挙げられる。この文献では、利用可能なベクターとして、キャンディダ・ユティリスの染色体ＤＮＡと相同な配列と、選択マーカー遺伝子とを含んでなり、相同組換えによって異種遺伝子をキャンディダ・ユティリスの染色体ＤＮＡに組み込むことができるもの、あるいは、キャンディダ・ユティリスで自律複製機能を有するＤＮＡ配列と、選択マーカー遺伝子とを含んでなり、高い頻度でキャンディダ・ユティリスを形質転換できるものが開発されている。

【００５０】

形質転換系で利用可能な選択マーカー遺伝子としては、酵母において機能し得る薬剤耐性マーカー、好ましくはシクロヘキシミド耐性型Ｌ４１遺伝子、ジェネティシン（Ｇ４１８）耐性を付与する遺伝子、またはハイグロマイシンＢ耐性を付与する遺伝子などがある。ジェネティシン（Ｇ４１８）耐性を付与する遺伝子、またはハイグロマイシンＢ耐性を付与する遺伝子は、野生の酵母には存在しない配列であるため、標的の遺伝子座に組込まれる確率が高いと考えられている。また、これらの薬剤耐性遺伝子は、その宿主の形質に与える影響も小さいと考えられている（Ｂａｇａｎｚ Ｆら，１３（１６）：１５６３－７３．，１９９７）（Ｃｏｒｄｅｒｏ Ｏｔｅｒｏ Ｒら，Ａｐｐｌ Ｍｉｃｒｏｂｉｏｌ Ｂｉｏｔｅｃｈｎｏｌ，４６（２）：１４３－８．，１９９６）。

【００５１】

他の形質転換系として、バクテリオファージＰ１由来のＣｒｅ－ｌｏｘＰ系が挙げられ

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 581 of 3077

る。これは、２つの３４ｂｐのｌｏｘＰ配列間での部位特異的組換えシステムであり、この組換えはＣｒｅ遺伝子がコードするＣｒｅ組換え酵素によって触媒される。このシステムは、サッカロマイセス・セレビシエなどの酵母細胞においても機能することが報告されており、２つのｌｏｘＰ配列の間に配置された選択マーカー遺伝子は、ｌｏｘＰ配列間の組換えによって除去されることが知られている（Ｇｕｌｄｅｎｅｒ，Ｕ．ら，Ｎｕｃｌｅｉｃ Ａｃｉｄｓ Ｒｅｓ．，２４，２５１９－２４．，１９９６）。このシステムはクルイベロマイセス・ラクティス（Ｋｌｕｙｖｅｒｏｍｙｃｅｓ ｌａｃｔｉｓ）など、複数の酵母種で利用されている（Ｓｔｅｅｎｓｍａ，Ｈ．Ｙ．ら，Ｙｅａｓｔ，１８，４６９－７２．，２００１）。

【００５２】

物質の製造方法

本発明による酵母菌株を適当な炭素源の存在下で培養することにより、培養物中に物質（例えばイソプロパノール）を製造することができる。本発明による物質（例えばイソプロパノール）製造法によれば、培養系から物質（例えばイソプロパノール）を分離する工程を実施することにより、物質（例えばイソプロパノール）を得ることができる。なお、本発明において培養物とは、培養上清の他、培養細胞あるいは菌体、細胞もしくは菌体の破砕物を包含している。

【００５３】

本発明による酵母菌株の培養にあたっては、酵母の種類に応じて培養方法や培養条件を選択することができる。培養方法としては、例えば、試験管、フラスコあるいはジャーファーメンターを用いた液体培養法を挙げることができ、回分培養、半回分培養などの培養形式を採用できる。振盪速度および通気量は、一般的には５０～２００ｒｐｍ、好ましくは８０～１５０ｒｐｍ、あるいはこの振盪速度に相当する通気量とされる。初期の菌体量としてのOD_{600}は、一般的には０．１～３０、好ましくは０．５～１０とされる。さらに、本発明による酵母菌株の培養にあたっては、振盪速度は速い方が好ましく（つまり通気量は多い方が好ましく）、初期OD_{600}は高い方が好ましい。例えば、試験管やフラスコでの培養では、振盪速度は約１２０ｒｐｍまたはそれ以上、初期OD_{600}は約１０とすることが好ましい。

【００５４】

本発明による物質（例えばイソプロパノール）製造法において、培地の組成は、酵母が生育し、且つ目的の物質を製造できる各種栄養素を含む組成であればよく、特に限定されない。培地に含まれる資化炭素源としては、例えば、グルコースのほかにもキシロースやスクロースでも資化できれば用いることができる。本発明の好ましい実施態様によれば、炭素源はグルコースを含む。培地中のグルコース濃度は、目的の物質の生産効率を確認しながら、当業者であれば適宜調節することができる。特に、本発明による酵母菌株の培養では、培地中のグルコース濃度は、比較的低濃度に保つことが好ましい。例えば、１９２時間の培養において、グルコースの初期濃度を１００ｇ／Ｌとし、培養７２時間後に終濃度１００ｇ／Ｌとなるようにグルコース固体を添加した場合に比べて、グルコースの初期濃度を５０ｇ／Ｌとし、４８時間後、９６時間後および１４４時間後に終濃度５０ｇ／Ｌとなるようにグルコース固体を添加した場合の方が、物質の生産効率が高い。

【００５５】

培地に含まれる栄養源としては、例えば、酵母エキス、ペプトン、ホエーなどが用いられるが、ＹＰ（１０ｇ／Ｌ酵母エキス、２０ｇ／Ｌペプトン）に上記の資化炭素源を加えた培地、例えばＹＰＤ５（５０ｇ／Ｌグルコース、１０ｇ／Ｌ酵母エキス、２０ｇ／Ｌペプトン）等の培地で、さらに適宜ｐＨ調整されたものが便利である。

【００５６】

発酵温度は、用いる酵母の生育可能な範囲で選択することができる。発酵温度は、例えば、約１５℃～４５℃とすることができ、より好ましくは２０～４０℃、さらに好ましくは２５～３５℃、最も好ましくは約３０℃とする。また、発酵過程における培地のｐＨは３～８に保持することが好ましく、より好ましくはｐＨ４～７であり、必要に応じてｐＨ

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 582 of 3077

調節剤を用いることができる。用いるｐＨ調節剤としては、炭酸カルシウム、水酸化ナトリウム、水酸化カリウム等が挙げられるが、好ましくは炭酸カルシウムとされる。

【００５７】

物質（例えばイソプロパノール）製造に要する反応時間は特に限定されず、本発明の効果が認められる限り任意の反応時間で実施される。これらの条件の最適化は、当業者であれば容易に行うことができる。

【００５８】

本発明による物質（例えばイソプロパノール）の製造法においては、このようにして製造した物質（例えばイソプロパノール）成分を培地から分離・回収するが、そのための具体的な手順は、分離・回収しようとする物質に応じて、当業者であれば容易に選択することができる。

【実施例】

【００５９】

以下、実施例を挙げて本発明をより具体例に説明するが、これら実施例は本発明の技術的範囲を制限するものではない。

【００６０】

実施例１：高いイソプロパノール耐性を示すキャンディダ・ユティリス株の選抜

キャンディダ・ユティリス　ＮＢＲＣ０３９６株、ＮＢＲＣ０６１９株、ＮＢＲＣ０９８８株、ＮＢＲＣ０６３９株、ＮＢＲＣ０６２６株、ＮＢＲＣ１０８６株、ＮＢＲＣ１０７０７株、ＡＴＣＣ９２５５株、ＡＴＣＣ２８９５５株、ＡＴＣＣ４４６３８株、ＡＴＣＣ９２０６株、ＡＴＣＣ９６６２１株、ＡＴＣＣ１８２０１株、およびＡＴＣＣ１５２３９株を、ＹＰＤ２（２０ｇ／Ｌペプトン、１０ｇ／Ｌ酵母エキス、２０ｇ／Ｌグルコース）に接種し、３０℃で培養を行った。定常期の酵母培養液を１０倍ごとに段階希釈し、１０～１００ｇ／ＬのイソプロパノールをふくむＹＰＤ２寒天培地にスポットし、３０℃で培養した。３日間培養を継続し、コロニーの形成を観察した。結果を表１に記載する。

【００６１】

【表１】

表１

菌株	生育限界 ＩＰＡ濃度（ｇ／ｌ）
ＮＢＲＣ０３９６	５０
ＮＢＲＣ０６１９	６０
ＮＢＲＣ０９８８	４０
ＮＢＲＣ０６３９	５０
ＮＢＲＣ０６２６	５０
ＮＢＲＣ１０８６	６５
ＮＢＲＣ１０７０７	５５
ＡＴＣＣ９２５５	６５
ＡＴＣＣ２８９５５	５５
ＡＴＣＣ４４６３８	６０
ＡＴＣＣ９２０６	６０
ＡＴＣＣ９６６２１	５５
ＡＴＣＣ１８２０１	５０
ＡＴＣＣ１５２３９	６５

【００６２】

以上の結果より、アッセイを行った株の中では、ＮＢＲＣ１０８６株、ＡＴＣＣ９２５５株、およびＡＴＣＣ１５２３９株が、最も高い６５ｇ／ＬのＩＰＡ耐性を示すことが明らかとなった。

【００６３】

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 583 of 3077

実施例２：ピルビン酸脱炭酸酵素遺伝子が破壊されたキャンディダ・ユティリスＡＴＣＣ１５２３９株の構築

キャンディダ・ユティリスＡＴＣＣ１５２３９株において、その内因性遺伝子であるピルビン酸脱炭酸酵素遺伝子（ＰＤＣ）を破壊した。この遺伝子破壊は、Ｃｒｅ－ｌｏｘＰシステムを用いて国際公開第２０１０／０９５７５０号パンフレットと同様の方法で行った。２回の破壊により、ＰＤＣ１アレルが完全に欠損した株（ｄＰＤＣ１株）を構築した。ｄＰＤＣ１株は完全にはエタノール生産能を欠損しなかったが、エタノール生産能は著しく低下していた。

【００６４】

実施例３：クロストリジウム・アセトブチリカム由来のＣｏＡ転移酵素サブユニットＡ遺伝子、同サブユニットＢ遺伝子、クロストリジウム・ベイジェリンキ由来のアセト酢酸脱炭酸酵素遺伝子および二級アルコール脱水素酵素遺伝子を発現するベクターの構築

イソプロパノール生合成に関わる４つの遺伝子、クロストリジウム・アセトブチリカム由来のＣｏＡ転移酵素遺伝子サブユニットＡ（ｃｔｆＡ）、同サブユニットＢ（ｃｔｆＢ）、クロストリジウム・ベイジェリンキ由来のアセト酢酸脱炭酸酵素（ａｄｃ）および二級アルコール脱水素酵素（ｓａｄｈ）を効率良くキャンディダ・ユティリスで発現させるために、以下の項目を指標として、天然に存在しない新規な遺伝子配列を設計し、合成した：

（１）キャンディダ・ユティリスにおいて多用されているコドンの使用；

（２）遺伝子クローニングに不適当な制限酵素部位の排除；および

（３）染色体組み込み型発現ベクターに導入するための、５’末端におけるＸｂａＩサイトおよび３’末端におけるＢａｍＨＩサイトの付与。

【００６５】

合成されたＣｏＡ転移酵素遺伝子サブユニットＡ遺伝子、同サブユニットＢ遺伝子、アセト酢酸脱炭酸酵素遺伝子および二級アルコール脱水素酵素遺伝子のヌクレオチド配列を、それぞれ配列番号１、３、５および７に示す。また、これらのヌクレオチド配列によってコードされるアミノ酸配列を、それぞれ配列番号２、４、６および８に示す。これらの遺伝子を有するプラスミドをそれぞれＧＡ０９１４２－ｃｔｆＡ、ＧＡ０９１４２－ｃｔｆＢ、ＧＡ０９１４２－ａｄｃ、ＧＡ０９１４２－ｓａｄｈと名付けた。

【００６６】

ｐＶＴ９２は、キャンディダ・ユティリスのオロチジン５’酸脱炭酸酵素遺伝子（ＵＲＡ３）座への組み込みベクターであり、ハイグロマイシン耐性遺伝子を選択マーカーとして有している（国際公開第２０１０／０９５７５０号パンフレット）。ｐＶＴ９２は、ＢｇｌＩＩ処理してリニアＤＮＡにすることにより、染色体導入することができる。ｐＶＴ９２は、外来遺伝子を発現させるために、ＸｂａＩ／ＢａｍＨＩギャップの外側にそれぞれ、グリセルアルデヒド脱水素酵素遺伝子プロモーターとグリセリン酸３’リン酸リン酸化酵素遺伝子ターミネーターを有している。また、ｐＶＴ９２は、プロモーターの５’末端側およびターミネーターの３’末端側には、それぞれＮｈｅＩサイトおよびＳｐｅＩサイトを有している。

【００６７】

ｐＶＴ２３９は、キャンディダ・ユティリスのアルファアミノアジピン酸還元酵素遺伝子（ＬＹＳ２）座への組み込みベクターであり、ジェネティシン耐性遺伝子を選択マーカーとして有している（国際公開第２０１０／０９５７５０号パンフレット）。ｐＶＴ２３９は、ＮｏｔＩ処理してリニアＤＮＡにすることにより、染色体導入することができる。また、ｐＶＴ２３９は、ジェネティシン耐性遺伝子の５’末端側にＳｐｅＩサイトを有している。

【００６８】

各イソプロパノール合成酵素遺伝子を含む４つのベクターＧＡ０９１４２－ｃｔｆＡ、ＧＡ０９１４２－ｃｔｆＢ、ＧＡ０９１４２－ａｄｃ、およびＧＡ０９１４２－ｓａｄｈを、ＸｂａＩおよびＢａｍＨＩ処理し、アガロースゲル抽出による精製を行った後、同制

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 584 of 3077

限酵素で処理したｐＶＴ９２にライゲーションして、ｐＶＴ２９６、ｐＶＴ２９８、ｐＶＴ３００およびｐＶＴ３０２を得た。ｐＶＴ２９６およびｐＶＴ３００はＳｐｅＩ処理し、脱リン酸化を行った。また、ｐＶＴ２９８およびｐＶＴ３０２はＮｈｅＩ、ＳｐｅＩ処理を行った後、ｃｔｆＢ発現カセット、ｓａｄｈ発現カセットをそれぞれ含むＤＮＡ断片の精製を行い、それぞれｐＶＴ２９６およびｐＶＴ３００に連結して、ｐＶＴ３２１およびｐＶＴ３２５を得た。ｐＶＴ３２５をＮｈｅＩおよびＳｐｅＩで処理した後、ａｄｃ発現カセットとｓａｄｈ発現カセットを含むＤＮＡ断片の精製を行い、ｐＶＴ２３９のＳｐｅＩサイトに導入し、ｐＶＴ３４６を得た。最終的に得られたｐＶＴ３２１はｃｔｆＡとｃｔｆＢのそれぞれの発現カセットとハイグロマイシン耐性遺伝子を有し、キャンディダ・ユティリスのＵＲＡ３座に組み込まれるベクターであり、ｐＶＴ３４６はａｄｃとｓａｄｈのそれぞれの発現カセットとジェネティシン耐性遺伝子を有し、キャンディダ・ユティリスのＬＹＳ２座に組み込まれるベクターである。

【００６９】

実施例４：４つのイソプロパノール合成遺伝子を有するキャンディダ・ユティリス株の構築

ｐＶＴ３２１をＢｇｌＩＩで消化し、エタノール沈殿により濃縮した。得られたＤＮＡ断片を用いて、Ｃ．ｕｔｉｌｉｓ　ＡＴＣＣ１５２３９株、および実施例２で得られたｄＰＤＣ１株を形質転換し、６００μｇ／ｍＬハイグロマイシンを含むＹＰＤ培地に塗抹し、３０℃で２日間培養した。導入した全ての遺伝子を有する株を選抜した。

【００７０】

さらにｐＶＴ３４６をＮｏｔＩで消化し、エタノール沈殿により濃縮した。得られたそれぞれのＤＮＡ断片を用いて上記で得られた２株をそれぞれ形質転換し、６００μｇ／ｍＬハイグロマイシンと２００μｇ／ｍＬジェネティシンを含むＹＰＤ培地に塗抹し、３０℃で２日間培養した。導入した全ての遺伝子を有する株を選抜し、ＡＴＣＣ１５２３９株由来のイソプロパノール生産株をＴＭＳ２７２株、ｄＰＤＣ株由来のイソプロパノール生産株をＴＭＳ２９０株とした。

【００７１】

実施例５：イソプロパノール生産試験

実施例４で得られた形質転換体のイソプロパノール生産能の評価を、以下に示すように実施した。

【００７２】

各種形質転換体を２ｍＬ　ＹＰＤ液体培地（２０ｇ／Ｌペプトン、１０ｇ／Ｌ酵母エキス、２０ｇ／Ｌグルコース）／１４ｍＬ容試験管に接種し、１４０ｒｐｍ、３０℃で２４時間振とう培養した。得られた前培養液を１００ｍＬ　ＹＰＤ１０（２０ｇ／Ｌペプトン、１０ｇ／Ｌ酵母エキス、１００ｇ／Ｌグルコース）／２００ｍＬ容三角フラスコもしくはＹＰＤ１０Ｃ（２０ｇ／Ｌペプトン、１０ｇ／Ｌ酵母エキス、１００ｇ／Ｌグルコース、４５ｇ／Ｌ炭酸カルシウム）／２００ｍＬ容三角フラスコに接種し、初期ＯＤ６００＝１、１００ｒｐｍ、３０℃で振とう培養した。培養液から経時的にサンプリングし、０．２μｍフィルターでろ過した後に、各種成分の定量分析を行った。グルコース、酢酸、エタノールおよびグリセロールは、高速液体クロマトフラフィー（島津製作所製）、ＩＳｅｐ－ＩＯＮ３００カラム（東京化成工業製）および示唆屈折計を用いて定量した。イソプロパノールおよびアセトンは、ガスクロマトフラフィーＧＣ－２０１０　Ｐｌｕｓ（島津製作所製）を用いて定量した。

【００７３】

また、フラスコでの簡易的な中和培養としてＹＰＤ１０Ｃ（２０ｇ／Ｌペプトン、１０ｇ／Ｌ酵母エキス、１００ｇ／Ｌグルコース、４５ｇ／Ｌ炭酸カルシウム）培地を用いても同様の実験を行った。

【００７４】

非中和条件下での培養試験の結果を図２、中和条件下での培養試験の結果を図３に記載する。なお、データは３反復試行の平均で示した。

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 585 of 3077

【００７５】
図２によれば、親株であるＡＴＣＣ１５２３９株およびｄＰＤＣ１株ではイソプロパノールの生産が認められなかったのに対し、イソプロパノール合成遺伝子を導入したＴＭＳ２７２株およびＴＭＳ２９０株では、それぞれ最大で０．１６ｇ／Ｌおよび０．１３ｇ／Ｌのイソプロパノールの生産が認められた。また、ＰＤＣ１遺伝子を破壊した株では糖の消費速度および増殖能が著しく低下していた。ＴＭＳ２７２株では培養２４時間でグルコースが完全に消費され、エタノールの蓄積が認められたが、その後の培養の継続に伴ってエタノール濃度が低下した。エタノール濃度が低下している間もイソプロパノールの生産が認められたことから、ＴＭＳ２７２株は、一度生産したエタノールを二次消費することによりイソプロパノールを生産していることが考えられた。ｄＰＤＣ１株に由来するＴＭＳ２９０株ではイソプロパノール生産速度が低下したことから、酵母におけるイソプロパノール生産においては、基質である細胞質のアセチルＣｏＡがエタノールから供給される経路が最も重要であることが示唆された。エタノールの同化は酸化的反応であることから、これらの結果は、酵母によるイソプロパノール生産においてはキャンディダ・ユティリスのような呼吸活性の高いクラブツリー陰性酵母が適している可能性を示唆している。
【００７６】
図３によれば、培地に炭酸カルシウムを添加したことにより、ＴＭＳ２７２株は１．１ｇ／Ｌのイソプロパノールを生産し、非中和条件より生産量が増大した。また、ＴＭＳ２７２株は、１６．２ｇ／Ｌの酢酸を蓄積しており、何らかの理由により酢酸が蓄積したことでｐＨが低下し、イソプロパノールの生産を阻害していたことが考えられた。また、ＴＭＳ２９０株においても、中和条件下ではイソプロパノール生産量が０．９ｇ／Ｌまで増大しており、一方で３７．４ｇ／Ｌという高濃度のグリセロールを蓄積していた。
【００７７】
実施例６：過剰発現によってイソプロパノール生産性を向上させる遺伝子の選抜
過剰発現することによりイソプロパノールの生産性を向上させる遺伝子の選抜を行った。候補遺伝子が触媒する代謝経路を図１に記載する。
【００７８】
キャンディダ・ユティリス内在性のアセチルＣｏＡ加水分解酵素遺伝子（ＡＣＨ１；ヌクレオチド配列：配列番号９；アミノ酸配列：配列番号１０）、アセチルＣｏＡ合成酵素遺伝子（ＡＣＳ１；ヌクレオチド配列：配列番号１１；アミノ酸配列：配列番号１２、ＡＣＳ２；ヌクレオチド配列：配列番号１３；アミノ酸配列：配列番号１４）、細胞質局在型エタノール脱水素酵素遺伝子（ＡＤＨ２；ヌクレオチド配列：配列番号１５；アミノ酸配列：配列番号１６）、ミトコンドリア局在型エタノール脱水素酵素遺伝子（ＡＤＨ３；ヌクレオチド配列：配列番号１７；アミノ酸配列：配列番号１８）、細胞質局在型アセトアルデヒド脱水素酵素遺伝子（ＡＬＤ１；ヌクレオチド配列：配列番号１９；アミノ酸配列：配列番号２０）、ミトコンドリア局在型アセトアルデヒド脱水素酵素遺伝子（ＡＬＤ２；ヌクレオチド配列：配列番号２１；アミノ酸配列：配列番号２２）、ミトコンドリア内膜カルニチン輸送体（ＣＲＣ１；ヌクレオチド配列：配列番号２３；アミノ酸配列：配列番号２４）、ミトコンドリア局在型カルニチンアセチルCoA転移酵素（ＣＡＴ２；ヌクレオチド配列：配列番号２５；アミノ酸配列：配列番号２６）、カルニチンアセチル転移酵素遺伝子（ＹＡＴ１；ヌクレオチド配列：配列番号２７；アミノ酸配列：配列番号２８、ＹＡＴ２；ヌクレオチド配列：配列番号２９；アミノ酸配列：配列番号３０）、アセチルＣｏＡアセチル基転移酵素（ＥＲＧ１０；ヌクレオチド配列：配列番号６１；アミノ酸配列：配列番号６２）、ヤロウィア・リポリティカ由来のＡＴＰ－クエン酸除去付加酵素遺伝子（ＡＣＬ１；ヌクレオチド配列：配列番号３１；アミノ酸配列：配列番号３２と、ＡＣＬＹ；ヌクレオチド配列：配列番号３３；アミノ酸配列：配列番号３４がサブユニットを構成する遺伝子）を、それぞれのゲノムＤＮＡを鋳型としてＰＣＲにより増幅した。ＰＣＲに用いたプライマーは以下の通りである：（ＡＣＨ１；配列番号３５と配列番号３６）、（ＡＣＳ１；配列番号３７と配列番号３８）、（ＡＣＳ２；配列番号３９と配列番号４０）、（ＡＤＨ２；配列番号４１と配列番号４２）、（ＡＤＨ３；配列番号４３と配

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 586 of 3077

列番号４４）、（ＡＬＤ１；配列番号４５と配列番号４６）、（ＡＬＤ２；配列番号４７と配列番号４８）、（ＣＡＴ２；配列番号４９と配列番号５０）、（ＣＲＣ１；配列番号５１と配列番号５２）、（ＹＡＴ１；配列番号５３と配列番号５４）、（ＹＡＴ２；配列番号５５と配列番号５６）、（ＡＣＬ１；配列番号５７と配列番号５８）、（ＡＣＬＹ；配列番号５９と配列番号６０）、（ＥＲＧ１０；配列番号６３と配列番号６４）。

【００７９】

得られたＰＣＲ産物をＸｂａＩ／ＢａｍＨＩ処理して、同制限酵素で処理したｐＣＵ１５５（Keiji Kondo, , Yutaka Miura, Hidetaka Sone, Kazuo Kobayashi & Hiroshi liji ma., High-level expression of a sweet protein, monellin, in the food yeast Candida utilis., Nature Biotechnology 15, 453 - 457 (1997)）にクローニングした。ｐＣＵ１５５は、シクロヘキシミド耐性遺伝子を選抜マーカーとして有し、ＵＲＡ３座にマルチコピーインテグレーションされるベクターであり、ＸｂａＩ／ＢａｍＨＩギャップの外側にそれぞれ、グリセルアルデヒド脱水素酵素遺伝子プロモーターと同ターミネーターを有している。これまでに構築しているイソプロパノール生産株のＵＲＡ３座にはイソプロパノール合成酵素遺伝子が組み込まれているが、Ｃ．ｕｔｉｌｉｓは高次倍数体であり、ＵＲＡ３座が複数コピー存在するため、さらにＵＲＡ３座に遺伝子の導入が可能である。また、イソプロパノール合成遺伝子はＵＲＡ３座に二点組み換えにより挿入されているため、ｐＣＵ１５５の組み込みにより、導入された遺伝子の置換が起こることはない。候補遺伝子がクローン化されたプラスミドはＢｇｌＩＩで消化し、エタノール沈殿により濃縮した。得られたＤＮＡ断片を用いてＣ．ｕｔｉｌｉｓ ＴＭＳ２７２株を形質転換し、６００μｇ／ｍＬハイグロマイシン、２００μｇ／ｍＬジェネティシンおよび２０μｇ／ｍＬシクロヘキシミドを含むＹＰＤ培地に塗抹し、３０℃で４日間培養した。なお、ＡＣＬ１とＡＣＬＹのＤＮＡは混合し、同時に形質転換を行った。得られた各４クローンについて発酵試験を行い、イソプロパノール生産性を調査した。結果を図４に示す。

【００８０】

図４によると、ＡＣＳ１、ＡＣＳ２、ＣＲＣ１、ＹＡＴ１、ＥＲＧ１０、またはＡＣＬの過剰発現により、イソプロパノール生産量が１．５～４．０倍になることが明らかとなった。ＡＣＳ１、ＡＣＳ２、ＹＡＴ１、およびＡＣＬの各遺伝子はアセチルＣｏＡの合成に関わる酵素遺伝子であり、ＣＲＣ１遺伝子はアセチルＣｏＡ合成の前駆物質をミトコンドリアから細胞質に輸送するキャリアタンパク質遺伝子であり、ＥＲＧ１０遺伝子はアセチルＣｏＡからアセトアセチルＣｏＡを合成する遺伝子である。これらの結果により、酵母によるイソプロパノール高生産のためには、細胞質のアセチルＣｏＡを効率的に生産するための酵素遺伝子の発現または活性をアップレギュレートすることが重要であることが示される。

【００８１】

実施例７：ＡＣＬ１／ＡＣＬＹ、ＡＣＳ２、ＥＲＧ１０、ＣＲＣ１およびＹＡＴ１を同時に過剰発現するイソプロパノール生産株の構築と培養試験

ＡＣＬ１、ＡＣＬＹ、ＡＣＳ２、ＥＲＧ１０、ＣＲＣ１およびＹＡＴ１の各遺伝子がクローン化されたプラスミドをＢｇｌＩＩＩで消化し、エタノール沈殿により濃縮した。得られた各５μｇのＤＮＡ断片を混合し、この混合物を用いてＴＭＳ２７２株を形質転換し、６００μｇ／ｍＬハイグロマイシン、２００μｇ／ｍＬジェネティシンおよび１０μｇ／ｍＬシクロヘキシミドを含むＹＰＤ培地に塗抹し、３０℃で４日間培養した。得られた８０クローンについて発酵試験を行い、イソプロパノール生産量が最も高い株を選抜し、ＴＭＳ３７８株とした。

【００８２】

ＴＭＳ３７８株の培養試験を行った結果を図５に示す。ＴＭＳ２７２株が１．３ｇ／Ｌのイソプロパノールを生産したのに対して、ＴＭＳ３７８株は９．８ｇ／Ｌと約７．５倍のイソプロパノール生産量を示した。また、ＴＭＳ２７２株では酢酸が１８．１ｇ／Ｌ蓄積したのに対して、ＴＭＳ３７８株は４．５ｇ／Ｌと、主要な副産物である酢酸が約４分の１に低下していた。

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 587 of 3077

【００８３】

実施例８：ＴＭＳ３７８株のイソプロパノール生産における初期ＯＤ$_{600}$および振とう速度の影響

イソプロパノール生産に適したＴＭＳ３７８株の培養条件の検討を行った。ＴＭＳ３７８株を２ｍＬ ＹＰＤ液体培地（２０ｇ／Ｌペプトン、１０ｇ／Ｌ酵母エキス、２０ｇ／Ｌグルコース）／１４ｍＬ容試験管に接種し、１４０ｒｐｍ、３０℃で２４時間振とう培養した。得られた前培養液を１００ｍＬ ＹＰＤ１０Ｃ（２０ｇ／Ｌペプトン、１０ｇ／Ｌ酵母エキス、１００ｇ／Ｌグルコース、４５ｇ／Ｌ炭酸カルシウム）／３００ｍＬ容三角フラスコに接種し、初期ＯＤ$_{600}$を１もしくは１０とし、１００ｒｐｍもしくは１２０ｒｐｍ、３０℃で振とう培養した。培養７２時間後に終濃度１００ｇ／Ｌとなるようにグルコース固体を培養液に添加し、さらに１４４時間まで培養を行った。経時的にサンプリングして０．２μｍフィルターでろ過した。各種定量分析は、実施例５に記載の方法に従って行った。なお、データは３反復試行の平均で示した。結果を図６に示す。

【００８４】

図６によると、初期ＯＤ６００＝１０、１２０ｒｐｍで培養したときに最も高いイソプロパノール生産量が見られ、培養１４４時間後に最大で１８．５ｇ／Ｌのイソプロパノールが生産された。その際の対糖収率は３０．８％であった。このことから、振とう速度が速いほど（通気量が多いほど）、また、初期ＯＤが高いほど、イソプロパノール生産には好ましいことが明らかとなった。

【００８５】

実施例９：ＴＭＳ３７８株のイソプロパノール生産における添加グルコース濃度の影響

イソプロパノール生産に適したＴＭＳ３７８株の培養条件の検討を行った。ＴＭＳ３７８株を２ｍＬ ＹＰＤ液体培地（２０ｇ／Ｌペプトン、１０ｇ／Ｌ酵母エキス、２０ｇ／Ｌグルコース）／１４ｍＬ容試験管に接種し、１４０ｒｐｍ、３０℃で２４時間振とう培養した。得られた前培養液を１００ｍＬ ＹＰＤ１０Ｃ（２０ｇ／Ｌペプトン、１０ｇ／Ｌ酵母エキス、１００ｇ／Ｌグルコース、４５ｇ／Ｌ炭酸カルシウム）／３００ｍＬ容三角フラスコもしくはＹＰＤ５Ｃ（２０ｇ／Ｌペプトン、１０ｇ／Ｌ酵母エキス、５０ｇ／Ｌグルコース、４５ｇ／Ｌ炭酸カルシウム）／３００ｍＬ容三角フラスコに接種し、初期ＯＤ６００＝１、１２０ｒｐｍ、３０℃で振とう培養した。ＹＰＤ１０Ｃで培養を行ったサンプルについては、培養７２時間後に終濃度１００ｇ／Ｌとなるようにグルコース固体を添加し、さらに１９２時間まで培養を行った。また、ＹＰＤ５Ｃで培養を行ったサンプルについては、培養４８時間おきに３回終濃度５０ｇ／Ｌとなるようにグルコース固体を添加し、さらに１９２時間まで培養を行った。いずれの条件も最終のグルコース濃度は２００ｇ／Ｌである。経時的にサンプリングして０．２μｍフィルターでろ過した。各種定量分析は、実施例５に記載の方法に従って行った。なお、データは３反復試行の平均で示した。結果を図７に示す。

【００８６】

図７によると、ＹＰＤ１０Ｃで培養し、培養７２時間後に１００ｇ／Ｌのグルコースを添加した場合は、培養１２２時間後に最大で１９．２ｇ／Ｌのイソプロパノールが生産された。しかし、それ以降はイソプロパノール生産量の増大は認められず、酢酸が蓄積した（２７．２ｇ／Ｌ）。イソプロパノールの生産における最大の対糖収率は２９．０％であった。一方、ＹＰＤ５Ｃで培養し、４８時間おきに５０ｇ／Ｌのグルコースを３回添加した場合は、培養１９２時間までイソプロパノールの増大が認められ、培養１９２時間後に最大で２７．４ｇ／Ｌのイソプロパノールが生産された。また、酢酸の蓄積はほとんど認められなかった。この条件でのイソプロパノールの生産における最大の対糖収率は４１．５％であった。このことから、添加するグルコースの濃度を比較的低濃度に保つことはイソプロパノール生産において好ましいことが明らかとなった。

【００８７】

実施例１０：イソプロパノール生産量向上に寄与の大きい遺伝子の特定

同時に過剰発現することによってイソプロパノール生産量を向上させる寄与の大きい遺

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 588 of 3077

伝子を特定するために、(i)ＡＣＬ１、ＡＣＬＹおよびＡＣＳ２の組み合わせを過剰発現する株、(ii)ＡＣＬ１、ＡＣＬＹおよびＥＲＧ１０の組み合わせを過剰発現する株、(iii)ＡＣＳ２およびＥＲＧ１０の組み合わせを過剰発現する株、(iv)ＡＣＬ１、ＡＣＬＹ、ＡＣＳ２およびＥＲＧ１０の組み合わせを過剰発現する株をそれぞれ構築し、ＡＣＬ１、ＡＣＬＹ、ＡＣＳ２、ＣＲＣ１、ＥＲＧ１０およびＹＡＴ１の組み合わせを過剰発現する株（ＴＭＳ３７８株）およびそれぞれの単独遺伝子の過剰発現株との間で、イソプロパノール生産量の比較を行った。組み換え株の構築は実施例６および実施例７と同様の方法を用いて行い、組み込みたい遺伝子ＤＮＡを混合し、同時に形質転換を行った。得られた各３２クローンについて発酵試験を行い、それぞれの導入遺伝子の組み合わせにおいて平均的なイソプロパノール生産量を示す株を選抜した。各構築株を２ｍＬ ＹＰＤ液体培地（２０ｇ／Ｌペプトン、１０ｇ／Ｌ酵母エキス、２０ｇ／Ｌグルコース）／１４ｍＬ容試験管に接種し、１４０ｒｐｍ、３０℃で２４時間振とう培養した。得られた前培養液を３ｍＬ ＹＰＤ１０Ｃ（２０ｇ／Ｌペプトン、１０ｇ／Ｌ酵母エキス、１００ｇ／Ｌグルコース、４５ｇ／Ｌ炭酸カルシウム）／ガラス試験管に接種し、初期ＯＤ６００＝１、１００ｒｐｍ、３０℃で振とう培養した。経時的にサンプリングを行い、高速液体クロマトフラフィーおよびガスクロマトフラフィーＧＣ－２０１０でイソプロパノール生産量を評価した。なお、データは３反復試行の平均で示した。

【００８８】

図８によると、ＡＣＳ１、ＡＣＳ２およびＥＲＧ１０のそれぞれ単独の過剰発現により、イソプロパノール生産量が２．２～２．７倍になることが明らかとなった。(i)ＡＣＬ１、ＡＣＬＹおよびＡＣＳ２の組み合わせを過剰発現する株、ならびに(ii)ＡＣＬ１、ＡＣＬＹおよびＥＲＧ１０の組み合わせを過剰発現する株のイソプロパノール生産量は、それぞれＡＣＳ２およびＥＲＧ１０を単独で過剰発現する株と有意な差は認められなかった。一方、(iii)ＡＣＳ２およびＥＲＧ１０の組み合わせを過剰発現する株ではイソプロパノール生産の飛躍的な向上が認められ、生産量が１１．６倍に向上した。また、(iii)ＡＣＳ２およびＥＲＧ１０の組み合わせを過剰発現する株、(iv)ＡＣＬ１、ＡＣＬＹ、ＡＣＳ２およびＥＲＧ１０の組み合わせを過剰発現する株、ＴＭＳ３７８株（ＡＣＬ１、ＡＣＬＹ、ＡＣＳ２、ＣＲＣ１、ＥＲＧ１０およびＹＡＴ１の組み合わせを過剰発現する株）の株間において、イソプロパノール生産量に有意な差は認められなかった。

【００８９】

これらの結果より、同時に過剰発現した際にはＡＣＬ、ＣＲＣ１およびＹＡＴ１の過剰発現がイソプロパノールの生産性向上に与える寄与は小さく、イソプロパノール生産性を向上させるためにはＡＣＳ２とＥＲＧ１０を同時に過剰発現することが最も重要であることが示された。以上の結果は、酵母によるイソプロパノール高生産のためにはエタノールからアセチルＣｏＡを合成する代謝経路、すなわち酢酸からアセチルＣｏＡを合成する代謝経路に関わる酵素遺伝子の発現または活性を増強することが重要であることを示している。

【図面の簡単な説明】

【００９０】

【図１】図１は、酵母によるイソプロパノール合成経路を示す図である。

【図２】図２は、イソプロパノール生産遺伝子を導入した株の非中和条件での発酵試験の結果を示す図であり、培養液中のイソプロパノール濃度（パネルＡ）、グルコース濃度（パネルＢ）、エタノール濃度（パネルＣ）、酢酸濃度（パネルＤ）、グリセロール濃度（パネルＥ）およびＯＤ６００（パネルＦ）の経時変化を示している。各パネルにおいて、黒丸はＡＴＣＣ１５２３９株の結果を示し、白丸はＴＭＳ２７２株（ＡＴＣＣ１５２３９株＋ＩＰＡ合成遺伝子）の結果を示し、黒三角はＰＤＣ１破壊株の結果を示し、白三角はＴＭＳ２９０株（ＰＤＣ破壊株＋ＩＰＡ合成遺伝子）の結果を示す。

【図３】図３は、イソプロパノール生産遺伝子を導入した株の中和条件での発酵試験の結果を示す図であり、培養液中のイソプロパノール濃度（パネルＡ）、グルコース濃度（パネルＢ）、エタノール濃度（パネルＣ）、酢酸濃度（パネルＤ）およびグリセロール濃度

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 589 of 3077

（パネルＥ）の経時変化を示している。各パネルにおいて、黒丸はＡＴＣＣ１５２３９株の結果を示し、白丸はＴＭＳ２７２株（ＡＴＣＣ１５２３９株＋ＩＰＡ合成遺伝子）の結果を示し、黒三角はＰＤＣ１破壊株の結果を示し、白三角はＴＭＳ２９０株（ＰＤＣ破壊株＋ＩＰＡ合成遺伝子）の結果を示す。

【図４】図４は、ＴＭＳ２７２株において、各種遺伝子を過剰発現させた場合におけるイソプロパノール生産量を、ＴＭＳ２７２株によるイソプロパノール生産量に対する相対量として示したグラフである。

【図５】図５は、イソプロパノール生酸性を向上させる遺伝子を全て過剰発現する株の発酵試験の結果を示す図であり、培養液中のイソプロパノール濃度（パネルＡ）、グルコース濃度（パネルＢ）、エタノール濃度（パネルＣ）、酢酸濃度（パネルＤ）、グリセロール濃度（パネルＥ）およびアセトン濃度（パネルＦ）の経時変化を示している。各パネルにおいて、黒丸はＴＭＳ２７２株の結果を示し、白丸はＴＭＳ３７８株（ＴＭＳ２７２株＋ＡＣＬ１：ＡＣＬＹ：ＡＣＳ２：ＥＲＧ１０：ＣＲＣ１：ＹＡＴ１）の結果を示す。

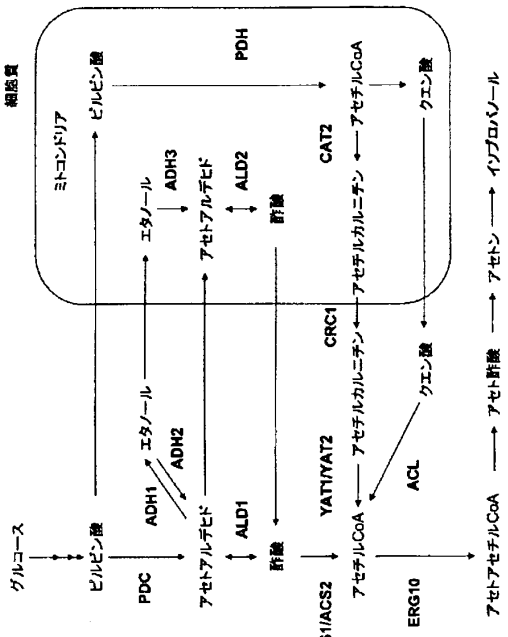
【図６】図６は、イソプロパノール生産性を向上させる遺伝子を全て過剰発現するＴＭＳ３７８株の３０℃での発酵試験（初期$ＯＤ_{６００}$および振とう速度の影響）の結果を示す図であり、培養液中のイソプロパノール濃度（パネルＡ）、グルコース濃度（パネルＢ）、エタノール濃度（パネルＣ）、酢酸濃度（パネルＤ）、グリセロール濃度（パネルＥ）およびアセトン濃度（パネルＦ）の経時変化を示している。各パネルにおいて、黒丸は初期ＯＤ６００＝１かつ撹拌数＝１００ｒｐｍの条件下での結果を示し、黒三角は初期ＯＤ６００＝１かつ撹拌数＝１２０ｒｐｍの条件下での結果を示し、白丸は初期ＯＤ６００＝１０かつ撹拌数＝１００ｒｐｍの条件下での結果を示し、白三角は初期ＯＤ６００＝１０かつ撹拌数＝１２０ｒｐｍの条件下での結果を示す。

【図７】図７は、イソプロパノール生産性を向上させる遺伝子を全て過剰発現するＴＭＳ３７８株の３０℃での発酵試験（流加グルコース濃度の影響）の結果を示す図である。パネルＡは、初期グルコース濃度を１００ｇ／Ｌとし、培養７２時間後に１００ｇ／ｌのグルコースを流加して、１９２時間培養を行った結果を示す。パネルＢは、初期グルコース濃度を５０ｇ／Ｌとし、培養４８時間後、９６時間後および１４４時間後にそれぞれ５０ｇ／ｌのグルコースを流加して、１９２時間培養を行った結果を示す。各パネルにおいて、イソプロパノール濃度（黒三角）、グルコース濃度（黒丸）、エタノール濃度（白丸）および酢酸濃度（白三角）の経時変化を示している。

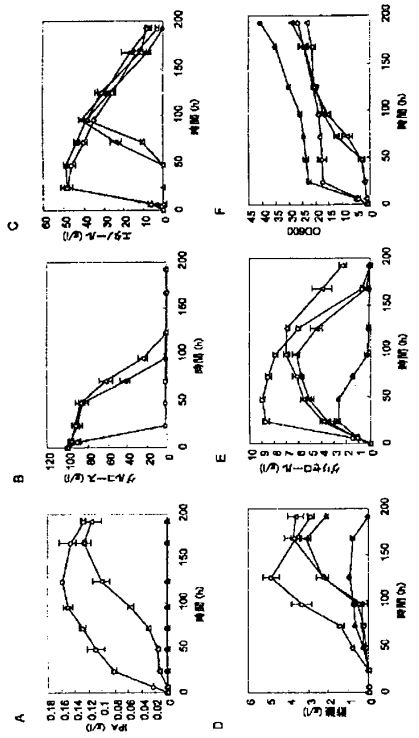
【図８】図８は、ＴＭＳ２７２株において、各種遺伝子を過剰発現させた場合におけるイソプロパノール生産量を示したグラフである。

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 590 of 3077

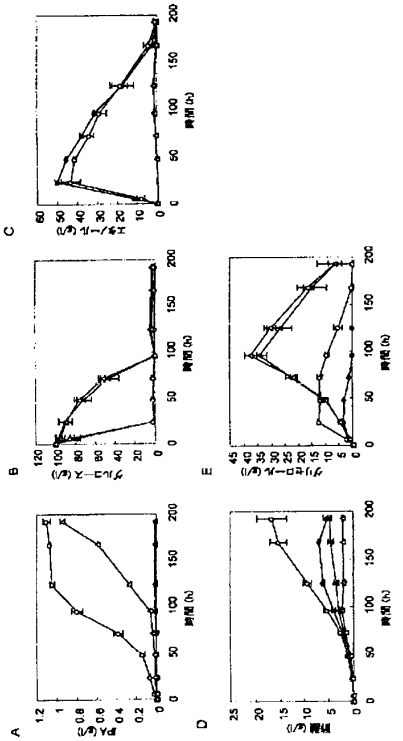
【図１】


細胞質
ミトコンドリア
グルコース
ピルビン酸
PDC
アセトアルデヒド
ADH1
ADH2
エタノール
ALD1
酢酸
ACS1/ACS2
アセチルCoA
YAT1/YAT2
アセチルカルニチン
CRC1
CAT2
PDH
ERG10
ACL
クエン酸
アセトアセチルCoA
CTF
アセト酢酸
ADC
アセトン
SADH
イソプロパノール

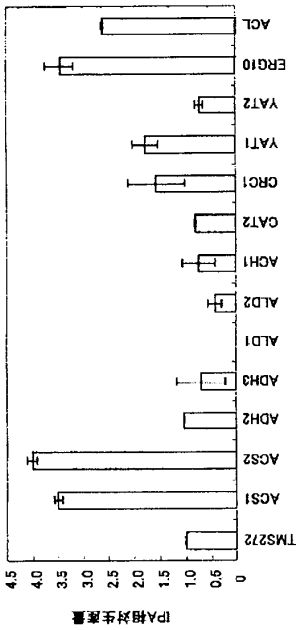
【図２】



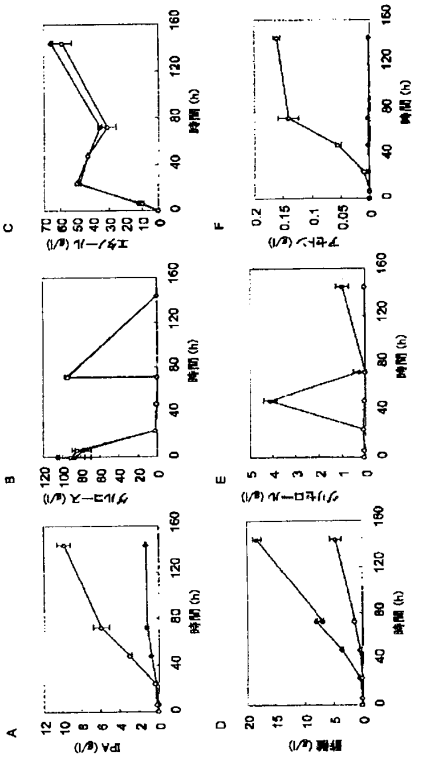
【図３】



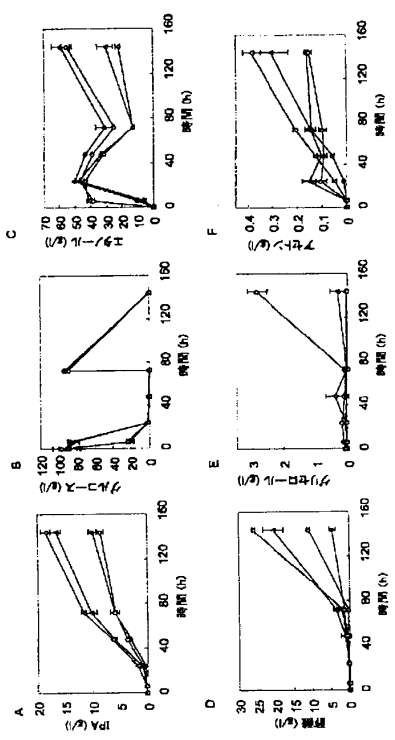
【図４】


TMS272
ACS1
ACS2
ADH2
ADH3
ALD1
ALD2
ACH1
CAT2
CRC1
YAT1
YAT2
ERG10
ACL
IPA相対生産量

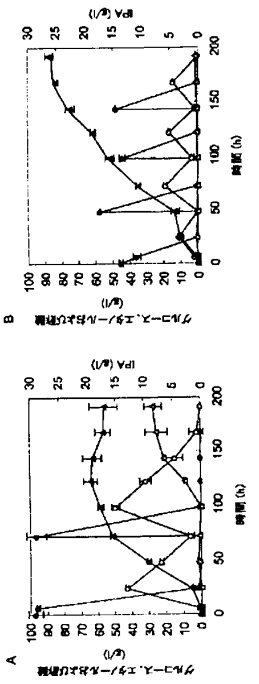
【図５】



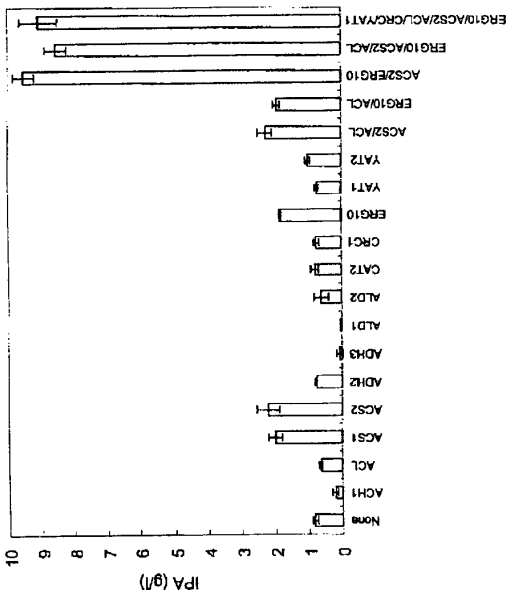
【図６】



【図７】



【図８】


IPA (g/l)
None
ACH1
ACL
ACS1
ACS2
ADH2
ADH3
ALD1
ALD2
CAT2
CRC1
ERG10
YAT1
YAT2
ACS2/ACL
ERG10/ACL
ACS2/ERG10
ERG10/ACS2/ACL
ERG10/ACS2/ACL/CRC/YAT1

【配列表】
2012254044000001.app

フロントページの続き

(72)発明者 生 嶋 茂 仁
東京都中央区新川二丁目１０番１号 キリンホールディングス株式会社内

Ｆターム(参考) 4B024 AA17 BA80 CA04 DA12 GA11
4B050 CC03 DD02 LL05
4B064 AC02 CA06 CA19 CC24 DA16
4B065 AA23Y AA73X AB01 AC14 BA02 CA05 CA54

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 594 of 3077

METHOD FOR PRODUCING SUBSTANCE USING METABOLIC PATHWAY GOING THROUGH ACETYL COA IN YEAST

ABSTRACT

PROBLEM TO BE SOLVED: To provide a yeast strain capable of highly producing metabolic products going through acetyl-CoA.SOLUTION: The yeast strain in which activity of the following enzymes: (a) an acetyl-CoA synthetase (ACS); and (b) an acetyl-CoA acetyltransferase (ERG10) are enhanced.

Pub. No.	JP2012254044A
App. No.	JP20110129527
App. No. Original	2011129527
Appl. Date	09-Jun-2011
Pub. Date	27-Dec-2012
Inventor(s)	TAMAGAWA HIDEYUKI IKUSHIMA SHIGEHITO
Assignee Norm.	KIRIN HOLDING CO LTD
Current Owner	MITSUBISHI GROUP
IPC	C12N1/19 , C12N9/10 , C12N15/09 , C12P7/04
CPC	Y02P20/52

CLAIMS

1. The following enzymes: (A) acetyl-CoA synthase (ACS); (b) and acetyl-CoA acetyltransferase (ERG10) Activity is enhanced, yeast strain.

2. The following genes: Gene encodes a polypeptide having an activity (a) acetyl-CoA synthase (ACS); expression of a gene encoding a polypeptide having an activity (b) acetyl-CoA acetyltransferase in (ERG10) and is enhanced The yeast strain according to any yl, in claim 1.

3. Polypeptide with ACS activity, The amino acid sequence represented by SEQ ID NO: 14 or (ii) SEQ ID NO: 12 or a polypeptide comprising the amino acid sequence represented by SEQ ID NO: 14, or (i) SEQ ID NO: 12, one or several amino acids are deleted, is a polypeptide having an ACS activity and comprises an amino acid sequence that has been inserted, a substituted or added, the yeast strain according to claim 2.

4. Is a polypeptide having ERG10 activity, Amino acids in the amino acid sequence represented by (ii) SEQ ID NO: 62 or a polypeptide comprising the amino acid sequence represented by (i) SEQ ID NO: 62, one or more amino acids is inserted deleted, substituted, added, or is a polypeptide having an ERG10 activity and comprises the sequence, the yeast strain according to claim 2 or 3.

5. In addition, the following genes: The gene encodes a polypeptide having an activity (c) ATP- citric acid removal enzyme addition (ACL); Selected from the group consisting of a gene encoding a polypeptide having an activity (e) carnitine acetyl transferase gene (YAT1) and; gene encoding a polypeptide having an activity (d) the inner mitochondrial membrane carnitine transporter (CRC1) Expression of one gene is increased at least a yeast strain as claimed in any one of claims 1 to 4.

6. It is intended a polypeptide having ACL activity, a polypeptide having ACLY activity and a polypeptide having the activity ACL1, Is a polypeptide having ACL1 activity, Amino acids in the amino acid sequence represented by (ii) SEQ ID NO: 32 or a polypeptide comprising the amino acid sequence represented by (i) SEQ ID NO: 32, one or more amino acids is inserted deleted, substituted, added, or It is a polypeptide having an activity ACL1 and comprises a sequence, Is a polypeptide having ACLY activity, Amino acids in the amino acid sequence represented by (ii) SEQ ID NO: 34 or a polypeptide

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 595 of 3077

comprising the amino acid sequence represented by (i) SEQ ID NO: 34, one or more amino acids is inserted deleted, substituted, added, or is a polypeptide having an ACLY activity and comprises the sequence, the yeast strain according to claim 5.

7. Is a polypeptide having CRC1 activity, Amino acids in the amino acid sequence represented by (ii) SEQ ID NO: 24 or a polypeptide comprising the amino acid sequence represented by (i) SEQ ID NO: 24, one or more amino acids is inserted deleted, substituted, added, or is a polypeptide having a CRC1 activity and comprises the sequence, the yeast strain according to claim 5 or 6.

8. Is a polypeptide having YAT1 activity, Amino acids in the amino acid sequence represented by (ii) SEQ ID NO: 28 or a polypeptide comprising the amino acid sequence represented by (i) SEQ ID NO: 28, one or more amino acids is inserted deleted, substituted, added, or is a polypeptide having a YAT1 activity and comprises the sequence, the yeast strain according to any one of claims 5 to 7.

9. A Crabtree negative yeast effect, the yeast strain according to any of claims 1 to 8.

10. Is Candida utilis, and a yeast strain as claimed in any one of claims 1 to 9.

11. Is one which has been transformed with genes for the synthesis of metabolites via the acetyl CoA, the yeast strain according to any one of claims 1 to 10.

12. 4 that encodes a polypeptide having the activity secondary alcohol dehydrogenase CoA transferase subunit A (ctfA), CoA transferase subunit B (ctfB), acetoacetic acid decarboxylase and (adc) of (sadh) The yeast strains according is to be transformed with the genes of the species to claim 11.

13. The method comprising the step of culturing the yeast strain according to any one of claims 1 to 12, to produce a material.

14. Is a metabolite through the acetyl CoA, the material A method according to claim 13.

15. Is isopropanol said substance, the method according to claim 13 or 14.

DESCRIPTION

Background of the invention

In yeast, TECHNICAL FIELD The present invention relates to a process for the production of a substance using the metabolic pathway via acetyl-CoA.

Against the background of environmental issues to the background art serious, biofuels have attracted attention as an alternative fuel for petroleum. Ethanol is well known as biofuels, it only has lower energy density than gasoline, it is necessary to be mixed with gasoline in a concentration range that is limited in order to function as a transportation fuel, as a fuel shortcomings in many cases. Also, having a higher octane number than the 1-butanol, isopropanol, a solvent having high versatility is also used as industrial raw materials. Isopropanol, which are fabricated by the hydration of propylene from naphtha, never been produced commercially from renewable sources.

Fermentative production of butanol and acetone by microorganisms, has been carried out industrially in the early 20th century. Can be produced at the same time the butanol and acetone Clostridium bacteria of some sort of strictly anaerobic bacteria (acetone-butanol fermentation). Acetone is produced as a raw material acetyl CoA present in the bacterial cell. Is converted to acetoacetyl-CoA by (thiolase) acetyl-CoA acetyltransferase, and are converted to acetoacetate by CoA transferase addition, acetyl-CoA is converted to acetone by acetoacetic acid decarboxylase enzyme. Clostridium certain bacteria has a dehydrogenase of secondary alcohols further, I want to produce isopropanol from acetone. Have reported tha

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 596 of 3077

in Clostridium bacteria 52 strains, including Clostridium Beijerinki, strains producing isopropanol 30mM of (1.8g / L) at the maximum exists Chen et al (1: Chen, J.-S., and SF Hiu Acetone-butanol-isopropanol production by Clostridium beijerinckii (synonym, Clostridium butylicum) Biotechnol Lett 1986. 8:.... 371-376). We have also introduced CoA transferase of Clostridium genus, acetoacetic acid decarboxylase, a secondary alcohol dehydrogenase in E. coli, Hanai et al. And Jojima colleagues, excess acetyl-CoA acetyltransferase of endogenous recently ... T Hanai, S. Atsumi, and JC Liao Engineered Synthetic Pathway for Isopropanol Production in Escherichia coli Appl Environ Microbiol,: 2 (non-patent literature have reported E. coli strain that can efficiently produce isopropanol by expressing ... Toru Jojima & Masayuki Inui & Hideaki Yukawa Production of isopropanol by metabolically engineered Escherichia coli Appl Microbiol Biotechnol 2008 77::; 2007, p 7814-7818 Dec. 3 non-patent literature 1219-1224). However, E. coli, various stress-resistant glucose tolerance, osmotic tolerance, salt tolerance, and organic solvent resistance is not very high in many cases, there is also a risk of phage contamination. Also, isopropanol resistance is low means that the inhibition of fermentation by product occurs. How to continue fermentation while removing from the medium the product by using, for example gas stripping process in order to avoid fermentation inhibitors are known, the use of the method not only to complication of the process of fermentation , leading to a cost increase. It can be considered as one means for solving these problems, and producing isopropanol using the microorganism having a high organic solvent resistance.

Yeast are unicellular eukaryotic microorganisms also showed a high organic solvent resistance, as well as the microorganisms such as E. coli and that can be cultured at low costs than the cells of higher organisms in terms of culture cost culture equipment, and culture time I have a feature. In particular, budding yeast is used in the production of bread and alcoholic beverages such as beer and wine, as the safety of the food has been confirmed.

In order to produce isopropanol in yeast, the yeast strain that is required to CoA transferases exogenous acetoacetate decarboxylase, and introducing a secondary alcohol dehydrogenase. Until now, the example of producing isopropanol by yeast, it was produced isopropanol by introducing the gene of the isopropyl alcohol-producing enzymes Saccharomyces cerevisiae to (Saccharomyces cerevisiae), and culturing the transformant has been reported (: Patent Publication No. 2011-97929 Patent Document 1). However, production of isopropanol by this method, only a small quantity of 100mg / L before and after the concentration in the culture medium. Yeast is a eukaryote, from that organelle is present, the supply route is more present in the cytoplasm acetyl-CoA is a substrate of isopropanol synthesis. There is no example of what yeast or is excellent in such capacity can be provided to the cytoplasm acetyl-CoA from which path or whether advantageous for isopropanol synthesis was investigated to date.

Candida utilis (Candida utilis), as well Saccharomyces cerevisiae and (Saccharomyces cerevisiae) Saccharomyces fragilis and (Saccharomyces fragilis), and has been certified as edible yeast FDA by (Federal Food and Drug Administration), as a food additive It is an edible yeast can be used safely permitted. In fact, even today, including the Japanese, it is produced in countries around the world such as the United States and Germany, Candida utilis, has been used as a food diet. In additior Candida utilis, has been used in a wide range of industries as producing strain bacterial production by deer pentose, ethyl acetate, L- glutamic acid, glutathione, and invertase. Hiroshi Shimada, Keiji Kondo, Paul D. Fraser, Yutaka: 4 (non-patent literature in recent years, the transformation method has been developed in Candida utilis, technologies various compounds are produced metabolically engineering has been developed .... Miura, Toshiko Saito, and Norihiko Misawa, Increased Carotenoid Production by the Food Yeast Candida utilis through Metabolic Engineering of the Isoprenoid Pathway, Appl Environ Microbiol, July 1998, p 2676-2680, Vol 64, 7 No.; non- Patent Document 5:. Yutaka Miura, Keiji Kondo, Toshiko Saito, Hiroshi Shimada, Paul D. Fraser, and Norihiko MisawaProduction of the Carotenoids Lycopene, β-Carotene, and Astaxanthin in the Food Yeast Candida utilis, Appl Environ Microbiol, April 1998, .. p 1226-1229, Vol 64, No. 4; 6 Non-Paten Document: Keiji Kondo,, Yutaka Miura, Hidetaka Sone, Kazuo Kobayashi & Hiroshi Iijima, High-level expression of a sweet protein, monellin, in the food yeast Candida utilis, 15, 453 Nature Biotechnology - 457 (1997); 7 Non-Patent Document:.. Yutaka Miura, Keiji Kondo, Hiroshi Shimada, Toshiko .. Saito, Katsumi Nakamura, Norihiko Misawa, Production of lycopene by the food yeast, Candida utilis that does not naturally synthesize carotenoid, Biotechnology and Bioengineering Volume 5(Issue 2-3, Pages 306-308; 8 Non-Patent Document: Shigehito IKUSHIMA, Toshio FUJII, Osamu KOBAYASHI, Satoshi

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 597 of 3077

YOSHIDA and Aruto YOSHIDA., Genetic Engineering of Candida utilis Yeast for Efficient Production of L-Lactic Acid., Bioscience, Biotechnology, and Biochemistry Vol. 73 (2009), No. 8 pp.1818-1824 ; Non-Patent Document 9:... Yi-Ren Hong, Ya-Lei Chen, Lynn Farh, Wen-Jen Yang, Chen-Hua Liao and David Shiuan, Recombinant Candida utilis for the production o biotin, Applied Microbiology and Biotechnology, Volume 71 (2006), Number 2, 211-221). In addition, Candida utilis, unlike Saccharomyces genus, from that Crabtree is (Crabtree) effect negative yeast, respiratory activity is not reduced even in the presence of glucose. Thus, by adopting the conditions for ventilating enough, it is possible to suppress generation of ethano Moreover, respiratory activity high Candida utilis, since the ethanol produced is also once again rapidly assimilated, thereby enabling the production of metabolites and cell efficiency.

When producing materials such as isopropanol in yeast, generally, generation of ethanol as a by-product becomes a problem. Patent document; No. 2010/095751 WO 2 (patent document in order to suppress the ethanol by-product, efforts to deficient (PDC) pyruvate decarboxylase gene is a key enzyme in the ethanol production have been made 3:..; 2010/095750 No. WO 10 non-patent literature: Flikweert MT, Van Der Zanden L, Janssen WM, et al Pyruvate decarboxylase: an indispensable enzyme for growth of Saccharomyces cerevisiae on glucose Yeast 12 (1996): ..; 247-257 Non-Patent Document 11: van Maris AJA, Winkler AA, Porro D, et al Homofermentative lactate production can not sustain anaerobic growth of engineered Saccharomyces cerevisiae: Possible consequence of energy-dependent lactate export Appl Environ Microbiol 70 (2004) : 2898-2905).

Chen, J.-S., and SF Hiu Acetone-butanol-isopropanol production by Clostridium beijerinckii (synonym, Clostridium butylicum Biotechnol Lett 1986. 8:.... 371-376 T. Hanai, S. Atsumi, and JC Liao. Engineered Synthetic Pathway for Isopropanol Production in Escherichia coli. Appl Environ Microbiol, Dec. 2007, p. 7814-7818 .. Toru Jojima & Masayuki Inui & Hideaki Yukawa Production of isopropanol by metabolically engineered Escherichia coli Appl Microbiol Biotechnol 2008 77: 1219-1224 Hiroshi Shimada, Keiji Kondo, Paul D. Fraser, Yutaka Miura, Toshiko Saito, and Norihiko Misawa., Increased Carotenoid Production by the Food Yeast Candida utilis through Metabolic Engineering of the Isoprenoid Pathway., Appl Environ Microbiol, July 1998, p. 2676 -2680, Vol. 64, No. 7 Yutaka Miura, Keiji Kondo, Toshiko Saito, Hiroshi Shimada, Paul D. Fraser, and Norihiko MisawaProduction of the Carotenoids Lycopene, β-Carotene, and Astaxanthin in the Food Yeast Candida utilis., Appl Environ Microbiol, April 1998, p. 1226- 1229, Vol. 64, No. 4 .. Keiji Kondo ,, Yutaka Miura, Hidetaka Sone, Kazuo Kobayashi & Hiroshi Iijima, High-level expression of a sweet protein, monellin, in the food yeast Candida utilis, Nature Biotechnology 15, 453 - 457 (1997) Yutaka Miura, Keiji Kondo, Hiroshi Shimada, Toshiko Saito, Katsumi Nakamura, Norihiko Misawa., Production of lycopene by the food yeast, Candida utilis that does not naturally synthesize carotenoid., Biotechnology and Bioengineering Volume 58 Issue 2-3, Pages 306- 308 Shigehito IKUSHIMA, Toshio FUJII, Osamu KOBAYASHI, Satoshi YOSHIDA and Aruto YOSHIDA., Genetic Engineering of Candida utilis Yeast for Efficient Production c L-Lactic Acid., Bioscience, Biotechnology, and Biochemistry Vol. 73 (2009), No. 8 pp. 1818-1824 Yi-Ren Hong, Ya-Lei Chen Lynn Farh, Wen-Jen Yang, Chen-Hua Liao and David Shiuan., Recombinant Candida utilis for the production of biotin., Applied Microbiology and Biotechnology., Volume 71 (2006), Number 2, 211-221 Flikweert MT, Van Der Zanden L, Janssen WM, et al Pyruvate decarboxylase:.. An indispensable enzyme for growth of Saccharomyces cerevisiae on glucose Yeast 1: (1996): 247-257 van Maris AJA, Winkler AA, Porro D, et al Homofermentative lactate production can not sustain anaerobic growth of engineered Saccharomyces cerevisiae:.. Possible consequence of energy-dependent lactate export Appl Environ Microbiol 70 (2004): 2898-2905 Japanese Patent Publication No. 2011-97929 No. 2010/095751 WO No. 2010/095750 WO

By enhancing the activity acetyl CoA acetyl transferase enzyme acetyl-CoA synthase in yeast (ACS) and of (ERG10), the present inventors have found that, to a significant increase in the production volume of the substance through the acetyl-Co, in yeast I have found that. The present invention is based on this finding.

Thus, yeast strains enzymatic activity according to the synthesis and metabolism of acetoacetyl-CoA and acetyl-CoA are involved as intermediates in the fermentation production of various substances is enhanced, an object of the present invention, the materials, as well as with yeast strain It is another object to provide a manufacturing process.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 598 of 3077

The yeast strain according to the present invention, the following enzymes:
(A) acetyl-CoA synthase (ACS); (b) and acetyl-CoA acetyltransferase (ERG10)
It is a yeast strain activity is enhanced.

Furthermore, the production method according to the invention is a method comprising the step of culturing a yeast strain according to the present invention, to produce the material.

In accordance with the present invention, the metabolic pathway of yeast acetoacetyl CoA and acetyl-CoA is involved is activated. Because of the wide variety of metabolites to generate also there is a wide variety of metabolic pathways which acetoacetyl-CoA and acetyl-CoA. Therefore, according to the present invention, it is possible to manufacture with high efficiency various substances by culturing the yeast strain. Also, production of the material by depending on the material to be produced, is incorporated into the yeast in an expressible form genes required enzyme, and culturing the yeast strain obtained is also possible.

DETAILED DESCRIPTION OF THE INVENTION

The yeast strain according to the invention, the following genes: a gene encoding a polypeptide having an activity (a) acetyl-CoA synthase (ACS); poly having an activity (b) acetyl-CoA acetyltransferase in (ERG10) and Expression of the gene encoding the peptides is enhanced.

As used herein, acetyl-CoA synthase and the above (ACS) gene acetyl-CoA acetyltransferase with (ERG10) gene, to prepare a yeast strain overexpressing the four genes for the synthesis of isopropanol, culturing this By the, the high concentration of isopropanol are produced very has been demonstrated. Isopropanol production pathway in the yeast strain is as shown in Figure 1. From path shown in Figure 1, a high production volume isopropanol, due to the enhancement of enzyme activity relating to synthesis and metabolism of acetoacetyl-CoA and acetyl-CoA is clear. Therefore, the present invention allows to produce a high various materials not only isopropanol, via acetoacetyl-CoA and acetyl-CoA. Also, as shown in Figure 1, ERG10 and ACS are enzymes that function in the cytoplasm.

Acetyl-CoA supplied to the cytoplasm, a core material compound having a flavonoid skeleton, compounds having a terpene terpenoid skeleton, a compound having a sterol backbone, such as fatty acids, secondary metabolites for various utilities. Thus, as long as the synthetic route of the secondary metabolites are in a state of being genetically modified artificially or native state yeast strains, it can be highly productive of these substances by the method of the present invention .

When in general, to perform the microbial production of substances, when suppressing effort is made a by-product production in many cases. For example, when producing the substance in yeast, In general, in order to suppress the formation of ethanol as a by-product, the enzyme genes involved in ethanol production, such efforts to be deficient in pyruvate decarboxylase gene (PDC) have been made .. 2010/095750 No. WO; No. 2010/095751 pamphlet (International Publication Flikweert MT, Van Der Zanden L, Janssen WM, et al Pyruvate decarboxylase: an indispensable enzyme for growth of Saccharomyces cerevisiae on glucose Yeast 12 (1996):.. 247-257; van Maris AJA, Winkler AA, Porro D, et al Homofermentative lactate production can not sustain anaerobic growth of engineered Saccharomyces cerevisiae: Possible consequence of energy-dependent lactate export Appl Environ Microbiol 70 (2004): 2898 -2905). However, herein, than the veast strain was deficient in the PDC described above. the direction of a veast strain did not lacking PDC indicates isopropanol production significantly higher have been demonstrated. Thus, the yeast strain according to the invention is a

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 599 of 3077

feature also is maintained without being deleted gene involved in ethanol production, such as PDC, isopropanol production, or its acetyl CoA in turn a result that the production volume of acetoacetyl-CoA is a metabolite can be improved remarkably, is surprising.

In this specification the enzyme activity being "increased" means that at one time at least in the course of culturing, the stronger the enzyme activity is observed compared with yeast which is not subjected to means for enhanced activity to. Enhancement of enzyme activity, can be implemented by a replacement of a higher enzyme specific activity of the gene, by modification of the enzyme gene by way engineering or protein that enhances expression of the enzyme gene. As used herein the expression of the gene being "enhanced", the expression level of that gene in yeast is increased by genetic modification of the yeast, that is, it is meant that overexpression of the gene. The method of genetic modification include, for example, (that is, a method for transfecting yeast in a form capable of expressing the gene) Methods for transforming yeast with the gene, gene present in the genome of the yeast method for enhancing the promoter activity, and a method of enhancing expression of the gene by replacing the promoter with other promoters of genes present in the genome of the yeast, and various methods are known, but preferably It is a method to transform yeast by the gene.

The yeast strain according to the invention the yeast is a yeast which has been genetically altered genes specific to overexpress. Well, yeast used for genetic modification is not particularly limited as long as it is a yeast having a metabolic pathway through the acetoacetyl-CoA or acetyl-CoA, the clubs and more preferably, is a Crabtree effect negative yeast with high respiratory activity, preferably it is a Candida utilis is a Tree negative yeast. Various strains of well-known, for example, NBRC0396 shares, NBRC0619 shares, NBRC0988 shares, NBRC0639 shares, NBRC0626 shares, NBRC1086 shares, NBRC10707 shares, ATCC9255 shares, ATCC28955 strain, ATCC44638 strains, strains of Candida utilis ATCC9206 in the art strain, ATCC96621 strain, ATCC18201 shares, and may be a ATCC15239 shares, etc., but is preferably NBRC1086 share, and ATCC15239 shares or ATCC9255 shares.

The following enzymes enzyme activity is enhanced in yeast strain caused by the gene present invention and the enzyme activity is enhanced,: (a) acetyl-CoA synthase (ACS); (b) and acetyl-CoA acetyltransferase (ERG10 is). In accordance with a preferred embodiment of the present invention, genes of these two enzymes are over-expressed in a yeast strain according to the present invention, that is, the following genes: poly have activity (a) acetyl-CoA synthase (ACS) gene encoding the peptide; expression of a gene encoding a polypeptide having an activity (b) acetyl-CoA acetyltransferase in (ERG10) and has enhanced.

The yeast strain according to the present invention, further,: a gene encoding a polypeptide having the activity (c) ATP- citric acid removal additional enzymes (ACL); (d) the inner mitochondrial membrane carnitine transporter and of (CRC1) Expression of one gene may have been increased at least is selected from the group consisting of a gene encoding a polypeptide having an activity (e) carnitine acetyl transferase gene (YAT1) and; gene encoding a polypeptide having an activity may be, are (c), (d) and the expression of all genes (e) is preferably increased.

Those derived from a variety of yeast are known genes and these enzymes, its origin is not particularly limited. Or may be the expression product and the endogenous genes of the yeast used in breeding, gene and enzyme may be a unit derived from yeast other than yeast used for breeding. For example, genes and enzymes these may be the one independently from Yarrowia lipolytica or Candida utilis, respectively. As to the origin of each enzyme, for example, acetyl-CoA synthase (ACS) Candida utilis yeast, acetyl-CoA acetyl transferase enzyme (ERG10) yeast Candida utilis, mitochondrial inner membrane

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 600 of 3077

carnitine transporter (CRC1) is ACLY and ACL1 of (Yarrowia lipolytica derived constitute a subunit Yarrowia lipolytica yeast (ACL) the yeast Candida utilis, ATP- citric acid removal adding enzyme Candida utilis yeast, carnitine acetyltransferase gen (YAT1) gene) that is preferred.

Acetyl-CoA synthase derived from Candida utilis to (ACS) gene, there is a gene ACS2 and ACS1 gene. Coding sequence o ACS1 gene is represented by SEQ ID NO: 11, the amino acid sequence encoded is shown in SEQ ID NO: 12. Coding sequence of ACS2 gene is represented by SEQ ID NO: 13, the amino acid sequence encoded is shown in SEQ ID NO: 14. Acetyl-CoA synthetase (ACS) is an enzyme that converts acetyl-CoA and acetate. May be any of ACS2 and ACS1 The ACS gene used in the present invention, but I are preferably ACS2. In addition, a polypeptide having the activity acetyl-CoA synthase (ACS) is an amino acid in the amino acid sequence represented by SEQ ID NO: 14 or SEQ ID NO: 12, one or more amino acids is inserted deleted, substituted, added, or It may be a polypeptide having an ACS activity and comprises a sequence.

Coding sequence acetyl CoA acetyl transferase enzyme derived from Candida utilis of (ERG10) gene is represented by SEQ ID NO: 61, the amino acid sequence encoded is represented by SEQ ID NO: 62. Acetyl-CoA acetyltransferase (ERG10) is an enzyme that converts acetoacetyl-CoA to acetyl-CoA. In addition, the amino acid sequence represented by SEQ ID NO: 62, a polypeptide having the activity acetyl-CoA acetyltransferase in (ERG10) is an amino acid sequence that is inserted deleted, substituted, added, or one or several amino acids It may be a polypeptide having an activity ERG10 and containing

The coding sequence of the gene ACL1 Yarrowia lipolytica derived is represented by SEQ ID NO: 31, the amino acid sequence encoded is shown in SEQ ID NO: 32. Also, in the amino acid sequence represented by SEQ ID NO: 32, a polypeptide having the activity of ACL1 poly having a ACL1 activity and comprising an amino acid sequence inserted deleted, substituted, added, or one or several amino acids may be a peptide. Moreover, the coding sequence of the gene ACLY Yarrowia lipolytica derived is represented by SEQ ID NO: 33, the amino acid sequence encoded is shown in SEQ ID NO: 34. In addition, in the amino acid sequence represented by SEQ ID NO: 34, a polypeptide having the activity of ACLY poly having a ACLY activity and comprising an amino acid sequence that has been inserted deletion, substitution, or addition of one or several amino acid residues may be a peptide. And subunit ACLY ACL1 and these ATP- citric acid removal adding enzyme (ACL) is an enzyme that converts acetyl CoA and citric acid.

The coding sequence mitochondrial inner membrane carnitine transporter derived from Candida utilis of (CRC1) gene is represented by SEQ ID NO: 23, the amino acid sequence encoded is shown in SEQ ID NO: 24. Mitochondrial inner membrane carnitine transporter (CRC1) is an enzyme that transport to the cytoplasm from mitochondria acetyl carnitine. In addition, in the amino acid sequence represented by SEQ ID NO: 24, a polypeptide having the activity mitochondrial inner membrane carnitine transporter (CRC1) is an amino acid sequence that is inserted deletion, substitution, or addition of one or several amino acid residues It may be a polypeptide having an activity CRC1 and containing.

Coding sequence carnitine acetyl transferase gene derived from Candida utilis of (YAT1) gene is represented by SEQ ID NO: 27, the amino acid sequence encoded is represented by SEQ ID NO: 28. Carnitine acetyl transferase gene (YAT1) is an enzyme that converts acetyl-CoA acetyl carnitine. In addition, the amino acid sequence represented by SEQ ID NO: 28, a polypeptide having the activity carnitine acetyl transferase gene (YAT1) comprises an amino acid sequence inserted deleted substituted, added, or one or several amino acids , and may be a polypeptide having an activity YAT1.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 601 of 3077

Amino acid deletions, substitutions, additions, or insertions can be done by modifying the techniques known in the art, a gene encoding the polypeptide. It is possible to introduce a mutation in a gene is carried out by a method or an equivalent known method Gapped duplex method, or the Kunkel method, for example mutagenesis kit that utilizes site-directed mutagenesis, Mutant-K or alternatively, using Mutant-G and (Takara Bio Inc.), and the like (Takara Bio Inc.), to introduce the mutation using LA PCR in vitro Mutagenesis series kit of Takara Bio Inc., KOD-Plus-Mutagenesis Kit and (TOYOBO) I can be. Activity of the enzyme described above can be confirmed by procedures known in the art.

Can also be substituted for the above genes, the use of equivalents of each. On condition that it has a function equivalent tc each gene, the equivalents, refers to a gene nucleotide residues of some different. The equivalents of such 70% or more an the respective nucleotide sequence, preferably 80% or more, more preferably 85% or more, more preferably 90% or more, homology of 95% or more and most preferably (sequence identity genes comprising a nucleotide sequence encoding a polypeptide having the activity of each enzyme, and the like) and is there. Furthermore, a gene comprising a nucleotide sequence encoding a polypeptide having the activity of each enzyme, and hybridized under stringent conditions with a complementary sequence thereof, or the nucleotide sequence of each can be mentioned as the equivalent. Furthermore, th nucleotide sequence encoding a polypeptide having an activity of each enzyme and comprises a sequence that has been inserted deletions, nucleotide residues of one or several substitution, addition, or in the nucleotide sequence of each as the equivalent gene, including the like.

Here, nucleotide deletions, substitutions, additions, or insertions can be done by modifying the techniques known in the art, genes containing the sequence. It is possible to introduce a mutation in a gene is carried out by a method or an equivalent known method Gapped duplex method, or the Kunkel method, for example, may be used mutagenesis kit that utilizes site-directed mutagenesis. For example, or, using Mutant-K Mutant-G and (Takara Bio Inc.), and the like (Takara Bio Inc.), using LA PCR in vitro Mutagenesis series kit of Takara Bio Inc., KOD-Plus-Mutagenesis Kit and (TOYOBO) It is possible to introduce mutations. Activity of each enzyme can be confirmed by procedures known in the art.

A number that indicates the (sequence identity) homology (percent), nucleotide sequence comparison program: are those using GENETYX-WIN7.0.0 for example, is calculated by the default parameters (initial setting). That is, each gene in the yeast chromosome, not the same, but functionally equivalent, that is optionally substituted via homologous recombination such as the gene encoding a polypeptide having the activity. Activity of each enzyme can be confirmed by procedures know in the art.

The stringent conditions, for example, using Rapid-Hyb Buffer (GE Healthcare Bio-Science Co., Ltd.), 40~70 ℃ preferably, as 60 ℃ and more preferably, high performed according to the protocol attached to other temperature conditions is the hybridization conditions. Then, for example, using common methods known to those skilled in the art, washing for 5 minutes with a solution consists of (w / v) SDS 0.1% and 2 × SSC, followed by 0.1% and 1 × SSC (refers to do a 10 minute wash with a solution consist w / v) SDS, washed for 10 minutes with a solution consists of (w / v) SDS 0.1% and 0.1 × SSC furthe However by appropriately setting the conditions of salt concentration of the solution used and the temperature condition during hybridization and washing of the membrane then one (70%, 80%, 85%, 90%, 95% certain I can be cloned DNA containing a nucleotide sequence having homology) or more. Gene obtained in this manner is not identical to sequence on but functionally equivalent, that is optionally substituted via homologous recombination such as the gene encoding a polypeptide having the activity. Activity of each enzyme can be confirmed by procedures known in the art.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 602 of 3077

For use in the production of useful substances desired, the yeast strain by gene present invention for the synthesis of useful substances may be transformed with a gene required for production of the substance. For example, in order to produce the metabolites via the acetyl-CoA by the method of the present invention, the yeast strain according to the invention may be transformed with genes for the synthesis of metabolites via acetyl CoA. Necessary genes, may be selected depending on the material to be produced, and appropriately selected by those skilled in the art.

For example, in the case to be produced and isopropanol, as an additional genetic alterations, transforming a yeast strain with isopropanol synthetic gene. The isopropanol synthetic gene, it is possible and enzyme genes described uses a gene known to those skilled in the JP 2011-97929 discloses, for example, CoA sulfotransferase subunit A (ctfA), the subunit B It is possible to use a combination of four genes secondary alcohol dehydrogenase (ctfB), acetoacetic acid decarboxylase and (adc) of (sadh).

According to one embodiment of the present invention, as a further genetic modification, CoA sulfotransferase subunit A (ctfA), CoA transferase subunit B (ctfB), the yeast strain according to the invention, acetoacetic acid decarboxylase it has been transformed by the four genes that encode a polypeptide having the activity secondary alcohol dehydrogenase and (adc) of (sadh). It can be used in the manufacture of isopropanol yeast strains such, this method of manufacture, comprising culturing the yeast strain according to the present embodiment.

Those derived from various organisms are known to these genes, their origin is not particularly limited. For example, these genes can be assumed that the independently derived from Clostridium Beijerinki or Clostridium acetobutylicum, respectively. As to the origin of each enzyme, for example, preferably Clostridium acetobutylicum is (ctfB) same subunit B CoA transferase gene subunit A and (ctfA), and secondary alcohol dehydrogenase acetoacetic acid decarboxylase and (adc Clostridium Beijerinki is preferred (sadh).

The amino acid sequence CoA transferase subunit A of Clostridium acetobutylicum derived (ctfA) is represented by SEQ ID NO: 2. CoA transferase subunit A (ctfA) is an enzyme that together with (ctfB) same subunit B, is converted to acetoacetate to acetoacetyl-CoA. It is possible such arrangement of the gene of Clostridium acetobutylicum from, appropriately selected by those skilled in the art, the gene sequence encoding the amino acid sequence represented by SEQ ID NO: 2 is not particularly limited, according to the invention preferably It is the sequences using codons frequently in species of the yeast strain. For example, gene sequences that conform to the codon frequency in Candida utilis, nucleotide sequence represented by SEQ ID NO: 1. In the amino acid sequence represented by SEQ ID NO: 2, a polypeptide having the activity CoA transferase subunit A of (ctfA) comprises an amino acid sequence inserted deleted, substituted, added, or one or several amino acid residues, and may be a polypeptide having an activity ctfA.

The amino acid sequence CoA transferase subunit B of Clostridium acetobutylicum derived (ctfB) is represented by SEQ ID NO: 4. CoA transferase subunit B (ctfB) is an enzyme that together with (ctfA) same subunit A, is converted to acetoacetate to acetoacetyl-CoA. It is possible such arrangement of the gene of Clostridium acetobutylicum from, appropriately selected by those skilled in the art, the gene sequence encoding the amino acid sequence represented by SEQ ID NO: 4 is not particularly limited, according to the invention preferably It is the sequences using codons frequently in species of the yeast strain. For example, gene sequences that conform to the codon frequency in Candida utilis, nucleotide sequence represented by SEQ ID NO: 3. In the amino acid sequence represented by SEQ ID NO: 4, a polypeptide having the activity CoA transferase subunit B of (ctfB) comprises an amino acid sequence inserted deleted, substituted, added, or one or several amino acid residues, and may be a polypeptide having an activity ctfB.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 603 of 3077

The amino acid sequence acetoacetic acid decarboxylase of Clostridium Beijerinki derived (adc) is represented by SEQ ID NO: 6. Acetoacetic acid decarboxylase (adc) is an enzyme that converts acetone and acetoacetic acid. It is possible such arrangement of the gene of Clostridium Beijerinki from, appropriately selected by those skilled in the art, the gene sequence encoding the amino acid sequence represented by SEQ ID NO: 6 is not particularly limited, according to the invention preferably It is the sequences using codons frequently in species of the yeast strain. For example, gene sequences that conform to the codon frequency in Candida utilis, nucleotide sequence represented by SEQ ID NO: 5, and the like. In the amino acid sequence represented by SEQ ID NO: 6, a polypeptide having the activity acetoacetic acid decarboxylase (adc) is and comprises the amino acid sequence inserted deleted, substituted, added, or one or several amino acids It may be a polypeptide having an activity adc.

The amino acid sequence secondary alcohol dehydrogenase of Clostridium Beijerinki derived (sadh) is represented by SEQ ID NO: 8. Secondary alcohol dehydrogenase (sadh) is an enzyme that converts isopropanol and acetone. It is possible such arrangement of the gene of Clostridium Beijerinki from, appropriately selected by those skilled in the art, the gene sequence encoding the amino acid sequence represented by SEQ ID NO: 8 is not particularly limited, according to the invention preferably It is the sequences using codons frequently in species of the yeast strain. For example, gene sequences that conform to the codon frequency in Candida utilis, nucleotide sequence represented by SEQ ID NO: 7, and the like. In the amino acid sequence represented by SEQ ID NO: 8, a polypeptide having an active secondary alcohol dehydrogenase (sadh) comprises an amino acid sequence inserted deletion, substitution, or addition of one or several amino acid residues and may be a polypeptide having an activity sadh.

Amino acid deletions, substitutions, additions, or insertions can be done by modifying the techniques known in the art, a gene encoding the polypeptide. It is possible to introduce a mutation in a gene is carried out by a method or an equivalent known method Gapped duplex method, or the Kunkel method, for example mutagenesis kit that utilizes site-directed mutagenesis, Mutant-K or alternatively, using Mutant-G and (Takara Bio Inc.), and the like (Takara Bio Inc.), to introduce the mutation using LA PCR in vitro Mutagenesis series kit of Takara Bio Inc., KOD-Plus-Mutagenesis Kit and (TOYOBO) I can be. Activity of the enzyme described above can be confirmed by procedures known in the art.

Can also be substituted for the above genes, the use of equivalents of each. On condition that it has a function equivalent to each gene, the equivalents, refers to a gene nucleotide residues of some different. The equivalents of such 70% or more an the respective nucleotide sequence, preferably 80% or more, more preferably 85% or more, more preferably 90% or more, homology of 95% or more and most preferably (sequence identity genes comprising a nucleotide sequence encoding a polypeptide having the activity of each enzyme, and the like) and is there. Furthermore, a gene comprising a nucleotide sequence encoding a polypeptide having the activity of each enzyme, and hybridized under stringent conditions with a complementary sequence thereof, or the nucleotide sequence of each can be mentioned as the equivalent. Furthermore, the nucleotide sequence encoding a polypeptide having an activity of each enzyme and comprises a sequence that has been inserted deletions, nucleotide residues of one or several substitution, addition, or in the nucleotide sequence of each as the equivalent gene, including the like.

Here, nucleotide deletions, substitutions, additions, or insertions can be done by modifying the techniques known in the art, genes containing the sequence. It is possible to introduce a mutation in a gene is carried out by a method or an equivalent known method Gapped duplex method, or the Kunkel method, for example, may be used mutagenesis kit that utilizes site-directed mutagenesis. For example, or, using Mutant-K Mutant-G and (Takara Bio Inc.), and the like (Takara Bio Inc.), using LA PCR in vitro Mutagenesis series kit of Takara Bio Inc., KOD-Plus-Mutagenesis Kit and (TOYOBO) It is possible to introduce mutations. Activity of each enzyme can be confirmed by procedures known in the art.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 604 of 3077

A number that indicates the (sequence identity) homology (percent), nucleotide sequence comparison program: are those using GENETYX-WIN7.0.0 for example, is calculated by the default parameters (initial setting). That is, each gene in the yeast chromosome, not the same, but functionally equivalent, that is optionally substituted via homologous recombination such as the gene encoding a polypeptide having the activity. Activity of each enzyme can be confirmed by procedures know in the art.

The stringent conditions, for example, using Rapid-Hyb Buffer (GE Healthcare Bio-Science Co., Ltd.), 40~70 ℃ preferably, as 60 ℃ and more preferably, high performed according to the protocol attached to other temperature conditions is the hybridization conditions. Then, for example, using common methods known to those skilled in the art, washing for 5 minutes with a solution consists of (w / v) SDS 0.1% and 2 × SSC, followed by 0.1% and 1 × SSC (refers to do a 10 minute wash with a solution consist w / v) SDS, washed for 10 minutes with a solution consists of (w / v) SDS 0.1% and 0.1 × SSC furthe However by appropriately setting the conditions of salt concentration of the solution used and the temperature condition during hybridization and washing of the membrane then one (70%, 80%, 85%, 90%, 95% certain I can be cloned DNA containing a nucleotide sequence having homology) or more. Gene obtained in this manner is not identical to sequence on but functionally equivalent, that is optionally substituted via homologous recombination such as the gene encoding a polypeptide having the activity. Activity of each enzyme can be confirmed by procedures known in the art.

Also isopropanol addition, it is possible to construct a yeast strain producing substances such as sterols particular produced a particular carotenoid, such as lycopene, squalene or squalene. For example, Candida utilis that express GGPP synthase gene of Erwinia uredovora origin, phytoene synthase gene, phytoene desaturase gene, overexpressing HMG-CoA reductas gene of endogenously (HMG1) high lycopene (Shimada H, Kondo K, Fraser PD, Miura Y, Saito T, Misawa N. Increased carotenoid production by the food yeast Candida utilis through metabolic engineering of the isoprenoid pathway. Appl Environ Microbiol. 1998 Jul which can accumulate has been reported ; 64 (7): 2676-80). Also, ergosta and Zaimosuteroru-24-methyl transferase gene 5,7,24, (28) - Saccharomyces cerevisiae that has a defect in the expression of trienol 22-dehydrogenase gene, has been operated HMG1 is to overexpress recombinant strain, (5-192184 JP) that can accumulate a high concentration of squalene are known. By using these techniques, it is possible to construct a yeast strain that produces lycopene, and squalene.

Applying the yeast strains genetically modified to overexpress the genes described above, for example, molecular breeding of yeast strains by molecular breeding invention yeast strains, yeast strains in an expressible state the gene clusters above can be done by introducing. The genes are, be incorporated into the chromosome of the yeast by the technique of homologous recombination or the like. It is well known in the art, the choice of homologous recombination for gene sequences to achieve integration into the chromosome such, and selecting the appropriate homologous recombination for gene sequence as required by those skilled in the art so that it is possible to construct a homologous recombination DNA fragment for Te.

The chromosomal location of the genes described above is integrated into the yeast genome is not particularly limited, for example, orotidine 5 'phosphate decarboxylase gene (URA3) locus, alpha amino adipic acid reductase gene (LYS2) seat ca be chosen such. In addition, these genes, promoters originally present in the locus incorporated, for example, may be place under the control of the URA3 gene promoter, such as LYS2 gene promoter, glyceraldehyde-3-phosphate dehydrogenase (may be placed under the control of other promoters such as GAP) gene promoter.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 605 of 3077

In accordance with a preferred embodiment of the present invention, the yeast used in breeding the yeast strain according t the invention is a Candida utilis. Its high ploidy, Candida utilis, does not form spores. If it is desired to introduce a mutation ir a gene of the strain with high ploidy, it is necessary compared to strain haploid, added to severe variations thereof but, in doing so, also the gene is not a gene that want to give the mutant I considered the possibility to be added mutation will increase. Therefore, in the case of introducing a mutation in a gene of Candida utilis, using a technique capable of applying mutation in multiple efficiently only the target gene is preferred. The transformation method of Candida utilis, technology include those described in JP Patent Publication No. 2003-144185. In this document, that the available vectors, and sequences homologous to the chromosomal DNA of Candida utilis, and comprises a selectable marker gene, is integrated into the chromosomal DNA of Candida utilis heterologous gene by homologous recombination or which can be, a DNA sequence having an autonomous replication in Candida utilis, which comprise a selection marker gene can transform Candida utilis at high frequency have been developed.

A selection marker gene available in the transformation system, drug resistance markers that can function in yeast, etc. gen conferring hygromycin B resistance gene or confers cycloheximide resistance L41 gene, geneticin (G418) resistance preferably I have a. Gene conferring hygromycin B resistance gene or confer (G418) resistant geneticin is considered because it is a sequence not present in the yeast wild probability which is incorporated locus of the target is high. In additior drug resistance genes, these are considered impact on the traits of the host also small: (Cordero Otero R, et al, (Baganz F e al, 13 (16). 1563-73, 1997) Appl Microbiol Biotechnol, 46 (2):. 143-8, 1996).

As transformation systems other, Cre-loxP system from bacteriophage P1, and the like. This is a site-specific recombination system in the loxP sequences of 34bp between the two, the recombination is catalysed by Cre recombinase Cre gene that encodes. The system can also function in yeast cells such as Saccharomyces cerevisiae has been reported, the selectable marker gene placed between the two loxP sites are known be removed by recombination of the loxP sequence between (Guldener, U., et al, Nucleic Acids Res., 24,2519-24., 1996) that it is. This system (Steensma, H.Y., Et al, Yeast, 18,469-72., 2001) it is being used Kluyveromyces lactis and (Kluyveromyces lactis), in multiple yeast species.

By culturing in the presence of a suitable carbon source and yeast strains by the production method of the present invention substance can be prepared (eg isopropanol) substances in the culture. According to the production method (eg isopropanol) substances according to the invention, by performing the step of separating (eg isopropanol) material from the culture system, it is possible to obtain (eg isopropanol) material. It should be noted that the culture in the present invention, in addition to the culture supernatant, I have included the debris of bacteria bacterial cells or cultured cells, or cells.

Upon culturing the yeast strain according to the present invention, it is possible to choose the culture conditions and culture method for different types of yeast. As a culture method, for example, can be employed to form the culture can be cited liqui culture method using a test tube, a jar fermenter or flask, culture batch, such as fed-batch culture. Aeration and shaking speed is a ventilation amount corresponding to this shaking speed 50~200rpm in general, or 80~150rpm, preferably. Is 0.1 t 30 in general, OD_{600} of the cell amount of the initial, and preferably 0.5 to 10. In addition, when the culture of the yeast strain according to the present invention, the faster is (preferably the greater the amount of ventilation that is) preferably shaking speed, the higher is preferable that the initial OD_{600}. For example, in a culture in a test tube or flask, shaking it speeds greater than or equal to about 120rpm, initial OD_{600} is about 10 is preferred.

In the production process (such as isopropanol) substances according to the invention, the yeast is grown, well, the composition of the medium is not particularly limited as long as it is a composition containing various nutrients can be produced and the desired material. The assimilable carbon source contained in the culture medium, for example, may be

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 606 of 3077

used as long assimilated in sucrose and xylose In addition to glucose. In accordance with a preferred embodiment of the present invention, the carbon source comprises glucose. Glucose concentration in the medium, can be confirmed while the production efficiency of the substance of interest is adjusted appropriately by those skilled in the art. In particular, the culture of the yeast strain according to the invention, glucose concentration in the medium, keeping the relatively low concentration is preferred. For example, in the culture of 192 hours, and 100g / L and the initial concentration of glucose, as compared wit the case where glucose was added solid to a final concentration of 100g / L after 72 hours culture, and 50g / L and the initia concentration of glucose , 48 hours later, those obtained when glucose was added to the solid to a final concentration of 50g / L after 144 hours and after 96 hours is higher production efficiency of the material.

The nutrient source contained in the medium, for example, yeast extract, peptone, and whey is used, but media supplemented with assimilation of carbon sources mentioned above (10g / L yeast extract, 20g / L peptone) in YP, for example YPD5 in medium (50g / L glucose, 10g / L yeast extract, 20g / L peptone) such, those appropriate pH adjustment is more convenient.

The fermentation temperature can be chosen in a range of viable yeast to be used. For example, it can be about 15 ℃ ~45 ℃, 20~40 ℃ more preferably, 25~35 ℃ more preferably the fermentation temperature is about 30 ℃ most preferably. Also, keeping the 3 to 8, more preferably is pH4~7, pH of the medium in the fermentation process can be used and the pH adjusting agent as needed. The pH adjusting agent used, calcium carbonate, sodium hydroxide, potassium hydroxide and the like, but are preferably calcium carbonate.

The reaction time required for manufacturing (eg isopropanol) material is not particularly limited, and is carried out in a reaction time insofar as the effects of the present invention is observed. Optimization of these conditions can be readily made by those skilled in the art.

In the production method (eg isopropanol) substances according to the invention, be separated and recovered from the culture medium ingredients (eg, isopropanol) material produced in this manner, the specific procedures for this reason, the substances to be separated and recovered It is possible according to and readily selected by those skilled in the art.

Will now be described in more detail examples of the present invention by examples which are not intended to limit the scope of the invention.

Example 1: NBRC0396 share selection Candida utilis Candida utilis strain exhibiting the isopropanol resistance high, NBRC0619 shares, NBRC0988 shares, NBRC0639 shares, NBRC0626 shares, NBRC1086 shares, NBRC10707 shares, ATCC9255 shares, ATCC28955 strain, ATCC44638 shares, was inoculated YPD2 (20g / L peptone, 10g / L yeast extract, 20g / L glucose) in, ATCC9206 shares, ATCC96621 strain, ATCC18201 shares, and the ATCC15239 strain, and culture was performed at 30 ℃. Serially diluted 10-fold for each yeast culture in stationary phase, were spotted in YPD2 agar medium containing isopropanol 10~100g / L, and cultured at 30 ℃. And continued for 3 days incubation, was observed the formation of colonies. I listed in Table 1 shows the results.

From the above results, among the strains assayed, that NBRC1086 strain, ATCC9255 strain and ATCC15239 strain, indicating the IPA resistance highest 65g / L revealed.

Example 2: In ATCC15239 shares build Candida utilis ATCC15239 strain Candida utilis pyruvate decarboxylase gene is destroyed, destroyed the (PDC) pyruvate decarboxylase gene and its endogenous gene. The gene disruption was performe by the same method as the first 2010/095750 WO with the Cre-loxP system. By the destruction of the two, I was constructed

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 607 of 3077

strains PDC1 allele was completely deficient in (dPDC1 shares). dPDC1 strain not lacking ethanol production ability to completely, ethanol producing ability was markedly decreased.

Building isopropanol generation of vectors expressing CoA transferase subunit A gene of Clostridium acetobutylicum derive from the same subunit B gene, a secondary alcohol dehydrogenase gene and acetoacetic acid decarboxylase gene of Clostridium Beijerinki from Example 3: secondary alcohol dehydrogenase acetoacetic acid decarboxylase four genes involved in the synthesis, CoA transferase gene subunit A of Clostridium acetobutylicum from (ctfA), the subunit B (ctfB), Clostridium Beijerinki and origin (adc) In order to be expressed in Candida utilis efficiently (sadh) enzymes, and to design a novel gene sequences as an indication that the following items do not exist in nature, were synthesized:
The use of codons that are frequently used in (1) Candida utilis;
For introducing (3) chromosome embedded expression vector, grant of BamH I site at the 3 'end and Xba I site in the 5' end and elimination of inappropriate restriction enzyme site (2) gene cloning.

I is shown in SEQ ID NO: 7 and 1, 3, 5, respectively, the nucleotide sequence CoA transferase gene subunit A gene that ha been synthesized, the B subunit gene, the secondary alcohol dehydrogenase gene and acetoacetic acid decarboxylase gene. Also, I show in 2, 4, 6 and 8 SEQ ID NOs, the amino acid sequence encoded by these nucleotide sequences. I was named, GA09142-sadh GA09142-ctfA respectively, GA09142-ctfB, GA09142-adc a plasmid carrying the gene of these.

It is a built-in vector orotidine 5 'acid decarboxylase gene of Candida utilis to (URA3) seat, pVT92 is, (the first 2010/095750 WO), which has as a selection marker hygromycin resistance gene. By employing the linear DNA by BgII II processing, pVT92 may be introduced chromosome. To express the foreign gene, respectively, pVT92 has a glycerate 3 'phosphate kinase gene terminator glyceraldehyde dehydrogenase gene promoter outside of the Xba I / BamH I gap. In addition, 5 of the promoter 'end and 3 Terminator' to the end side, pVT92 has a Spe I site and Nhe I site, respectively.

It is a built-in vector alpha amino adipic acid reductase gene of Candida utilis to (LYS2) seat, pVT239 is, (the first 2010/095750 WO), which has as a selectable marker geneticin resistance gene. By employing the linear DNA by Not I processing, pVT239 may be introduced chromosome. Also, pVT239 has a Spe I site at the 5 'end of the geneticin resistance gene.

After the BamH I processing and Xba I, the GA09142-sadh and, GA09142-adc, four vectors containing each isopropanol synthase gene GA09142-ctfA, GA09142-ctfB, it was purified by agarose gel extraction, with the same restriction enzyme Th ligated into pVT92 which has been treated to give a pVT302 and pVT296, pVT298, pVT300. pVT300 and pVT296 and Spe I process, I went dephosphorylation. Also, after Nhe I, and Spe I treatment is performed purification of DNA fragments containing each ctfB expression cassette, the sadh expression cassette, and ligated into pVT300 and pVT296 respectively, pVT302 and pVT298 was obtained pVT325 and pVT321. After treatment with Spe I and Nhe I and pVT325, and Purification of a DNA fragment containing the sadh expression cassette and adc expression cassette is introduced into the Spe I site of pVT239, to obtain a pVT346. PVT321 finally obtained is a vector having a hygromycin resistance gene expression cassette of each of ctfB and ctfA, are integrated into the URA3 locus of Candida utilis, pVT346 is the expression cassette of each of sadh and adc It is a vector having a geneticin resistance gene is incorporated into the LYS2 locus of Candida utilis.

Example 4 was digested with BgII II building pVT321 of Candida utilis strains with isopropanol synthetic gene of four, was concentrated by ethanol precipitation. Using the DNA fragment obtained, ATCC15239 strain C. utilis, and was transformed dPDC1 strain obtained in Example 2 was spread on YPD medium containing 600μg / mL hygromycin, and incubated 2 days at 30 ℃ . Were selected a strain having all genes introduced.

Was digested with Not I and pVT346 was further concentrated by ethanol precipitation. Transforming each of the two strains obtained above by using a DNA fragment of each of the obtained was spread on YPD medium containing 200μg / mL geneticin and 600μg / mL hygromycin, and incubated 2 days at 30 ℃. And select a strain with all genes were introduced, TMS272 shares, was the TMS290 shares of isopropanol-producing strain of dPDC strain isopropanol production strains derived from strain ATCC15239.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 608 of 3077

Was carried out as shown below, the evaluation of the isopropanol-producing ability of the transformants obtained with isopropanol production test Example 4; Example 5.

Inoculated (20g / L peptone, 10g / L yeast extract, 20g / L glucose) in / 14mL volume test tube 2mL YPD liquid medium the various transformants were cultured with shaking for 24 hours at, 30 ℃ 140rpm. 100mL YPD10 YPD10C (20g / L peptone, 10g / L yeast extract, 100g / L glucose) or / 200mL Erlenmeyer flask (20g / L peptone, 10g / L yeast extract, 100g / L glucos pre-culture solution obtained, was inoculated to 45g / L calcium carbonate) / 200mL Erlenmeyer flask, and cultured with shaking OD600 = 1,100rpm initial, at 30 ℃. After sampling with time from the culture was filtered through a 0.2μm filter, and subjected to quantitative analysis of the various components. Glucose, glycerol, acetic acid, and ethanol, were quantified using the suggested refractometer high performance liquid chromatographic (manufactured by Shimadzu Corporation), ISep ION300 column (manufactured by Tokyo Chemical Industry Co., Ltd.). Acetone and isopropanol, were quantified using gas chromatographic GC-2010 Plus the (manufactured by Shimadzu Corporation).

Also, I did a similar experiment using (20g / L peptone, 10g / L yeast extract, 100g / L glucose, 45g / L calcium carbonate) a a medium YPD10C neutralizing culture simple in the flask.

I described in Figure 3 shows the results Figure 2, of the culture of the neutralization test conditions the results of the culture test in the non-neutralizing conditions. In addition, I was shown by the average of three repeated trials data.

According to Figure 2, the production of isopropanol while not observed in dPDC1 strain and ATCC15239 strain, which is th parent strain, the TMS290 strains and TMS272 strains introduced isopropanol synthetic genes, 0 and 0.16g / L at each up production of isopropanol .13g / L was observed. In addition, proliferative capacity and rate of consumption of sugar was decreased significantly in the strain that destroyed PDC1 gene. Glucose is completely consumed in the 24 hours of culture, the accumulation of ethanol was observed in the TMS272 share, but ethanol concentration was decreased with the continuation of subsequent culture. Since the production of isopropanol was observed while the ethanol concentration is reduced, the TMS272 strain be producing isopropanol by the secondary consume ethanol was produced at a time is considered. From the isopropanol production rate is reduced in the TMS290 strain derived from dPDC1 strain it in isopropanol production in yeast pathway acetyl-CoA in the cytoplasm, a substrate is fed from ethanol are the most importan was suggested. Since the assimilation of ethanol is an oxidation reaction, these results suggest the possibility that Crabtree negative yeast high respiratory activity such as Candida utilis is suitable in isopropanol production by yeast.

According to Figure 3, by the addition of calcium carbonate to the medium, to produce isopropanol 1.1g / L, production was increased from the non-neutralizing conditions TMS272 strain. Also, TMS272 strain has accumulated acetic acid 16.2g / L, the pH is lowered by acetic acid was accumulated for some reason, it was inhibiting the production of isopropanol was considered. In addition, in the TMS290 share, isopropanol production volume has increased to 0.9g / L in neutralizing conditions, I had accumulated a glycerol concentration as high as 37.4g / L on the other hand.

Followed by selection of the gene to improve the productivity of isopropanol by the over-expression of selected genes for improving the productivity of isopropanol by overexpression Example 6. I described in Figure 1 the metabolic pathway of the candidate gene is catalyst.

Array:; amino acid sequence SEQ ID NO: 11: nucleotide sequence; (ACS1 (;:;; SEQ ID NO: 9 amino acid sequence SEQ ID NO: 10 nucleotide sequence ACH1), acetyl-CoA synthase gene acetyl-CoA hydrolase gene of Candida utilis endogenous SEQ ID NO: 14), (ADH2 cytoplasmic localization types ethanol dehydrogenase gene; nucleotide sequence:; amino acid sequence SEQ ID NO: 13: nucleotide sequence; 12, ACS2 number SEQ ID NO: 15; amino acid sequence SEQ ID NO: 16), mitochondrial localization (SEQ ID NO: 18;:; amino acid sequence SEQ ID NO: 17 nucleotide sequence ADH3), cytoplasmic localization type acetaldehyde dehydrogenase gene (ALD1; nucleotide sequence:;: SEQ ID NO: 20 amino acid sequence SEQ ID NO: 19) type ethanol dehydrogenase gene , mitochondria localized acetaldehyde dehydrogenase gene (ALD2; nucleotide sequence:,: SEQ ID NO: 22 amino acid sequence SEQ ID NO: 21), (CRC1 mitochondrial inner membrane carnitine transporter; nucleotide sequence:;: SEQ ID NO: 24 amino acid sequence SEQ ID NO: 23), (CAT2 mitochondrial

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 609 of 3077

localization carnitine-acetyl-CoA transferase; nucleotide sequence:; SEQ ID NO: 26), (YAT1 carnitine acetyl transferase gene; nucleotide sequence: the amino acid sequence SEQ ID NO: 25; SEQ ID NO: 28 amino acid sequence SEQ ID NO: 2 , YAT2; nucleotide sequence: SEQ ID NO: 61; amino acid sequence:;:; nucleotide sequence (ERG10 acetyl-CoA acetyltransferase SEQ ID NO: 30), the amino acid sequence SEQ ID NO: 29 SEQ ID NO: 62), of Yarrowia lipolytica from ATP- citric each, (SEQ ID NO: 34 genes constituting the subunit to SEQ ID NO: 32, ACLY;:;; nucleotide sequence:;:; amino acid sequence SEQ ID NO: 31 amino acid sequence SEQ ID NO: 33 nucleotide sequence ACL1) the acid removal adding enzyme gene I was amplified by PCR as a template genomic DNA. SEQ ID NO: 36 and SEQ ID NO: 35), (ACS1;; Primers used for PCR ;(ACH1 are as follows: SEQ ID NO: 38 and SEQ ID NO: 37), (ACS2; SEQ ID NO: 40 and SEQ ID NO: 39), (ADH2 ; SEQ ID NO: 42 and SEQ ID NO: 41), (ADH3; SEQ ID NO: 44 and SEQ ID NO: 43), (ALD1; SEQ ID NO: 46 and SEQ ID NO: 45), (ALD2; SEQ ID NO: 48 and SEQ ID NO: 47), (CAT2; array SEQ ID NO: 50 and number 49), (CRC1; SEQ ID NO: 52 to SEQ ID NO: 51), (YAT1; SEQ ID NO: 54 to SEQ ID NO: 53), (YAT2; SEQ ID NO: 56 to SEQ ID NO: 55), (ACL1 SEQ ID NO: 57 SEQ ID NO: 58), (ACLY; SEQ ID NO: 60 to SEQ ID NO: 59), (ERG10; SEQ ID NO: 64 and SEQ ID NO: 63)

PCU155 (Keiji Kondo,, Yutaka Miura, Hidetaka Sone, Kazuo Kobayashi & Hiroshi Iijima., High-level expression of a sweet protein, monellin that with Xba I / BamH I treated PCR product was obtained, which was treated with the same restriction enzyme , in the food yeast Candida utilis, 15, 453 Nature Biotechnology - I was cloned into the 457 (1997)).. have as a selection marker cycloheximide resistance gene, is a vector that is multi-copy integration in the URA3 locus, respectively, pCU155 has the same terminator and glyceraldehyde dehydrogenase gene promoter on the outside of the Xba I / BamH I gap have. Isopropanol synthase gene is incorporated into the URA3 locus of isopropanol production strains are building so far, C.utilis is higher polyploid, because the URA3 locus multiple copies exist, the URA3 locus further gene transfer is possible. Also, since the isopropanol synthetic gene is inserted by the two-point recombination URA3 locus, no incorporatior of pCU155, replacement of the introduced genes occurs. was digested with BglI II plasmid candidate gene has been cloned. and concentrated by ethanol precipitation. Transforming the TMS272 strain C.utilis using a DNA fragment obtained was spread on YPD medium containing 20μg / mL cycloheximide 600μg / mL hygromycin, and 200μg / mL geneticin and culture for 4 days at 30 ℃. In addition, DNA of ACLY and ACL1 were mixed, I transformation was carried out at the same time. The fermentation test carried out for each 4 clones obtained were investigated isopropanol productivity. I The results are shown in FIG.

According to Figure 4, due to overexpression of the ACL or ACS1, ACS2, CRC1, YAT1, ERG10,, isopropanol production volume will be 1.5 to 4.0 times has been revealed. Each gene of the ACL is an enzyme genes involved in the synthesis of acetyl-CoA and ACS1, ACS2, YAT1,, CRC1 gene is a carrier protein gene that exported to the cytoplasm from the mitochondria the precursor of acetyl-CoA synthesis, ERG10 gene acetyl CoA and a gene of synthesizing acetoacetyl-CoA. I These results, in order of isopropanol high production by yeast, to up-regulate the expression or activity of an enzyme gene for efficiently producing acetyl-CoA in the cytoplasm is important is shown.

Culture test building and isopropanol-producing strains overexpressing simultaneously YAT1 and, ACS2, ERG10, CRC1 ACL1 / ACLY: 7 Example

Was digested with BglI II plasmids each gene of YAT1 and, ACS2, ERG10, CRC1 ACL1, ACLY has been cloned, and concentrated by ethanol precipitation. The mixed DNA fragments of each 5μg obtained were transformed TMS272 strain by using this mixture was spread on YPD medium containing 10μg / mL cycloheximide 600μg / mL hygromycin, and 200μg / m geneticin at 30 ℃ I was cultured for 4 days. The fermentation test conducted on 80 clones obtained, and select a strain having the highest production of isopropanol, was the TMS378 strain.

Are shown in Figure 5 the result of the culture test of TMS378 shares. TMS272 shares whereas the production of isopropanol 1.3g / L, TMS378 shares showed isopropanol production volume of about 7.5 times and 9.8g / L. Moreover, whereas ethyl accumulated 18.1g / L The strain TMS272, TMS378 strain and 4.5g / L, acetate is a major by-product was reduced to about one-quarter.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 610 of 3077

And examined the culture conditions of TMS378 strains suitable for the production of isopropanol impact speed shaking anc initial OD 600 in isopropanol production of TMS378 share: Example 8. Inoculated (20g / L peptone, 10g / L yeast extract, 20 / L glucose) in / 14mL volume test tube 2mL YPD liquid medium TMS378 strain was cultured with shaking for 24 hours at, 3(℃ 140rpm. Were inoculated (20g / L peptone, 10g / L yeast extract, 100g / L of glucose, 45g / L calcium carbonate) in / 300mL Erlenmeyer flask 100mL YPD10C a preculture obtained was 10 or 1 initial OD 600, were cultured with shaking at, 3(℃ 120rpm or 100rpm. Was added to the culture medium of glucose solids to a final concentration of 100g / L at 72 hours of culture, and cultured up to 144 hours. Was filtered through a 0.2μm filter by sampling over time. Various quantitative analysis, I was performed as described in Example 5. In addition, I was shown by the average of three repeated trials data. Figure 6 shows the results.

According to Figure 6, the highest isopropanol production was observed when cultured at an initial OD600 = 10,120rpm, isopropanol 18.5g / L was produced up to 144 hours of culture. Yield based on sugar at that time was 30.8%. From this, alsc (as the aeration amount is large) as the speed is higher, the higher the initial OD is higher, it is preferred that the isopropano production revealed shaking.

I have examined the culture conditions of TMS378 strains suitable for the production of impact isopropanol addition glucose concentration in isopropanol production of TMS378 share: Example 9. Inoculated (20g / L peptone, 10g / L yeast extract, 20 / L glucose) in / 14mL volume test tube 2mL YPD liquid medium TMS378 strain was cultured with shaking for 24 hours at, 3(℃ 140rpm. 100mL YPD10C YPD5C (20g / L peptone, 10g / L yeast extract, 100g / L glucose, 45g / L calcium carbonate) or 300mL Erlenmeyer flask (20g / L peptone, 10g / L yeast extract pre-culture solution obtained , was inoculated 50g / L glucose, to 45g / L calcium carbonate) / 300mL Erlenmeyer flask, and cultured with shaking OD600 = 1,120rpm early, at 30 ℃. For samples that were cultured in YPD10C, and glucose was added solid to a final concentration of 100g / L in 72 hours of culture, and cultured up to 192 hours. As for samples and cultured in YPD5C, and glucose was added to the solid to be a 3 Kaitsui concentration 50g / L at 48 hour intervals the culture, and cultured up to 192 hours. Glucose concentration in the final is a 200g / L any of the conditions. Was filtered through a 0.2μm filter by sampling over time. Various quantitative analysis, I was performed as described in Example 5. In addition, I was shown by the average of three repeated trials data. Figure 7 shows the results.

According to Figure 7, when cultured in YPD10C, followed by the addition of glucose 100g / L after 72 h incubation, isopropanol 19.2g / L was produced up to 122 hours of culture. However, the subsequent increase in isopropanol productior was not observed, acetic acid is accumulated (27.2g / L). Yield based on sugar of the maximum in the production of isopropanol was 29.0%. On the other hand, if it is cultured in YPD5C, was added 3 times of glucose 50g / L in 48-hour intervals, increase of isopropanol was observed up to 192 hours of culture, isopropanol of 27.4g / L production up to 192 hours after culture has been. In addition, there was no recognized these are little accumulation of acetic acid. Yield based oı sugar of the maximum in the production of isopropanol under these conditions was 41.5%. Therefore, preferable revealed ir isopropanol production keeping the relatively low concentration of glucose concentration to be added.

Example 10: In order to identify genes having a large contribution to improving the isopropanol production by over-expression at the same time specific gene large contribution isopropanol production increase, overexpression combinations ACS2 (i) and ACL1, ACLY strains overexpressing strain overexpressing lines over-expressing strains to be a combination of ERG10 (ii) and ACL1, ACLY, combination of ERG10 and (iii) ACS2, combination of ERG10 and, ACS2 (iv) ACL1, ACLY between the overexpression strains of single gene of each stock to build, respectively, overexpressing a combination of YAT1 and, ACS2, CRC1, ERG10 ACL1, ACLY and (TMS378 shares), a comparison was made of isopropanol production volume. Was performed by using the same method as construction of recombinant strain in Example 7 and Example 6, was mixed with genomic DNA to be built, and transformation was carried out at the same time. The fermentation test carried out for each 32 resulting clones were selected lines displaying isopropanol production average in the combination of the transgene, respectively. Inoculated (20g / L peptone, 10g / L yeast extract, 20g / L glucose) in / 14mL volume test tube 2mL YPD liquid medium build each strain was cultured with shaking for 24 hours at, 30 ℃ 140rpm. Inoculated (20g / L peptone, 10g / L yeast extract, 100g / L glucose, 45g / L calcium carbonate) in / glass test tube 3mL YPD10C pre-culture solution

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 611 of 3077

obtained, OD600 = 1,100rpm early, at 30 ℃ were cultured with shaking. Sampling was carried out over time, to evaluate the amount of isopropanol production by gas chromatographic GC-2010 and high-performance liquid chromatographic. In addition, I was shown by the average of three repeated trials data.

According to Figure 8, the overexpression of a single, isopropanol production is 2.2 to 2.7 times in each of the ERG10 and ACS1, ACS2 revealed. Isopropanol production volume of a strain overexpressing a combination of ERG10 (ii) and ACL1, ACLY and strain overexpressing a combination of ACS2 (i) and ACL1, ACLY includes a strain overexpressing alone ERG10 and ACS2, respectively significant difference was not observed. On the other hand, a dramatic improvement of isopropanol production observed in strains overexpressing a combination of ERG10 and (iii) ACS2, production volume was improved to 11.6 times. In addition, line over-expressing strains overexpressing a combination of ERG10 and (iii) ACS2, a combination c ERG10 and, ACS2 (iv) ACL1, ACLY, the combination of YAT1 and, ACS2, CRC1, ERG10 (ACL1, ACLY TMS378 shares In strains of the strain) overexpressing, no significant difference was observed in the isopropanol production volume.

Contribution overexpression of YAT1 and, CRC1 ACL gives improve productivity isopropanol is small, it can be simultaneously overexpressed the ERG10 and ACS2 in order to improve the isopropanol productivity when the results of these were simultaneously overexpressed But it is the most important has been shown. The above results, the metabolic pathways for the synthesis of acetyl-CoA from ethanol, namely to enhance the expression or activity of genes involved in metabolic pathways for the synthesis of acetyl-CoA from acetic acid is important for isopropanol high production by yeast I have shown that.

Figure 1 shows the synthetic route of isopropanol by yeast. Is a graph showing the results of a fermentation test for non-neutralizing conditions of strain introduced isopropanol production gene, isopropanol concentration in the medium (panel A) glucose concentration (panel B), ethanol concentration Figure 2 (panel shows the time course of (panel F) OD600 C), acid concentration (panel D), and glycerol concentration (panel E). In each panel, filled circles shows the results of the ATCC15239 strain, the white circles indicate the results TMS272 strain (ATCC15239 shares + IPA synthetic gene), black triangle shows the results of the PDC1 disrupted strain, TMS290 shares open triangles (PDC disrupted strain + IPA I shows the results of the synthetic gene). Is a graph showing the results of a fermentation test in the neutralization conditions of the strains introduced isopropanol production gene, isopropanol concentration in the medium (panel A), glucose concentration (panel B), the ethanol concentration Figure 3 (panels C I shows the time-dependent change glycerol concentration and acetic acid concentration (panel D) of (panel E),). In each panel, filled circles shows the results of the ATCC15239 strain, the white circles indicate the results TMS272 strain (ATCC15239 shares + IPA synthetic gene), black triangle shows the results of the PDC1 disrupted strain, TMS290 shares open triangles (PDC disrupted strain + IPA I shows the results of the synthetic gene). Figure 4 is a graph in TMS272 strain showed a relative amount with respect to isopropanol production by strain TMS272, isopropanol production in case of being over-expressed the various genes. Is a graph showing the results of the fermentation test of the strain overexpressing all genes for improving the isopropanol production acidic isopropanol concentration in the medium (panel A), glucose concentration (panel B), the ethanol concentration Figure 5 (Panel shows changes over time acetone concentration C), acetic acid concentration (panel D), glycerol concentration (panel E) of (panel F). In each panel, filled circles shows the results of the TMS272 shares, open circles (TMS272 shares + ACL1 YAT1; ACLY: ACS2: ERG10:: CRC1) TMS378 shares shows the results of. Is a graph showing the results of (effect of shaking speed and initial OD_{600}) fermentation test at 30 ℃ the TMS378 strains that overexpress all the genes for improving the isopropanol productivity, 6, isopropanol concentration in the culture medium shows changes over time acetone concentration (panel A), glucose concentration (panel B), ethanol concentration (panel C), acetic acid concentration (panel D), glycerol concentratior (panel E) of (panel F). In each panel, solid circle shows the results under the conditions of stirring rate = 100rpm, and initial OD600 = 1, the black triangles show the results under the condition of stirring rate = 120rpm, and initial OD600 = 1, the ope circles are the initial OD600 = shows the results under the conditions of stirring rate = 100rpm and 10, white triangle shows the results under the conditions of stirring rate = 120rpm and initial OD600 = 10. Figure 7 shows the results of (effect of fed glucose) fermentation test at 30 ℃ the TMS378 strain overexpressing all genes for improving the productivity of isopropanol And 100g / L initial glucose concentration, and fed glucose 100g / l after 72 h incubation, Panel A shows the result of the 19

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 612 of 3077

hours of culture. It is a 50g / L initial glucose concentration, 48 hours after culture in fed glucose 50g / l respectively after 14 hours and 96 hours later, Panel B shows the result of the 192 hours of culture. In each panel, and shows the time course concentration of acetic acid concentration of isopropanol (closed triangles), glucose concentration (closed circles), and ethanol concentration (open circles) (white triangles). 8 is a graph in TMS272 strain showed isopropanol production in case of being over-expressed the various genes.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 613 of 3077

PATENT COOPERATION TREATY

From the INTERNATIONAL PRELIMINARY EXAMINING AUTHORITY

To:
SILVIA SALVADORI, ESQ.
SILVIA SALVADORI, P.C.
270 MADISON AVENUE, 8TH FLOOR
NEW YORK, NEW YORK 10016
UNITED STATES OF AMERICA

PCT

WRITTEN OPINION OF THE INTERNATIONAL PRELIMINARY EXAMINING AUTHORITY

(PCT Rule 66)

Date of mailing *(day/month/year)*	13 SEP 2019
Applicant's or agent's file reference: **SS118-1026**	**REPLY DUE** within **2 MONTHS** from the above date of mailing.

International application No.	International filing date *(day/month/year)*	Priority date *(day/month/year)*
PCT/US2018/041579	**11 July 2018 (11.07.2018)**	**13 July 2017 (13.07.2017)**

International Patent Classification (IPC) or both national classification and IPC
IPC: C12N 1/16, 5/10, 9/88, 9/96, 15/09, 15/52; C07H 21/04; C07K 14/415
CPC: See Supplemental Sheet

Applicant: **RADICI CHIMICA S.P.A.**

1. ☑ The written opinion established by the International Searching Authority:
   ☑ is ☐ is not
   considered to be a written opinion of the International Preliminary Examining Authority.
2. This second opinion contains indications relating to the following items:
   - ☑ Box No. I Basis of the opinion
   - ☐ Box No. II Priority
   - ☑ Box No. III Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
   - ☐ Box No. IV Lack of unity of invention
   - ☑ Box No. V Reasoned statement under Rule 66.2(a)(ii) with regard to novelty, inventive step and industrial applicability; citations and explanations supporting such statement
   - ☐ Box No. VI Certain documents cited
   - ☑ Box No. VII Certain defects in the international application
   - ☐ Box No. VIII Certain observations on the international application
3. The applicant is hereby **invited to reply** to this opinion
   When? See the time limit indicated above. The applicant may, before the expiration of that time limit, request this Authority to grant an extension, see Rule 66.2(e).
   How? By submitting a written reply, accompanied, where appropriate, by amendments, according to Rule 66.3. For the form and the language of the amendments, see Rules 55.3 and 66.8.
   Also For the examiner's obligation to consider amendments and/or arguments, see Rule 66.4bis. For an informal communication with the examiner, see Rule 66.6. For an additional opportunity to submit amendments, see Rule 66.4.
   If no reply is filed, the international preliminary examination report will be established on the basis of this opinion
4. The final date by which the international preliminary report on patentability (Chapter II of the PCT) must be established according to Rule 69.2 is: 13 November 2019 (13.11.2019)

Name and mailing address of the IPEA/US	Authorized officer
COMMISSIONER FOR PATENTS MAIL STOP PCT, ATTN: IPEA/US P.O. BOX 1450 ALEXANDRIA, VA 22313-1450, UNITED STATES OF AMERICA	**MANJUNATH RAO**
Facsimile No. **(571)273-8300**	Telephone No. **571-272-0939**

Form PCT/IPEA/408 (Cover sheet) (January 2015)

WRITTEN OPINION OF THE INTERNATIONAL PRELIMINARY EXAMINING AUTHORITY

International application No.
PCT/US2018/041579

Box No. I Basis of the opinion

1. With regard to the **language**, this opinion has been established on the basis of:
   - [x] the international application in the language in which it was filed.
   - [ ] a translation of the international application into _____ which is the language of a translation furnished for the purposes of:
     - [ ] international search (Rules 12.3(a) and 23.1(b))
     - [ ] publication of the international application (Rule 12.4(a))
     - [ ] international preliminary examination (Rules 55.2(a) and/or 55.3(a) and (b))
2. With regard to the **elements** of the international application, this opinion has been established on the basis of *(replacement sheets which have been furnished to the receiving Office in response to an invitation under Article 14 are referred to in this opinion as "originally filed.")*:
   - [ ] the international application as originally filed/furnished
   - [x] the description:
     pages **1-550** as originally filed/furnished.
     pages _____ received by this Authority on _____
     pages _____ received by this Authority on _____
   - [x] the claims:
     Nos. _____ as originally filed/furnished.
     Nos. _____ as amended (together with any statement) under Article 19
     Nos. **1-155** received by this Authority on **10 June 2019 (10.06.2019)**
   - [x] the drawings:
     pages **1/59-59/59** as originally filed/furnished.
     pages _____ received by this Authority on _____
     pages _____ received by this Authority on _____
   - [ ] a sequence listing - see Supplemental Box Relating to Sequence Listing.
3. [x] The amendments have resulted in the cancellation of:
   - [ ] the description, pages _____
   - [x] the claims, Nos. **5,44,45,84-86,149**
   - [ ] the drawings, sheets/figs _____
   - [ ] the sequence listing *(specify)*: _____
4. [ ] This opinion has been established as if (some of) the amendments listed below had not been made, since either they are considered to go beyond the disclosure as filed, or they were not accompanied by a letter indicating the basis for the amendments in the application as filed, as indicated in the Supplemental Box (Rules 70.2(c) and (c-*bis*)):
   - [ ] the description, pages _____
   - [ ] the claims, Nos. _____
   - [ ] the drawings, sheets/figs _____
   - [ ] the sequence listing *(specify)*: _____
5. [ ] This opinion has been established:
   - [ ] taking into account the **rectification of an obvious mistake** authorized by or notified to this Authority under Rule 91 (Rules 66.1(d-*bis*)).
   - [ ] without taking into account the **rectification of an obvious mistake** authorized by or notified to this Authority under Rule 91(Rules 66.4*bis*).
6. [ ] Supplementary international search report(s) from Authority(ies) _____ has/ have been received and taken into account in establishing this opinion (Rule 45*bis*.8(b) and (c)).

WRITTEN OPINION OF THE INTERNATIONAL PRELIMINARY EXAMINING AUTHORITY

International application No.
PCT/US2018/041579

Box No. III Non-establishment of opinion with regard to novelty, inventive step and industrial applicability

The questions whether the claimed invention appears to be novel, to involve an inventive step (to be non obvious), or to be industrially applicable have not been examined in respect of:

☐ the entire international application.

☑ claims Nos. **113 and 114**

because:

☐ the said international application, or the said claims Nos. _____ relate to the following subject matter which does not require an international preliminary examination (*specify*):

☑ the description, claims or drawings *(indicate particular elements below)* or said claims Nos. **113 and 114** are so unclear that no meaningful opinion could be formed (*specify*):

Claims 113 and 114 are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

☐ the claims, or said claims Nos. _____ are so inadequately supported by the description that no meaningful opinion could be formed (*specify*):

☐ no international search report has been established for said claims Nos. _____

☐ a meaningful opinion could not be formed without the sequence listing; the applicant did not, within the prescribed time limit:

☐ furnish a sequence listing in the form of an Annex C/ST.25 text file, and such listing was not available to the International Preliminary Examining Authority in the form and manner acceptable to it; or the sequence listing furnished did not comply with the standard provided for in Annex C of the Administrative Instructions.

☐ furnish a sequence listing on paper or in the form of an image file complying with the standard provided for in Annex C of the Administrative Instructions, and such listing was not available to the International Preliminary Examining Authority in the form and manner acceptable to it; or the sequence listing furnished did not comply with the standard provided for in Annex C of the Administrative Instructions.

☐ pay the required late furnishing fee for the furnishing of a sequence listing in response to an invitation under Rules 13*ter*.1(a) or (b) and 13*ter*.2.

☐ See Supplemental Box for further details.

WRITTEN OPINION OF THE INTERNATIONAL PRELIMINARY EXAMINING AUTHORITY

International application No.
PCT/US2018/041579

Box No. V **Reasoned statement under Rule 66.2(a)(ii) with regard to novelty, inventive step and industrial applicability; citations and explanations supporting such statement**

1. Statement

Novelty (N)	Claims	1-112 and 115-155	YES
	Claims		NO
Inventive step (IS)	Claims	1-112 and 115-155	YES
	Claims		NO
Industrial applicability (IA)	Claims	1-112 and 115-155	YES
	Claims		NO

2. Citations and explanations:

---------- **EXPLANATION** ----------

The claims possess novelty and an inventive step under PCT Article 33(2) & 33(3) as the prior art does not teach nor suggest a genetically modified microorganism, such as a yeast/*Candida* having one or more heterologous terpene biosynthesis enzymes selected from those in claim 1, wherein the expression of said enzymes is provided by a nucleic acid that is responsive to fatty acid or alkane induction (the former for cytochrome P450).

The claims possess industial applicability under PCT Article 33(4) as production of terpenes are important in many industrial fields such as foods, cosmetics, pharmaceutical, etc.

WRITTEN OPINION OF THE INTERNATIONAL PRELIMINARY EXAMINING AUTHORITY

International application No.
PCT/US2018/041579

Box No. VII Certain defects in the international application

The following defects in the form or contents of the international application have been noted:

In claims 140-142, beginning of second lines, a space should exist between "encoding" and the following word.

WRITTEN OPINION OF THE INTERNATIONAL PRELIMINARY EXAMINING AUTHORITY

International application No.
PCT/US2018/041579

Supplemental Box

In case **the space in any of the preceding boxes is not sufficient.**

Continuation of CPC:
C12N 1/16, 5/10, 9/88, 9/96, 15/09, 15/52, 15/63, 15/70, 15/74, 15/82, 15/8249; C07H 21/04; C07K 14/415; C12P 7/02, 7/26

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
26 June 2014 (26.06.2014)

WIPO | PCT

(10) International Publication Number
WO 2014/100504 A2

(51) International Patent Classification:
C12N 1/16 (2006.01)

(21) International Application Number:
PCT/US2013/076739

(22) International Filing Date:
19 December 2013 (19. 12.2013)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/739,661 19 December 2012 (19. 12.2012) US

(71) Applicant: **VERDEZYNE, INC.** [US/US]; 2715 Loker Avenue West, Carlsbad, CA 92010 (US).

(72) Inventors: **BEARDSLEE, Tom;** 2715 Loker Avenue West, Carlsbad, CA 92010 (US). **HUTAGALUNG, Alex;** 2715 Loker Avenue West, Carlsbad, CA 92015 (US). **PICATAGGIO, Stephen;** 2715 Loker Avenue West, Carlsbad, CA 92010 (US).

(74) Agents: **GIRINATH, Prashant** et al; Wilson Sonsini Goodrich & Rosati, 650 Page Mill Road, Palo Alto, CA 94304-1050 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available):* AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available):* ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

(54) Title: BIOLOGICAL METHODS FOR PREPARING A FATTY DICARBOXYLIC ACID

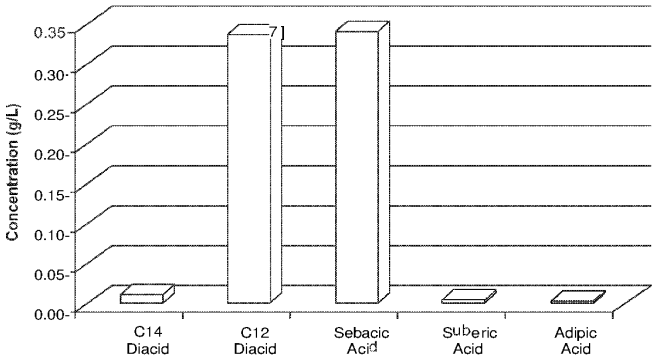


FIG. 17

(57) Abstract: The technology relates in part to biological methods for producing a fatty dicarboxylic acid and engineered microorganisms capable of such production. Provided are engineered microorganisms capable of producing fatty dicarboxylic acids and products expressed by such microorganisms. Also provided are biological methods for producing fatty dicarboxylic acids.

WO 2014/100504 A2

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 620 of 3077

BIOLOGICAL METHODS FOR PREPARING A FATTY DICARBOXYLIC ACID

Cross Reference to Related Applications

This application claims the benefit under 35 U.S.C. § 1 19(e) of U.S. Provisional Patent Application No. 61/739,661, filed December 19, 2012 which is herein incorporated by reference in its entirety.

Field

The technology relates in part to biological methods for producing a fatty dicarboxylic acid and engineered microorganisms capable of such production.

Background

Microorganisms employ various enzyme-driven biological pathways to support their own metabolism and growth. A cell synthesizes native proteins, including enzymes, in vivo from deoxyribonucleic acid (DNA). DNA first is transcribed into a complementary ribonucleic acid (RNA) that comprises a ribonucleotide sequence encoding the protein. RNA then directs translation of the encoded protein by interaction with various cellular components, such as ribosomes. The resulting enzymes participate as biological catalysts in pathways involved in production of molecules by the organism.

These pathways can be exploited for the harvesting of the naturally produced products. The pathways also can be altered to increase production or to produce different products that may be commercially valuable. Advances in recombinant molecular biology methodology allow researchers to isolate DNA from one organism and insert it into another organism, thus altering the cellular synthesis of enzymes or other proteins. Advances in recombinant molecular biology methodology also allow endogenous genes, carried in the genomic DNA of a microorganism, to be increased in copy number, thus altering the cellular synthesis of enzymes or other proteins. Such genetic engineering can change the biological pathways within the host organism, causing it to produce a desired product. Microorganic industrial production can minimize the use of caustic chemicals and the production of toxic byproducts, thus providing a "clean" source for certain compounds. The use of appropriate plant derived feedstocks allows production of "green" compounds while further minimizing the need for and use of petroleum derived compounds.

Summary

Provided in certain aspects is a genetically modified yeast comprising (i) an active, modified endogenous acyl-coA oxidase polypeptide or an active, modified endogenous acyl-coA dehydrogenase polypeptide; (ii) a genetic modification that alters an enoyl co-A isomerase activity; and (iii) a genetic modification that alters a dienoyl CoA reductase (DCR) activity, wherein the yeast is capable of producing a diacid from a feedstock comprising one or more components from a vegetable oil. Also provided in some aspects is a method for using such a genetically modified yeast to produce a diacid from a feedstock comprising one or more components from a vegetable oil.

Also provided in certain aspects is a method for producing a diacid, comprising: (a) contacting a genetically modified yeast of any of embodiments A1 to A44, B1 to B33, CI to C15 and G1 to G65 with a feedstock comprising an unsaturated vegetable fatty acid capable of being converted by the yeast to a diacid; (b) culturing the yeast under conditions in which diacid containing one or more unsaturations is produced from the feedstock; and (c) hydrogenating and removing one or more of the unsaturations.
Certain embodiments are described further in the following description, examples, claims and drawings.

Brief Description of the Drawings

The drawings illustrate embodiments of the technology and are not limiting. For clarity and ease of illustration, the drawings are not made to scale and, in some instances, various aspects may be shown exaggerated or enlarged to facilitate an understanding of particular embodiments.
FIG. 1 is a schematic representation of the conversion of decane to sebacic acid in a beta-oxidation blocked microorganism. Capric acid is formed as an intermediate during omega oxidation.

FIG. 2 is a schematic representation of the conversion of dodecane to dodecanedioic acid in a beta-oxidation blocked microorganism. Laurie acid is formed as an intermediate during omega oxidation.

FIG. 3 is a schematic representation of the conversion of a feedstock containing mixed chain-length alkanes to mixed diacids products, including sebacic acid in a beta-oxidation blocked microorganism. Mixed chain-length fatty acids are formed as intermediates during omega oxidation. Sebacic acid can be separated from other diacid products by the use of appropriate separation techniques.

FIG. 4 is a schematic representation of the conversion of a feedstock containing mixed chain-length alkanes to mixed diacids products, including dodecanedioic acid in a beta-oxidation blocked microorganism. Mixed chain-length fatty acids are formed as intermediates during omega oxidation. Dodecanedioic acid can be separated from other diacid products by the use of appropriate separation techniques.

FIG. 5 is a schematic representation of the conversion of a long-chain alkane into sebacic acid in a partially beta-oxidation blocked microorganism. The long-chain alkane is first converted into a long-chain fatty acid and then into a long-chain diacid by activities in the omega-oxidation pathway. The long-chain diacid can be converted to sebacic acid by activities in the beta-oxidation pathway, with the simultaneous generation of acetyl-CoA.

FIG. 6 is a schematic representation of the conversion of a long-chain alkane into dodecanedioic acid in a partially beta-oxidation blocked microorganism. The long-chain alkane is first converted into a long-chain fatty acid and then into a long-chain diacid by activities in the omega-oxidation pathway. The long-chain diacid can be converted to dodecanedioic acid by activities in the beta-oxidation pathway, with the simultaneous generation of acetyl-CoA.

FIG. 7 is a schematic representation of the conversion of a feedstock containing mixed chain-length alkanes into sebacic acid in a partially beta-oxidation blocked microorganism. The mixed chain-length alkanes are first converted into mixed chain-length fatty acids and then mixed diacids by activities in the omega-oxidation pathway. Mixed diacids can be converted to sebacic acid by activities in the beta-oxidation pathway, with the simultaneous generation of acetyl-CoA.

FIG. 8 is a schematic representation of the conversion of a feedstock containing mixed chain-length alkanes into dodecanedioic acid in a partially beta-oxidation blocked microorganism.

The mixed chain-length alkanes are first converted into mixed chain-length fatty acids and then mixed diacids by activities in the omega-oxidation pathway. Mixed diacids can be converted to dodecanedioic acid by activities in the beta-oxidation pathway, with the simultaneous generation of acetyl-CoA.

FIG. 9 graphically illustrates the conversion of decane to sebacic acid in a fully beta-oxidation blocked Candida yeast strain. After incubation for the times shown in the graph, the media was subjected to gas chromatography. The results indicate that greater than 99% of the decane was converted into sebacic acid, with a minimal amount of capric acid also detected by gas chromatography. No significant accumulation of any other monoacid or diacid was detected by gas chromatography. Experimental details and results are given in Example 1.

FIG. 10 graphically illustrates the conversion of capric acid to sebacic acid in a Candida yeast strain. GC analysis was performed after a predetermined period of growth. Nearly all the capric acid added was converted to sebacic acid using a starting concentration of capric acid. Experimental details and results are given in Example 2.

FIG. 11 graphically illustrates the distribution of diacids produced during the conversion of long-chain fatty acids to mixed diacids under fermentation conditions using a partially beta-oxidation blocked Candida tropicalis strain (e.g., sAA106). Experimental details and results are given in Example 5.

FIG. 12 graphically illustrates the conversion of decane to sebacic acid in a fully beta-oxidation blocked Candida yeast strain having additional genetic modifications. Strain sAA003 is the fully beta-oxidation blocked control strain. +CPR indicates the fully beta-oxidation blocked strain also includes an increased number of copies of cytochrome P450 reductase. +CPR+A12 indicates starting strain sAA003 includes the addition genetic modifications of an increased number of copies of cytochrome P450 reductase and also includes an increased number of copies of cytochrome P450 A12 (e.g., CYP52A12). +CPR+A18 indicates starting strain sAA003 includes the addition genetic modifications of an increased number of copies of cytochrome P450 reductase and also includes an increased number of copies of cytochrome P450 A18 (e.g., CYP52A18). +CPR+A19 indicates starting strain sAA003 includes the addition genetic modifications of an increased number of copies of cytochrome P450 reductase and also includes an increased number of copies of cytochrome P450 A19 (e.g., CYP52A19).

+CPR+A20 indicates starting strain sAA003 includes the addition genetic modifications of an increased number of copies of cytochrome P450 reductase and also includes an increased number of copies of cytochrome P450 A20 (e.g., CYP52A20). The y-axis of FIG. 12 is percent of theoretical maximum yield. Experimental details and results are given in Example 7.

FIG. 13 graphically illustrates the results of conversion of methyl laurate to dodecanedioic acid in a fully beta-oxidation blocked Candida yeast strain also contain genetic alterations to a monooxygenase reductase activity, a monooxygenase activity, or a monooxygenase reductase activity and a monooxygenase activity. After 48 hours of incubation the media was subjected to gas chromatography. The results indicate that Candida strains containing an increased number of copies of a CYP52A18 monooxygenase activity and an increased number of copies of a monooxygenase reductase activity (e.g., CPR750) gave the highest yield of dodecanedioic acid (e.g., DDDA), in shake flask fermentation experiments. Experimental details and results are given in Example 8.

FIG. 14 and FIG. 15 schematically illustrate a screening and/or selection method for identifying acyl-CoA oxidase activities with specific substrate specificities. The method can be utilized in conjunction with generating and/or identifying acyl-CoA oxidase activities with altered chain-length substrate specificities. Screening/selection method details are given in Example 9.

FIG. 16 graphically illustrates the results of engineered microorganisms described herein converting decane to sebacic acid under fermentation conditions using different amounts of decane as the feedstock. Experimental details and results are given in Example 3.

FIG. 17 graphically illustrates the results of engineered microorganisms described herein converting a mixed fatty acid feedstock (e.g., mixed chain-length fatty acids) to sebacic acid under fermentation conditions. Experimental details and results are given in Example 4.

FIG. 18 shows a diagram of a plasmid designated pAA073 containing a POX4 promoter and a POX4 terminator.

FIG. 19 illustrates the generation of a full-length deletion cassette for ECU using PCR overlap extension.

FIG. 20 illustrates the generation of a full-length deletion cassette for the second allele of ECU using PCR overlap extension.

FIG. 2 1 shows an acyl CoA oxidase activity profile for Pox5 isolated from a Candida strain.

FIG. 22 illustrates a PCR overlap extension method for introducing site-directed point mutations into Acyl-CoA Oxidase genes.

FIG. 23 - shows a sequence alignment of the N-terminal 180 amino acids of Acol and AcoII. The amino acids highlighted in grey are located within alpha helices and those in bold are located within beta sheets. The center sequence represents the consensus sequence showing conserved residues.

FIG. 24 - illustrates a HotSpot Wizard analysis of Pox4 (FIG. 24A) and Pox5 (FIG. 24B) from Candida strain ATCC20336. Residues highlighted in dark grey or light grey are mutagenic "hot spots". Dark grey residues show greater variability at that position than light grey residues. Residues in bold are found within or close to the substrate binding pocket.

FIG. 25 - illustrates a HotSpot Wizard analysis of RnAcoII. Residues highlighted in dark grey or light grey are mutagenic "hot spots". Dark grey residues show greater variability at that position than light grey residues. Residues in bold are found within or close to the substrate binding pocket.

FIG. 26 shows a multiple sequence alignment of all three proteins. The underlined portion of RnAcoII (AcoII from *R. norvegicus)* represents the alternatively spliced exon 3.

FIG. 27 shows the acyl CoA activity profile associated with Pox5 mutants. The substrate referred to as "CI 8" is shortened and pertains to a CI8:1 substrate.
FIG. 28 shows the acyl CoA activity profile associated with Pox4 mutants.

FIG. 29 shows a diagram of a plasmid designated pAA298.

FIG.30A and FIG.30B show an amino acid alignment of selected Acyl Co-A enzymes from Candida viswanathii (POX4 and POX5) and Yarrowia lipolytica (Yl POX2, Yl POX3 and Yl POX5). Also shown is a consensus sequence alignment.

FIG. 31 shows the synthesis of sebacic acid from linoleic acid.

FIG. 32 shows plasmid vector map of pAA918.

FIG. 33 shows the activity profiles of six Y. lipolytica acyl-CoA oxidase proteins in addition to POX4 and POX5 from ATCC20336. The activity profiles for these enzymes was determined using an in vitro acyl-CoA oxidase assay (e.g., as described in Ex. 41) and acyl-CoA substrates of different chain lengths (e.g., C6, six carbons; C8, eight carbons; CIO, ten carbons; C12, twelve carbons; C14, fourteen carbons; and C18-1, eighteen carbons).

FIG. 34 shows a schematic example of a reaction pathway that includes ACO, ECI, DCI (Dienoyl CoA Isomerase) and a DCR as well as their reaction products and substrates.

Detailed Description

Certain fatty dicarboxylic acids (i.e., diacids, e.g., dodecanedioic acid or sebacic acid) are chemical intermediates in manufacturing processes used to make certain polyamides, polyurethanes and plasticizers, all of which have wide applications in producing items such as antiseptics, top-grade coatings, hot-melt coating and adhesives, painting materials, corrosion inhibitor, surfactant, engineering plastics and can also be used as a starting material in the manufacture of fragrances, for example. For example dodecanedioic acid, also known as 1,12 dodecanedioic acid, and DDDA, is a 12 carbon organic molecule that is a fatty dicarboxylic acid. In another example, sebacic acid, also known as 1,10 decanedioic acid, and 1,8 octanedicarboxylic acid, is a 10 carbon organic molecule that is a fatty dicarboxylic acid.

Provided herein are methods for producing a fatty dicarboxylic acid (also referred to herein as a diacid). Any suitable diacid can be produced, and a diacid produced often includes acid moieties at each terminus of the molecule (e.g., alpha omega diacids). A diacid sometimes is a C4 to a C24 diacid (i.e., a diacid containing 4 carbons to 24 carbons) and sometimes is a C8, CIO, C12, C14, C16, C18, or C20 diacid. Yeast and processes herein are capable of producing

a diacid containing an odd number of carbons, and sometimes a product contains one or more diacids chosen from a C5, C7, C9, C11, CI3, CI5, CI7, CI9, C21 and C23 diacid. A hydrocarbon portion of a diacid sometimes is fully saturated and sometimes a diacid includes one or more unsaturations (e.g., double bonds).

Non-limiting examples of diacids include octadecanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid) and other organic intermediates using biological systems. Non-limiting examples of fatty dicarboxylic acids include suberic acid (i.e., octanedioic acid, 1,8-octanedioic acid, octanedioic acid, octane-l,8-dioic acid, 1,6-hexanedicarboxylic acid, capryllic diacids), sebacic acid (i.e., 1,10-decanedioic acid, decanedioic acid, decane-l,10-dioic acid, 1,8-octanedicarboxylic acid, capric diacid), dodecanedioic acid (i.e., DDDA, 1,12-dodecanedioic acid, dodecanedioic acid, dodecane-l,12-dioic acid, 1,10-decanedicarboxylic acid, decamethylenedicaboxylic acid, 1,10-dicarboxydecane, lauric diacid), tetradecanedioic acid (i.e., TDDA, 1,14-tetradecanedioic acid, tetradecanedioic acid, tetradecane-l,14-dioic acid, 1,12-dodecanedicarboxylic acid, myristic diacid), thapsic acid (i.e., hexadecanedioic acid, 1,16-hexadecanedioic acid, hexadecanedioic acid, hexadecane-l,16-dioic acid, 1,14-tetradecanedicarboxylic acid, palmitic diacid), cis-9-hexadecenedioic acid (i.e., palmitoleic diacids), octanedioic acid (i.e., 1,18-octadecanedioic acid, octadecanedioic acid, octadecane-1,18-dioic acid, 1,16-hexadecanedicarboxylic acid, stearic diacid), cis-9-octadecenedioic acid (i.e., oleic diacids), cis-9,12-octadecenedioic acid (i.e., linoleic diacids), cis-9, 12,15-octadecenedioic acid (i.e., linolenic diacids), arachidic diacid (i.e., eicosanoic diacid, icosanoic diacid), 11-eicosenoic diacid (i.e., cis-1 1-eicosenedioic acid), 13-eicosenoic diacids (i.e., cis-13-eicosenedioic acid), arachidonic diacid (i.e., cis-5,8,1 1,14-eicosatetraenedioic acid).

A genetically modified yeast can be provided with a feedstock to produce a diacid, and the feedstock sometimes includes a substantially pure aliphatic molecule from which the diacid is produced. In certain embodiments, the feedstock contains a mixed set of aliphatic molecules from which diacids may be produced. In some embodiments, an aliphatic molecule in the feedstock is the predominant aliphatic species and sometimes a particular diacid produced from that aliphatic molecule is the predominant diacid species produced. A predominant species generally is 51% or more by weight of aliphatic molecule species in a feedstock or 51% or more by weight of diacid species in a product (e.g., about 55% or more, 60%> or more, 65%> or more, 70%> or more, 75%> or more, 80%> or more, 85%> or more, 90%> or more or 95%> or more).

Such production systems may have significantly less environmental impact and could be economically competitive with current manufacturing systems. Thus, provided in part herein are methods for manufacturing a fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid) by engineered microorganisms. In some embodiments microorganisms are engineered to contain at least one heterologous gene encoding an enzyme, where the enzyme is a member of a novel and/or altered pathway engineered into the microorganism. In certain embodiments, an organism may be selected for elevated activity of a native enzyme.

Microorganisms

A microorganism selected often is suitable for genetic manipulation and often can be cultured at cell densities useful for industrial production of a target fatty dicarboxylic acid product. A microorganism selected often can be maintained in a fermentation device.

The term "engineered microorganism" as used herein refers to a modified microorganism that includes one or more activities distinct from an activity present in a microorganism utilized as a starting point (hereafter a "host microorganism"). An engineered microorganism includes a heterologous polynucleotide in some embodiments, and in certain embodiments, an engineered organism has been subjected to selective conditions that alter an activity, or introduce an activity, relative to the host microorganism. Thus, an engineered microorganism has been altered directly or indirectly by a human being. A host microorganism sometimes is a native microorganism, and at times is a microorganism that has been engineered to a certain point.

In some embodiments an engineered microorganism is a single cell organism, often capable of dividing and proliferating. A microorganism can include one or more of the following features: aerobe, anaerobe, filamentous, non-filamentous, monoploid, dipoid, auxotrophic and/or non-auxotrophic. In certain embodiments, an engineered microorganism is a prokaryotic microorganism (e.g., bacterium), and in certain embodiments, an engineered microorganism is a non-prokaryotic microorganism. In some embodiments, an engineered microorganism is a eukaryotic microorganism (e.g., yeast, fungi, amoeba). In some embodiments, an engineered microorganism is a fungus. In some embodiments, an engineered organism is a yeast.

Any suitable yeast may be selected as a host microorganism, engineered microorganism, genetically modified organism or source for a heterologous or modified polynucleotide. Yeast include, but are not limited to, Yarrowia yeast (e.g., Y. lipolytica (formerly classified as Candida lipolytica)), Candida yeast (e.g., C. revkaufi, C. viswanathii, C. pulcherrima, C. tropicalis, C. utilis), Rhodotorula yeast (e.g., R. glutinus, R. graminis), Rhodosporidium yeast (e.g., R. toruloides), Saccharomyces yeast (e.g., S. cerevisiae, S. bayanus, S. pastorianus, S. carlsbergensis), Cryptococcus yeast, Trichosporon yeast (e.g., T. pullans, T. cutaneum), Pichia yeast (e.g., P. pastoris) and Lipomyces yeast (e.g., L. starkeyii, L. lipoferus). In some embodiments, a suitable yeast is of the genus Arachniotus, Aspergillus, Aureobasidium, Auxarthron, Blastomyces, Candida, Chrysosporuim, Chrysosporuim Debaryomyces, Coccidiodes, Cryptococcus, Gymnoascus, Hansenula, Histoplasma, Issatchenkia, Kluyveromyces, Lipomyces, Lssatchenkia, Microsporum, Myxotrichum, Myxozyma, Oidiodendron, Pachysolen, Penicillium, Pichia, Rhodosporidium, Rhodotorula, Rhodotorula, Saccharomyces , Schizosaccharomyces, Scopulariopsis, Sepedonium, Trichosporon, or Yarrowia. In some embodiments, a suitable yeast is of the species Arachniotus flavoluteus, Aspergillus flavus, Aspergillus fumigatus, Aspergillus niger, Aureobasidium pullulans, Auxarthron thaxteri, Blastomyces dermatitidis, Candida albicans, Candida dubliniensis, Candida famata, Candida glabrata, Candida guilliermondii, Candida kefyr, Candida krusei, Candida lambica, Candida lipolytica, Candida lustitaniae, Candida parapsilosis, Candida pulcherrima, Candida revkaufi, Candida rugosa, Candida tropicalis, Candida utilis, Candida viswanathii, Candida xestobii, Chrysosporuim keratinophilum, Coccidiodes immitis, Cryptococcus albidus var. diffiuens, Cryptococcus laurentii, Cryptococcus neofomans, Debaryomyces hansenii, Gymnoascus dugwayensis, Hansenula anomala, Histoplasma capsulatum, Issatchenkia occidentalis, Isstachenkia orientalis, Kluyveromyces lactis, Kluyveromyces marxianus, Kluyveromyces thermotolerans, Kluyveromyces waltii, Lipomyces lipoferus, Lipomyces starkeyii, Microsporum gypseum, Myxotrichum deflexum, Oidiodendron echinulatum, Pachysolen tannophilis, Penicillium notatum, Pichia anomala, Pichia pastoris, Pichia stipitis, Rhodosporidium toruloides, Rhodotorula glutinus, Rhodotorula graminis, Saccharomyces cerevisiae, Saccharomyces kluyveri, Schizosaccharomyces pombe, Scopulariopsis acremonium, Sepedonium chrysospermum, Trichosporon cutaneum, Trichosporon pullans, Yarrowia lipolytica, or Yarrowia lipolytica (formerly classified as Candida lipolytica). In some embodiments, a yeast is a Y. lipolytica strain that includes, but is not limited to, ATCC20362, ATCC8862, ATCC 18944, ATCC20228, ATCC76982 and LGAM S(7)1 strains (Papanikolaou S., and Aggelis G., Bioresour. Technol. 82(1):43-9 (2002)). In

certain embodiments, a yeast is a Candida species (i.e., Candida spp.) yeast. Any suitable Candida species can be used and/or genetically modified for production of a fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid). In some embodiments, suitable Candida species include, but are not limited to Candida albicans, Candida dubliniensis, Candida famata, Candida glabrata, Candida guilliermondii, Candida kefyr, Candida krusei, Candida lambica, Candida lipolytica, Candida lustitaniae, Candida parapsilosis, Candida pulcherrima, Candida revkaufi, Candida rugosa, Candida tropicalis, Candida utilis, Candida viswanathii, Candida xestobii and any other Candida spp. yeast described herein. Non-limiting examples of Candida spp. strains include, but are not limited to, sAAOOl (ATCC20336), SAA002 (ATCC20913), sAA003 (ATCC20962), sAA496 (US20 12/0077252), sAA106 (US20 12/0077252), SU-2 (ura3-/ura3-), H5343 (beta oxidation blocked; US Patent No. 5648247) strains. Any suitable strains from Candida spp. yeast may be utilized as parental strains for genetic modification.

Yeast genera, species and strains are often so closely related in genetic content that they can be difficult to distinguish, classify and/or name. In some cases strains of C. lipolytica and Y. lipolytica can be difficult to distinguish, classify and/or name and can be, in some cases, considered the same organism. In some cases, various strains of C.tropicalis and C.viswanathii can be difficult to distinguish, classify and/or name (for example see Arie et.al., J. Gen. Appl. Microbiol., 46, 257-262 (2000). Some C. tropicalis and C.viswanathii strains obtained from ATCC as well as from other commercial or academic sources can be considered equivalent and equally suitable for the embodiments described herein. In some embodiments, some parental strains of C.tropicalis and C.viswanathii are considered to differ in name only.

Any suitable fungus may be selected as a host microorganism, engineered microorganism or source for a heterologous polynucleotide. Non-limiting examples of fungi include, but are not limited to, Aspergillus fungi (e.g., A. parasiticus, A. nidulans), Thraustochytrium fungi, Schizochytrium fungi and Rhizopus fungi (e.g., R. arrhizus, R. oryzae, R. nigricans). In some embodiments, a fungus is an A. parasiticus strain that includes, but is not limited to, strain ATCC24690, and in certain embodiments, a fungus is an A. nidulans strain that includes, but is not limited to, strain ATCC38163.

Any suitable prokaryote may be selected as a host microorganism, engineered microorganism or source for a heterologous polynucleotide. A Gram negative or Gram positive bacteria may be selected. Examples of bacteria include, but are not limited to, Bacillus bacteria (e.g., B. subtilis, B. megaterium), Acinetobacter bacteria, Norcardia baceteria, Xanthobacter bacteria, Escherichia bacteria (e.g., E. coli (e.g., strains DH10B, Stbl2, DH5-alpha, DB3, DB3.1), DB4, DB5, JDP682 and ccdA-over (e.g., U.S. Application No. 09/518,188))), Streptomyces bacteria, Erwinia bacteria, Klebsiella bacteria, Serratia bacteria (e.g., S. marcessans), Pseudomonas bacteria (e.g., P. aeruginosa), Salmonella bacteria (e.g., S. typhimurium, S. typhi), Megasphaera bacteria (e.g., Megasphaera elsdenii). Bacteria also include, but are not limited to, photosynthetic bacteria (e.g., green non-sulfur bacteria (e.g., Choroflexus bacteria (e.g., C. aurantiacus), Chloronema bacteria (e.g., C. gigateum)), green sulfur bacteria (e.g., Chlorobium bacteria (e.g., C. limicola), Pelodictyon bacteria (e.g., P. luteolum), purple sulfur bacteria (e.g., Chromatium bacteria (e.g., C. okenii)), and purple non-sulfur bacteria (e.g., Rhodospirillum bacteria (e.g., R. rubrum), Rhodobacter bacteria (e.g., R. sphaeroides, R. capsulatus), and Rhodomicrobium bacteria (e.g., R. vanellii)).

Cells from non-microbial organisms can be utilized as a host microorganism, engineered microorganism or source for a heterologous polynucleotide. Examples of such cells, include, but are not limited to, insect cells (e.g., Drosophila (e.g., D. melanogaster), Spodoptera (e.g., S. frugiperda Sf9 or Sf21 cells) and Trichoplusa (e.g., High-Five cells); nematode cells (e.g., C. elegans cells); avian cells; amphibian cells (e.g., Xenopus laevis cells); reptilian cells; mammalian cells (e.g., NIH3T3, 293, CHO, COS, VERO, C127, BHK, Per-C6, Bowes melanoma and HeLa cells); and plant cells (e.g., Arabidopsis thaliana, Nicotania tabacum, Cuphea acinifolia, Cuphea aequipetala, Cuphea angustifolia, Cuphea appendiculata, Cuphea avigera, Cuphea avigera var. pulcherrima, Cuphea axilliflora, Cuphea bahiensis, Cuphea baillonis, Cuphea brachypoda, Cuphea bustamanta, Cuphea calcarata, Cuphea calophylla, Cuphea calophylla subsp. mesostemon, Cuphea carthagenensis, Cuphea circaeoides, Cuphea confertiflora, Cuphea cordata, Cuphea crassiflora, Cuphea cyanea, Cuphea decandra, Cuphea denticulata, Cuphea disperma, Cuphea epilobiifolia, Cuphea ericoides, Cuphea flava, Cuphea flavisetula, Cuphea fuchsiifolia, Cuphea gaumeri, Cuphea glutinosa, Cuphea heterophylla, Cuphea hookeriana, Cuphea hyssopifolia (Mexican-heather), Cuphea hyssopoides, Cuphea ignea, Cuphea ingrata, Cuphea jorullensis, Cuphea lanceolata, Cuphea linarioides, Cuphea llavea, Cuphea lophostoma, Cuphea lutea, Cuphea lutescens, Cuphea melanium, Cuphea melvilla, Cuphea micrantha, Cuphea micropetala, Cuphea mimuloides, Cuphea nitidula,

Cuphea palustris, Cuphea parsonsia, Cuphea pascuorum, Cuphea paucipetala, Cuphea procumbens, Cuphea pseudosilene, Cuphea pseudovaccinium, Cuphea pulchra, Cuphea racemosa, Cuphea repens, Cuphea salicifolia, Cuphea salvadorensis, Cuphea schumannii, Cuphea sessiliflora, Cuphea sessilifolia, Cuphea setosa, Cuphea spectabilis, Cuphea spermacoce, Cuphea splendida, Cuphea splendida var. viridiflava, Cuphea strigulosa, Cuphea subuligera, Cuphea teleandra, Cuphea thymoides, Cuphea tolucana, Cuphea urens, Cuphea utriculosa, Cuphea viscosissima, Cuphea watsoniana, Cuphea wrightii, Cuphea lanceolata)

Microorganisms or cells used as host organisms or source for a heterologous polynucleotide are commercially available. Microorganisms and cells described herein, and other suitable microorganisms and cells are available, for example, from Invitrogen Corporation, (Carlsbad, CA), American Type Culture Collection (Manassas, Virginia), and Agricultural Research Culture Collection (NRRL; Peoria, Illinois).

Host microorganisms and engineered microorganisms may be provided in any suitable form. For example, such microorganisms may be provided in liquid culture or solid culture (e.g., agar-based medium), which may be a primary culture or may have been passaged (e.g., diluted and cultured) one or more times. Microorganisms also may be provided in frozen form or dry form (e.g., lyophilized). Microorganisms may be provided at any suitable concentration.

Carbon Processing Pathways and Activities

Figures 1-8 schematically illustrate non-limiting embodiments of engineered pathways that can be used to produce a fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid) from various starting carbon sources or feedstocks. FIG. 1 depicts an embodiment of a non-limiting engineered biological pathway for producing sebacic acid in microorganisms having a fully blocked beta-oxidation pathway, using decane as the carbon source starting material. FIG. 2 depicts an embodiment of a non-limiting engineered biological pathway for producing dodecanedioic acid in microorganisms having a fully blocked beta-oxidation pathway, using dodecane as the carbon source starting material. Figures 3 and 4 depict an embodiment of a non-limiting engineered biological pathway for producing mixed chain-length diacids in a microorganism having a fully blocked beta-oxidation pathway, using mixed chain-length alkanes as the carbon source starting material. Sebacic acid (FIG. 3) and dodecanedioic acid (FIG. 4) can be separated and/or purified away from other diacid products

using a suitable combination of centrifugation, organic solvent extraction, chromatography, and/or other purification/separation techniques. Figures 5 and 6 depict an embodiment of a non-limiting engineered biological pathway for producing sebacic acid (FIG. 5) and dodecanedioic acid (FIG. 6) in microorganisms having a partially blocked beta oxidation pathway, using long-chain alkanes as the carbon source starting material. Figures 7 and 8 depict an embodiment of a non-limiting engineered biological pathway for producing sebacic acid (FIG. 7) and dodecanedioic acid (FIG. 8) in microorganisms having a partially blocked beta oxidation pathway, using mixed-chain length alkanes as the carbon source starting material.

The alkane carbon source starting materials are initially metabolized using naturally occurring and/or engineered activities in naturally occurring and/or engineered pathways to yield an intermediate alcohol which can then be converted to a carboxylic acid (e.g., fatty acid) by the action of other naturally occurring and/or engineered activities in the omega-oxidation pathway depicted in Figures 1-8.

Alkanes are omega-hydroxylated by the activity of cytochrome P450 enzymes, thereby generating the equivalent chain-length alcohol derivative of the starting alkane carbon source material. In certain embodiments, a cytochrome P450 activity can be increased by increasing the number of copies of a cytochrome P450 gene (e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the gene), by increasing the activity of a promoter that regulates transcription of a cytochrome P450 gene, or by increasing the number of copies of a cytochrome P450 gene and increasing the activity of a promoter that regulates transcription of a cytochrome P450 gene, thereby increasing the production of target product (e.g., sebacic or dodecanedioic acid) via increased activity of one or more cytochrome P450 enzymes. In some embodiments, a cytochrome P450 enzyme is endogenous to the host microorganism. One or more cytochrome P450 activities can be added and/or increased dependent on the carbon source starting material, in certain embodiments. Cytochrome P450's sometimes exhibit increased activities in response to stimulation by certain feedstocks or carbon source starting materials. In some embodiments, an engineered microorganism includes an increased number of copies of one or more cytochrome P450s that are stimulated by a chosen carbon source starting material or feedstock. Cytochrome P450 responsiveness to a chosen starting carbon source or feedstock can be determined using any suitable assay. Non-limiting examples of assays suitable for identification of cytochrome P450 responsiveness to a starting carbon source or feedstock

include RT-PCR or qRT-PCR after the host microorganism has been exposed to the chosen carbon source or feedstock for varying amounts of time.

Cytochrome P450 is reduced by the activity of cytochrome P450 reductase (CPR), thereby recycling cytochrome P450 to allow further enzymatic activity. In certain embodiments, the CPR enzyme is endogenous to the host microorganism. In some embodiments, host CPR activity can be increased by increasing the number of copies of a CPR gene (e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the gene), by increasing the activity of a promoter that regulates transcription of a CPR gene, or by increasing the number of copies of a CPR gene and increasing the activity of a promoter that regulates transcription of a CPR gene, thereby increasing the production of target product (e.g., sebacic or dodecanedioic acid) via increased recycling of cytochrome P450. In certain embodiments, the promoter can be a heterologous promoter (e.g., endogenous or exogenous promoter). In some embodiments, the CPR gene is heterologous and exogenous and can be isolated from any suitable organism. Non-limiting examples of organisms from which a CPR gene can be isolated include C. tropicalis, S. cerevisiae and Bacillus megaterium.

Oxidation of the alcohol to an aldehyde may be performed by an enzyme in the fatty alcohol oxidase family (e.g., long-chain fatty alcohol oxidase EC 1.1.3.20), or an enzyme in the alcohol dehydrogenase family (e.g., fatty alcohol dehydrogenase; EC 1.1.1.1). The aldehyde may be oxidized to a carboxylic acid (e.g., sebacic or dodecanedioic acid) by the activity of the enzyme aldehyde dehydrogenase (e.g., long-chain-aldehyde dehydrogenase or fatty aldehyde dehydrogenase; EC 1.2.1.48). In some embodiments, the long chain fatty alcohol oxidase, fatty alcohol dehydrogenase and/or the long-chain-aldehyde dehydrogenase exist in a host organism. Flux through these two steps may sometimes be augmented by increasing the copy number of the enzymes, or by increasing the activity of the promoter transcribing the genes. In some embodiments alcohol and aldehyde dehydrogenases specific for 10, 12 or 14 carbon substrates may be isolated from another organism, and inserted into the host organism.

FIG. 1 depicts a non-limiting embodiment of an engineered biological pathway for making sebacic acid using decane (e.g., a CIO alkane) as the carbon source starting material. Due to the carbon chain length of decane, no chain shortening is necessary to arrive at the 10 carbon diacid, sebacic acid. Thus a fully beta oxidation blocked microorganism can be utilized to minimize conversion of the desired 10 carbon diacid into diacids having shorter chain lengths.

FIG. 2 depicts a non-limiting embodiment of an engineered biological pathway for making dodecanedioic acid using dodecane (e.g., a C12 alkane) as the carbon source starting material. Due to the carbon chain length of dodecane, no chain shortening is necessary to arrive at the 12 carbon diacid, dodecanedioic acid. Thus a fully beta oxidation blocked microorganism can be utilized to minimize conversion of the desired 12 carbon diacid into diacids having shorter chain lengths.

Figures 3 and 4 depict a non-limiting embodiment of an engineered biological pathway for generating a mixed population of diacid (fatty dicarboxylic acid) products, including sebacic acid (FIG. 3) and dodecanedioic acid (FIG. 4), using a carbon source or feedstock that contains mixed-chain-length alkanes as the carbon source starting material. Any suitable mixed-chain-length alkane, fatty alcohol, mixed chain length fatty alcohol feedstock, fatty acid, mixed fatty acid feedstock, paraffin, fat or oil can be used. In some embodiments, the distribution of carbon chain lengths in the starting material is substantially similar to the desired carbon chain length distribution in the mixed diacid product. In certain embodiments, the feedstock is enriched for a desired chain length. In some embodiments, the enriched fraction is enriched for carbon chain lengths of about 10 carbons. In some embodiments, the enriched fraction is enriched for carbon chain lengths of about 12 carbons. Because, in some embodiments, the diacids generated have substantially the same chain lengths as the chain lengths found in the carbon source starting material, a fully beta-oxidation blocked microorganism can be utilized to minimize conversion of the diacids of desired chain length into diacids of shorter chain lengths. The lower part of the pathways in FIG. 3 AND FIG. 4 show the separation of sebacic acid and dodecanedioic acid, respectively, away from the mixed diacid products by the use of separation techniques described herein, or those known in the art.

In certain embodiments involving genetically modified organisms having partially blocked beta-oxidation pathways (see Figures 5 - 8), feedstocks suitable for use include, but are not limited to, fatty acid distillates or soapstocks of renewable oils (palm oil fatty acid distillate, soybean oil soapstock, coconut oil soapstock), renewable oils (coconut oil, palm oil, palm kernel oil, soybean oil, corn oil, etc.), fatty acids of chain length equal to or greater than CIO (in substantially single form (e.g., in substantially pure form) or in mixture form, alkanes of chain length equal to or greater than CIO in substantially single form (e.g., substantially pure form) or in mixture form.

Carbon sources with longer chain lengths (e.g., 12 carbons or greater in length) can be metabolized using naturally occurring and/or engineered pathways to yield molecules that can be further metabolized using the beta oxidation pathway shown in the lower portion of Figures 5 through 8. In some embodiments, beta-oxidation activities in the pathways shown in Figures 5 through 8 also can be engineered (e.g., as described herein) to enhance metabolism and target product formation. In some embodiments, one acyl-CoA oxidase activity of the beta-oxidation pathway is engineered to be enhanced, and in certain embodiments, the other acyl-CoA oxidase activity in the beta-oxidation pathway is altered to reduce or eliminate the activity, thereby optimizing the production of a diacid of a desired chain-length or diacids with a distribution of desired chain lengths. In some embodiments, an acyl-CoA oxidase is selected and/or engineered to alter the substrate specificity of the enzyme. In certain embodiments, the substrate specificity of a heterologous and/or engineered acyl-CoA oxidase is for carbon chain lengths of between about 12 carbons and about 18 carbons, and in some embodiments a heterologous and/or engineered acyl-CoA oxidase exhibits no activity on substrates below 12 carbons in length. In certain embodiments, a heterologous acyl-CoA oxidase with a desired chain length specificity can be isolated from any suitable organism. Non-limiting examples of organisms that include, or can be used as donors for acyl-CoA oxidase enzymes include yeast (e.g., Candida, Saccharomyces, Debaryomyces, Meyerozyma, Lodderomyces, Scheffersomyces, Clavispora, Yarrowia, Pichia, Kluyveromyces, Eremothecium, Zygosaccharomyces, Lachancea, Nakaseomyces), animals (e.g., Homo, Rattus), bacteria (e.g., Escherichia, Pseudomonas, Bacillus), or plants (e.g., Arabidopsis, Nictotania, Cuphea).

In certain embodiments, a carbon source starting material (e.g., alkane, fatty acid, fatty alcohol, dicarboxylic acid) of intermediate or long chain length (e.g., between about 10 carbons and 22 carbons) is converted into an acyl-CoA derivative for entry into the beta-oxidation pathway. The acyl-CoA derivative can be generated by the activity of an acyl-CoA ligase enzyme, in some embodiments. The acyl-CoA derivative is subsequently oxidized by the activity of an acyl-CoA oxidase enzyme (e.g., also known as acyl-CoA oxidoreductase and fatty acyl-coenzyme A oxidase) of natural or altered substrate specificity, in certain embodiments. The trans-2,3-dehydroacyl-CoA derivative long chain fatty alcohol, fatty acid or dicarboxylic acid may be further converted to 3-hydroxyacyl-CoA by the activity of enoyl-CoA hydratase. 3-hydroxyacyl-CoA can be converted to 3-oxoacyl-CoA by the activity of 3-hydroxyacyl-CoA dehydrogenase. 3-oxoacyl-CoA may be converted to an acyl-CoA molecule, shortened by 2

carbons and an acetyl-CoA, by the activity of Acetyl-CoA C-acyltransferase (e.g., also known as beta-ketothiolase and beta-ketothiolase). In some embodiments, acyl-CoA molecules may be repeatedly shortened by beta oxidation until a desired carbon chain length is generated (e.g., 10 or 12 carbons, sebacic acid or dodecanedioic acid, respectively). A shortened fatty acid can be further processed using omega oxidation to yield a dicarboxylic acid (e.g., dodecanedioic acid).

Beta-oxidation activities

The term "beta oxidation pathway" as used herein, refers to a series of enzymatic activities utilized to metabolize fatty alcohols, fatty acids, or dicarboxylic acids. The activities utilized to metabolize fatty alcohols, fatty acids, or dicarboxylic acids include, but are not limited to, acyl-CoA ligase activity, acyl-CoA oxidase activity, acyl-CoA hydrolase activity, acyl-CoA thioesterase activity, enoyl-CoA hydratase activity, 3-hydroxyacyl-CoA dehydrogenase activity and acetyl-CoA C-acyltransferase activity. The term "beta oxidation activity" refers to any of the activities in the beta oxidation pathway utilized to metabolize fatty alcohols, fatty acids or dicarboxylic acids.

Beta-oxidation - Acyl-CoA Ligase

An acyl-CoA ligase enzyme sometimes is encoded by the host organism and can be added to generate an engineered organism. In some embodiments, host acyl-CoA ligase activity can be increased by increasing the number of copies of an acyl-CoA ligase gene, by increasing the activity of a promoter that regulates transcription of an acyl-CoA ligase gene, or by increasing the number copies of the gene and by increasing the activity of a promoter that regulates transcription of the gene, thereby increasing production of target product (e.g., sebacic or dodecanedioic acid) due to increased carbon flux through the pathway. In certain embodiments, the acyl-CoA ligase gene can be isolated from any suitable organism. Non-limiting examples of organisms that include, or can be used as donors for, acyl-CoA ligase enzymes include Candida, Saccharomyces, or Yarrowia.

Beta-oxidation - Enoyl-CoA Hydratase

An enoyl-CoA hydratase enzyme catalyzes the addition of a hydroxyl group and a proton to the unsaturated β-carbon on a fatty-acyl CoA and sometimes is encoded by the host organism and

sometimes can be added to generate an engineered organism. In certain embodiments, the enoyl-CoA hydratase activity is unchanged in a host or engineered organism. In some embodiments, the host enoyl-CoA hydratase activity can be increased by increasing the number of copies of an enoyl-CoA hydratase gene, by increasing the activity of a promoter that regulates transcription of an enoyl-CoA hydratase gene, or by increasing the number copies of the gene and by increasing the activity of a promoter that regulates transcription of the gene, thereby increasing the production of target product (e.g., sebacic or dodecanedioic acid) due to increased carbon flux through the pathway. In certain embodiments, the enoyl-CoA hydratase gene can be isolated from any suitable organism. Non-limiting examples of organisms that include, or can be used as donors for, enoyl-CoA hydratase enzymes include Candida, Saccharomyces, or Yarrowia.

Beta-oxidation - 3-Hydroxyacyl-CoA Dehydrogenase

3-hydroxyacyl-CoA dehydrogenase enzyme catalyzes the formation of a 3-ketoacyl-CoA by removal of a hydrogen from the newly formed hydroxyl group created by the activity of enoyl-CoA hydratase. In some embodiments, the activity is encoded by the host organism and sometimes can be added or increased to generate an engineered organism. In certain embodiments, the 3-hydroxyacyl-CoA activity is unchanged in a host or engineered organism. In some embodiments, the host 3-hydroxyacyl-CoA dehydrogenase activity can be increased by increasing the number of copies of a 3-hydroxyacyl-CoA dehydrogenase gene, by increasing the activity of a promoter that regulates transcription of a 3-hydroxyacyl-CoA dehydrogenase gene, or by increasing the number copies of the gene and by increasing the activity of a promoter that regulates transcription of the gene, thereby increasing production of target product (e.g., sebacic or dodecanedioic acid) due to increased carbon flux through the pathway. In certain embodiments, the 3-hydroxyacyl-CoA dehydrogenase gene can be isolated from any suitable organism. Non-limiting examples of organisms that include, or can be used as donors for, 3-hydroxyacyl-CoA dehydrogenase enzymes include Candida, Saccharomyces, or Yarrowia.

Beta-oxidation - Acetyl-CoA C-Acyltransferase

An Acetyl-CoA C-acyltransferase (e.g., beta-ketothiolase) enzyme catalyzes the formation of a fatty acyl-CoA shortened by 2 carbons by cleavage of the 3-ketoacyl-CoA by the thiol group of

another molecule of CoA. The thiol is inserted between C-2 and C-3, which yields an acetyl CoA molecule and an acyl CoA molecule that is two carbons shorter. An Acetyl-CoA C-acyltransferase sometimes is encoded by the host organism and sometimes can be added to generate an engineered organism. In certain embodiments, the acetyl-CoA C-acyltransferase activity is unchanged in a host or engineered organism. In some embodiments, the host acetyl-CoA C-acyltransferase activity can be increased by increasing the number of copies of an acetyl-CoA C-acyltransferase gene, or by increasing the activity of a promoter that regulates transcription of an acetyl-CoA C-acyltransferase gene, thereby increasing the production of target product (e.g., sebacic or dodecanedioic acid) due to increased carbon flux through the pathway. In certain embodiments, the acetyl-CoA C-acyltransferase gene can be isolated from any suitable organism. Non-limiting examples of organisms that include, or can be used as donors for, acetyl-CoA C-acyltransferase enzymes include Candida, Saccharomyces, or Yarrowia. One type of acetyl-CoA C-acyltransferase is an acetoacetyl CoA thiolase (e.g., "acoat").

Beta-oxidation - Enoyl CoA Isomerase

Feedstocks, such as fatty acid distillates and soapstocks can comprise unsaturated fatty acids, for example, such as oleic acid (C18:1), linoleic acid (C18:2), and linolenic acid (C18:3). In some embodiments, unsaturated fatty acids are converted to dicarboxylic acids that maintain the position and orientation of the double bonds. Cells can employ additional enzymes to allow the oxidation (e.g., beta oxidation) of these types of unsaturated fatty acids or diacids. In some instances, an enzyme enoyl-CoA isomerase (ECI) is required for the beta-oxidation of substrates with double bonds at odd numbered positions. In some instances, the enzyme dienoyl-CoA reductase (DCR) is required for the beta-oxidation of substrates with double bonds at even numbered positions).

Enoyl CoA Isomerase (ECI) can also be known as enoyl-CoA delta isomerase 1, dodecenoyl-CoA isomerase, 3,2 trans-enoyl-CoA isomerase, acetylene-allene isomerase, delta3, delta2-enoyl-CoA isomerase, dodecenoyl-CoA delta isomerase, and EC 5.3.3.8 (in human for example). Several alternatively spliced transcript variants are also known. ECI is a member of the hydratase/isomerase superfamily. ECI can be a key mitochondrial enzyme involved in beta-oxidation of unsaturated fatty acids. This enzyme can isomerize both 3-cis and 3-trans double bonds into the 2-trans form in a range of ECI enzymes from different species. ECI can catalyze

the transformation of 3-cis and 3-trans-enoyl-CoA esters arising during the stepwise degradation of cis-, mono-, and polyunsaturated fatty acids to the 2-trans-enoyl-CoA intermediates. ECI is present in many microorganisms and several species of yeast have at least two ECI enzymes. The percent amino acid sequence identities of two ECI enzymes from Candida viswanathii (Cv), two ECI enzymes from Candida tropicalis (Ct), one ECI enzyme from Yarrowia lipolytica (Yl) , and one ECI enzyme from Saccharomyces cerevisiae (Sc) are set forth in the TABLE 1 below. In TABLE 1, "p" refers to "polypeptide":

TABLE 1

	CtEci1p	CvEci1p	CtEci2p	CvEci2p	ScEci1p	YlEci1p
CtEci1p	100	92	58	57	39	19
CvEci1p		100	57	57	40	20
CtEci2p			100	84	36	18
CvEci2p				100	37	19
ScEci1p					100	18
YlEci1p						100

In some embodiments, ECI is a critical enzyme because of its activity and the normal position of double bonds in some feedstocks (e.g., soapstocks and fatty acid distillates). Many unsaturated fatty acids have a cis-A9 double bond. During the beta-oxidation of an 18-carbon diacid with a cis-A9 double bond, the double bond is encountered when it has been chain shortened to 12 carbons. At this stage the 12-carbon molecule can have a cis-A3 double bond that is not a substrate for an acyl-CoA oxidase. Therefore, in some instances, ECI is required to convert the cis-A3 double bond to a trans-A2 double bond. In some instances, the product of the ECI reaction is a substrate for the second step in beta-oxidation, and ECI can effectively bypass acyl-CoA oxidase in a particular round of beta-oxidation. In some instances, even if a strain comprises an acyl-CoA oxidase that is not active on feedstocks of less than or equal to C12 (i.e., 12 carbons), an active ECI can effect the shortening of one more rounds of beta-oxidation, which can produce a 10-carbon product for substrates with a cis-A9 double bond. Therefore, in some embodiments the ECI gene is disrupted (e.g., knocked out or deleted) in a yeast (e.g., in a Candida strain) to prevent chain shortening past a desired chain-length (e.g., in this instance, 12 carbons). In some embodiments, disrupting the expression (e.g. knocking out the expression) of an ECI gene can result in an increase in the production of a fatty dicarboxylic

acid comprising 10 to 18 carbons. In some embodiments, disrupting the expression (e.g. knocking out the expression) of an ECI gene can result in an increase in the production of a fatty dicarboxylic acid comprising 10, 12, 14, 16 or 18 carbons. In some embodiments, disrupting the expression of an enoyl CoA isomerase can increase the production of fatty dicarboxylic acid comprising 10, 12, 14, 16 or 18 carbons when using certain feedstocks (e.g., certain soapstocks or fatty acid distillates).

In some embodiments, an ECI knock out (i.e. eciA) strain is able to produce DDDA from oleic acid even in the presence of acyl-CoA oxidase with activity on substrates of chain-length less than 12 carbons.

In some embodiments, a 12 carbon dicarboxylic acid produced from fatty acid feedstocks comprising unsaturated fatty acids require hydrogenation to arrive at the fully saturated DDDA product.

Beta-oxidation - Dienoyl-CoA Reductase

Dienoyl CoA reductase (DCR, e.g., EC 1.3.1.34) is a peripheral enzyme that converts trans-2, cis-4 dienoyl-CoA substrates to trans-3-enoyl-CoA products (Gurvitz A, et al, (1997) JBC 272:22140-22147).

The trans-3-enoyl-CoA is then converted by the enzyme Enoyl-CoA Isomerase (ECI) to trans-2-enoyl-CoA which is then the substrate for the second enzyme in β-oxidation. The DCR reaction is often necessary for complete β-oxidation of fatty acids with double bonds at even numbered positions (e.g., linoleic acid (CI 8:2) and linolenic acid (CI 8:3)). Sometimes, the diacids produced from these fatty acids also require the activity of DCR for complete β-oxidation. Diacids are capable of being oxidized starting from either end (diterminal β-oxidation) and sometimes the enzymes required to deal with the double bonds are the same from either direction. This is because even-numbered diacids with double bonds at even-numbered positions maintain the even-numbered position from either end (similarly with double bonds at odd-numbered positions).

This reaction can be essential for complete beta-oxidation of fatty acids with double bonds at even numbered positions, such as linoleic acid and linolenic acid. Depending on the carbon chain length of the desired final di-acid product, it may be useful to either amplify or reduce the activity of one or more DCR enzymes in the host cell or engineered organism. For di-acid products that have a carbon chain length of eight or greater, it may be useful or desirable to reduce or eliminate one or all DCR enzymes in the host cell. For di-acid products that have a carbon chain length of less than eight carbons, it may be useful or desirable to amplify the activity of one or more DCR enzymes in the host cell or engineered organism.

TABLE 2 is a table of diacid products that may be produced from unsaturated fatty acids using a strain comprising a deletion in ECI and/or DCR. A Dcr- strain typically produces exclusively a C8:3 diacid. In some embodiments, a DCR polypeptide is not decreased, such as by disrupting a DCR-encoding polynucleotide, in a strain utilized to produce a diacid product (e.g., adipic acid sebacic acid, DDDA). In certain embodiments, DCR polypeptide production is increased (e.g., introducing additional copy numbers of an endogenous DCR-encoding polynucleotide; introducing one or more copies of a heterologous DCR-encoding polynucleotide) to produce adipic acid from polyunsaturated fatty acids, such as those prevalent in soybean or corn oil.

TABLE 2

		Diacids produced in Pox4+, Pox5+ background		
Fatty acid	**Carbons**	**Eci- product**	**Dcr- product**	**Eci-, Dcr- product**
Oleic	C18:1	C6:1	none	C6:1
Linoleic	C18:2	C10:2	C8:3	C10:3
Linolenic	C18:3	C12:3	C8:3	C12:3
Eicosenoic	C20:1	C6:1	none	C6:1
Erucic	C22:1	C6:1	none	C6:1

In yeast such as Candida tropicalis and Candida viswanathii, there are two DCR homologs, often referred to as DCR1 and DCR2. The yeast Saccharomyces cerevisiae includes one DCR enzyme, while the yeast Yarrowia lipolytica includes at least three DCR homologs, referred to herein as "DCR1", "DCR2", and "DCR3". The percent identities of some DCR enzymes are shown in TABLE 3 and TABLE 4 below where the prefix Ct is Candida tropicalis, Cv is Candida viswanathii, Sc is Saccharomyces cerevisiae and YI is Yarrowia lipolytica.

TABLE 3

	Ct XP_002545283	CvDcr1p	Ct XP_002550344	CvDcr2p	ScSps19p	Yl XP_501382	Yl XP_503937	Yl XP_502296
Ct XP_002545283	100	91	78	76	59	66	64	61
CvDcr1p		100	75	75	59	67	65	61
Ct XP_002550344			100	92	62	65	63	58
CvDcr2p				100	62	66	64	58
ScSps19p					100	60	56	54
Yl XP_501382						100	77	63
Yl XP_503937							100	60
Yl XP_502296								100

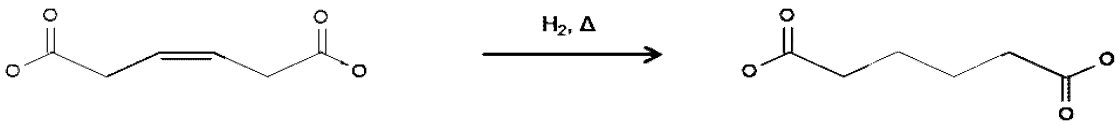
TABLE 4

	CtDcr1p	CvDcr1p	CtDcr2p	CvDcr2p	ScSps19p	YlDcr1p	YlDcr2p	YlDcr3p
CtDcr1p	100	91	78	76	59	66	64	61
CvDcr1p		100	75	75	59	67	65	61
CtDcr2p			100	92	62	65	63	58
CvDcr2p				100	62	66	64	58
ScSps19p					100	60	56	54
YlDcr1p						100	77	63
YlDcr2p							100	60
YlDcr3p								100

Hydrogenation

An unsaturated diacid sometimes is produced from a feedstock comprising an unsaturated fatty acid, and production of a fully saturated diacid in such situations often requires hydrogenation of the unsaturated diacid. For example, an unsaturated C6: 1 diacid generated from one or more long chain unsaturated fatty acids in an Eci-, poχ4Δ strain, can be converted to a fully saturated **C6:0** diacid by reducing the double bond by a suitable method. Non-limiting examples of hydrogenation methods include the use of a metallic chemical catalyst, non-metallic chemical catalyst, enzymatic catalyst, the like or combination thereof.

A non-limiting example of a hydrogenation reaction is shown below. Sometimes source hydrogen is provided from molecular hydrogen (e.g., in the case of chemical catalysis) and sometimes source hydrogen is provided from enzymatic cofactors, non-limiting examples of which include NADH, NADPH, FADH2, the like or combination thereof (e.g., in the case of enzymatic catalysis).



In some embodiments catalytic hydrogenation is carried out with a suitable metallic catalyst, non-limiting examples of which include platinum, palladium, rhodium, ruthenium, nickel, the like or combination thereof. Sometimes a catalyst is a homogenous catalyst and sometimes a catalyst is a heterogeneous catalyst. An elevated temperature and/or pressure can be employed to increase reaction rate. For example, an unsaturated diacid (e.g., cis, cis-muconic acid) can be hydrogenated and converted to adipic acid using a 10% Pt on carbon catalyst at 3400 kPa for 2.5 hours at ambient temperature (Niu et al., (2002) *Biotechnol.Prog.* 18:201-21 1). In some embodiments, catalytic hydrogenation can occur with nonmetallic catalysts such as frustrated Lewis pair compounds (Welch et al, (2006) Science 314:1 124-1 126).

In certain embodiments, enzymatic hydrogenation is conducted in vivo or in vitro with a suitable native or engineered enzyme that can catalyze a redox reaction with an unsaturated diacid or fatty acid as a substrate or a product. An enzyme can be utilized in vivo in some embodiments by increasing expression of a native enzyme or expressing a non-native enzyme capable of catalyzing a desired hydrogenation reaction in an organism that produces an unsaturated diacid precursor of a saturated diacid product. A lysate of an organism containing an enzyme capable of catalyzing a desired hydrogenation reaction, or a purified or isolated enzyme preparation, sometimes is utilized in an in vitro reaction. Non-limiting examples of a suitable native or engineered enzyme include acyl-CoA dehydrogenase (EC# 1.3.1.8), trans-2-enoyl-CoA reductase (EC# 1.3.1.44), stearoyl-CoA 9-desaturase (EC# 1.14.19.1), the like or combination thereof. In some embodiments, a desired reaction product (e.g., a saturated diacid) is produced by an enzyme operating in a forward or a reverse direction (e.g., a forward or reverse reaction).

Dienoyl CoA Isomerase

Dienoyl CoA Isomerase (DO, e.g., EC# 5.3.3, A3,5,A2,4-dienoyl-CoA isomerase, Δ3,5,Δ2,4-dienoyl-coenzyme A isomerase) is a peripheral β-oxidation enzyme that catalyzes the isomerization of a A3,5-dienoyl-CoA to a A2,4-dienoyl-CoA (e.g., see FIG. 33). This reaction is part of a minor β-oxidation pathway that occurs when the 3,2-enoyl-CoA isomerase (ECI) converts a A2,5-dienoyl-CoA to a A3,5-dienoyl-CoA. In order to fully oxidize this product DO

converts the A3,5-dienoyl-CoA to a A2,4-dienoyl-CoA, the latter of which is a substrate for the 2,4-dienoyl-CoA reductase (DCR). The product of the DCR reaction is a 3-enoyl-CoA, which is a substrate for ECI that converts it to a 2-enoyl-CoA that can be fully oxidized through β-oxidation.

A DCI enzyme was identified in mammalian systems. While an S. cerevisiae gene YOR180c has been identified as a DCI (Gurvitz et al, J. Biol. Chem. (1999) 274: 24514-24521) it also has been suggested that, while S. cerevisiae does contain DCI enzyme activity, such activity is not provided by the gene product of YOR180c (Ntamack et al., (2009) Biochim. Biophys. Acta 1791: 371-378).

The percent identity of the Yorl80c to the known DCI enzymes from rat and mouse is quite low (TABLE 5). Performing a a sequence alignment search of the Candida tropicalis and Candida viswanathii genomes for similarity to the mouse Dcilp identified sequences having greater identity than the S. cerevisiae Yorl80c. All of such sequences identified have C-terminal peroxisomal targeting sequences indicating that they are targeted to the peroxisomal compartment. The putative Candida sp. DCI enzymes identified may comprise dienoyl-CoA isomerase activity and also enoyl-CoA isomerase activity. They may also be peripheral β-oxidation enzymes of currently unknown function.

TABLE 5

	CtDci1p	CvDci1p	MmDci1p	RnDci1p	ScDci1p
CtDci1p		83	27	27	13
CvDci1p			28	28	13
MmDci1p				94	14
RnDci1p					14
ScDci1p					

The sequence of some putative DCI proteins are provided in SEQ. ID. NO: 3809-3812. In some embodiments, a polypeptide having DCI activity (e.g., in C. viswanathii), is reduced or amplified, depending upon the chain-length of a desired diacid product. For example, for adipic acid production, a DCI activity can be increased to improve productivity of unsaturated fatty acids in a host organism (e.g., DCI activity can be increased by inserting one or more copies of a polynucleotide encoding a polypeptide having DCI (e.g., inserting one or more copies of an endogenous or exogenous polynucleotide)). In some embodiments, for C8 and longer diacids, a DCI activity is reduced (e.g., by introducing a disruption, deletion or knockout

of (i) a polynucleotide that encodes a polypeptide having DCI activity, or (ii) a promoter operably linked to a polynucleotide that encodes a polypeptide having DCI activity).

Omega oxidation activities

The term "omega oxidation activity" refers to any of the activities in the omega oxidation pathway utilized to metabolize alkanes, fatty alcohols, fatty acids, dicarboxylic acids, or sugars. The activities utilized to metabolize fatty alcohols, fatty acids, or dicarboxylic acids include, but are not limited to, monooxygenase activity (e.g., cytochrome P450 activity), monooxygenase reductase activity (e.g., cytochrome P450 reductase activity), alcohol dehydrogenase activity (e.g., fatty alcohol dehydrogenase activity, or long-chain alcohol dehydrogenase activity), fatty alcohol oxidase activity, fatty aldehyde dehydrogenase activity, and thioesterase activity.

Omega Oxidation - Monooxygenases

A cytochrome P450 enzyme (e.g., monooxygenase activity) often catalyzes the insertion of one atom of oxygen into an organic substrate (RH) while the other oxygen atom is reduced to water. Insertion of the oxygen atom near the omega carbon of a substrate yields an alcohol derivative of the original starting substrate (e.g., yields a fatty alcohol). A cytochrome P450 sometimes is encoded by the host organism and sometimes can be added to generate an engineered organism.

In certain embodiments, the monooxygenase activity is unchanged in a host or engineered organism. In some embodiments, the host monooxygenase activity can be increased by increasing the number of copies of a cytochrome P450 gene, or by increasing the activity of a promoter that regulates transcription of a cytochrome P450 gene, thereby increasing the production of target product (e.g., sebacic or dodecanedioic acid) due to increased carbon flux through the pathway. In certain embodiments, the cytochrome P450 gene can be isolated from any suitable organism. Non-limiting examples of organisms that include, or can be used as donors for, cytochrome P450 enzymes include yeast (e.g., Candida, Saccharomyces, Debaryomyces, Meyerozyma, Lodderomyces, Scheffersomyces, Clavispora, Yarrowia, Pichia, Kluyveromyces, Eremothecium, Zygosaccharomyces, Lachancea, Nakaseomyces), animals (e.g., Homo, Rattus), bacteria (e.g., Escherichia, Pseudomonas, Bacillus), or plants (e.g., Arabidopsis, Nictotania, Cuphea).

Omega Oxidation - Monooxygenase Reductases

A cytochrome P450 reductase (e.g., monooxygenase reductase activity) catalyzes the reduction of the heme-thiolate moiety in cytochrome P450 by transferring an electron to the cytochrome P450. A cytochrome P450 reductase sometimes is encoded by the host organism and sometimes can be added to generate an engineered organism. In certain embodiments, the monooxygenase reductase activity is unchanged in a host or engineered organism. In some embodiments, the host monooxygenase reductase activity can be increased by increasing the number of copies of a cytochrome P450 reductase gene, or by increasing the activity of a promoter that regulates transcription of a cytochrome P450 reductase gene, thereby increasing the production of target product (e.g., sebacic or dodecanedioic acid) due to increased carbon flux through the pathway. In certain embodiments, the cytochrome P450 reductase gene can be isolated from any suitable organism. Non-limiting examples of organisms that include, or can be used as donors for, cytochrome P450 reductase enzymes include yeast (e.g., Candida, Saccharomyces, Debaryomyces, Meyerozyma, Lodderomyces, Scheffersomyces, Clavispora, Yarrowia, Pichia, Kluyveromyces, Eremothecium, Zygosaccharomyces, Lachancea, Nakaseomyces), animals (e.g., Homo, Rattus), bacteria (e.g., Escherichia, Pseudomonas, Bacillus), or plants (e.g., Arabidopsis, Nictotania, Cuphea).

Omega Oxidation - Alcohol Dehydrogenases

An alcohol dehydrogenase (e.g., fatty alcohol dehydrogenase, long-chain alcohol dehydrogenase) catalyzes the removal of a hydrogen from an alcohol to yield an aldehyde or ketone and a hydrogen atom and NADH, in the endoplasmic reticulum of a cell. An alcohol dehydrogenase sometimes is encoded by the host organism and sometimes can be added to generate an engineered organism. In certain embodiments, the alcohol dehydrogenase activity is unchanged in a host or engineered organism. In some embodiments, the host alcohol dehydrogenase activity can be increased by increasing the number of copies of an alcohol dehydrogenase gene, or by increasing the activity of a promoter that regulates transcription of an alcohol dehydrogenase gene, thereby increasing the production of target product (e.g., sebacic or dodecanedioic acid) due to increased carbon flux through the pathway. In certain embodiments, the alcohol dehydrogenase gene can be isolated from any suitable organism. Non-limiting examples of organisms that include, or can be used as donors for, alcohol dehydrogenase enzymes include yeast (e.g., Candida, Saccharomyces, Debaryomyces,

Meyerozyma, Lodderomyces, Scheffersomyces, Clavispora, Yarrowia, Pichia, Kluyveromyces, Eremothecium, Zygosaccharomyces, Lachancea, Nakaseomyces), animals (e.g., Homo, Rattus), bacteria (e.g., Escherichia, Pseudomonas, Bacillus), or plants (e.g., Arabidopsis, Nictotania, Cuphea).

Omega Oxidation - Fatty Alcohol Oxidases

A fatty alcohol oxidase (e.g., long-chain alcohol oxidase, EC 1.1.3.20) enzyme catalyzes the addition of oxygen to two molecules of a long-chain alcohol to yield 2 long chain aldehydes and 2 molecules of water, in the peroxisome of a cell. A fatty alcohol oxidase sometimes is encoded by the host organism and sometimes can be added to generate an engineered organism. In certain embodiments, the fatty alcohol oxidase activity is unchanged in a host or engineered organism. In some embodiments, the host fatty alcohol oxidase activity can be increased by increasing the number of copies of a fatty alcohol oxidase gene, or by increasing the activity of a promoter that regulates transcription of a fatty alcohol oxidase gene, thereby increasing the production of target product (e.g., sebacic or dodecanedioic acid) due to increased carbon flux through the pathway. In certain embodiments, the fatty alcohol oxidase gene can be isolated from any suitable organism. Non-limiting examples of organisms that include, or can be used as donors for, fatty alcohol oxidase enzymes include yeast (e.g., Candida, Saccharomyces, Debaryomyces, Meyerozyma, Lodderomyces, Scheffersomyces, Clavispora, Yarrowia, Pichia, Kluyveromyces, Eremothecium, Zygosaccharomyces, Lachancea, Nakaseomyces), animals (e.g., Homo, Rattus), bacteria (e.g., Escherichia, Pseudomonas, Bacillus), or plants (e.g., Arabidopsis, Nictotania, Cuphea).

Omega Oxidation - Aldehyde Dehydrogenases

A fatty aldehyde dehydrogenase (e.g., long chain aldehyde dehydrogenase) enzyme catalyzes the oxidation of long chain aldehydes to a long chain carboxylic acid, NADH and H^+. A fatty aldehyde dehydrogenase sometimes is encoded by the host organism and sometimes can be added to generate an engineered organism. In certain embodiments, the fatty aldehyde dehydrogenase activity is unchanged in a host or engineered organism. In some embodiments, the host fatty aldehyde dehydrogenase activity can be increased by increasing the number of copies of a fatty aldehyde dehydrogenase gene, or by increasing the activity of a promoter that regulates transcription of a fatty aldehyde dehydrogenase gene, thereby increasing the

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 649 of 3077

production of target product (e.g., sebacic or dodecanedioic acid) due to increased carbon flux through the pathway. In certain embodiments, the fatty aldehyde dehydrogenase gene can be isolated from any suitable organism. Non-limiting examples of organisms that include, or can be used as donors for, fatty aldehyde dehydrogenase enzymes include yeast (e.g., Candida, Saccharomyces, Debaryomyces, Meyerozyma, Lodderomyces, Scheffersomyces, Clavispora, Yarrowia, Pichia, Kluyveromyces, Eremothecium, Zygosaccharomyces, Lachancea, Nakaseomyces), animals (e.g., Homo, Rattus), bacteria (e.g., Escherichia, Pseudomonas, Bacillus), or plants (e.g., Arabidopsis, Nictotania, Cuphea).

Omega Oxidation - Thioesterases

A thioesterase enzyme (e.g., acyl-CoA thioesterase activity, acyl-ACP thioesterase activity) catalyzes the removal of Coenzyme A or acyl carrier protein (e.g., ACP) from a fatty acid including acyl-CoA or acyl carrier protein (e.g., esterified fatty acid) to yield a fatty acid and an alcohol. The reaction occurs in the presence of water and Coenzyme A or acyl carrier protein is specifically removed at a thiol group. A thioesterase sometimes is encoded by the host organism and sometimes can be added to generate an engineered organism. In certain embodiments, the thioesterase activity is unchanged in a host or engineered organism. In some embodiments, the host thioesterase activity can be increased by increasing the number of copies of a thioesterase gene, or by increasing the activity of a promoter that regulates transcription of a thioesterase gene, thereby increasing the production of target product (e.g., sebacic or dodecanedioic acid) due to increased carbon flux through the pathway. In certain embodiments, a thioesterase gene can be isolated from any suitable organism. Non-limiting examples of organisms that include, or can be used as donors for, thioesterase enzymes include yeast (e.g., Candida, Saccharomyces, Debaryomyces, Meyerozyma, Lodderomyces, Scheffersomyces, Clavispora, Yarrowia, Pichia, Kluyveromyces, Eremothecium, Zygosaccharomyces, Lachancea, Nakaseomyces), animals (e.g., Homo, Rattus), bacteria (e.g., Escherichia, Pseudomonas, Bacillus), or plants (e.g., Arabidopsis, Nictotania, Cuphea).

Engineered Pathways

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 650 of 3077

Figures 1-8 depict embodiments of biological pathways for making sebacic acid and dodecanedioic acid, using various alkanes, fatty acids, fatty alcohols or combinations thereof. Any suitable alkane, fatty acid, fatty alcohol, plant based oil, seed based oil, non-petroleum derived soap stock or the like can be used as the feedstock for the organism (e.g., dodecane, methyl laurate, lauric acid, carbon sources having 10 or greater carbons (e.g. for sebacic acid production) or carbon sources having 12 or greater carbons (e.g. for dodecanedioic acid production). In some embodiments, carbon sources with greater than 12 carbons can be metabolized using naturally occurring and/or engineered pathways to yield molecules that can be further metabolized using the beta oxidation pathway shown in the lower portion of Figures 5-8. In some embodiments, the activities in the pathways depicted in Figures 1-8 can be engineered, as described herein, to enhance metabolism and target product formation.

In certain embodiments, one or more activities in one or more metabolic pathways can be engineered to increase carbon flux through the engineered pathways to produce a desired product (e.g., sebacic or dodecanedioic acid). The engineered activities can be chosen to allow increased production of metabolic intermediates that can be utilized in one or more other engineered pathways to achieve increased production of a desired product with respect to the unmodified host organism. The engineered activities also can be chosen to allow decreased activity of enzymes that reduce production of a desired intermediate or end product (e.g., reverse activities). This "carbon flux management" can be optimized for any chosen feedstock, by engineering the appropriate activities in the appropriate pathways. Non-limiting examples are given herein using pure alkanes (e.g., single chain length alkanes, dodecane or example), mixed chain-length alkanes, long-chain alkanes, pure fatty acids (e.g., single chain length fatty acids, capric acid for example) and mixed chain length fatty acids (see Figures 1-8). The process of "carbon flux management" through engineered pathways produces a dicarboxylic acid (e.g. sebacic acid or dodecanedioic acid) at a level and rate closer to the calculated maximum theoretical yield for any given feedstock, in certain embodiments. The terms "theoretical yield" or "maximum theoretical yield" as used herein refer to the yield of product of a chemical or biological reaction that can be formed if the reaction went to completion. Theoretical yield is based on the stoichiometry of the reaction and ideal conditions in which starting material is completely consumed, undesired side reactions do not occur, the reverse reaction does not occur, and there no losses in the work-up procedure.

A microorganism may be modified and engineered to include or regulate one or more activities in a fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid) pathway. The term "activity" as used herein refers to the functioning of a microorganism's natural or engineered biological pathways to yield various products including a fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid) and its precursors. A fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid) producing activity can be provided by any non-mammalian source in certain embodiments. Such sources include, without limitation, eukaryotes such as yeast and fungi and prokaryotes such as bacteria. In some embodiments, a reverse activity in a pathway described herein can be altered (e.g., disrupted, reduced) to increase carbon flux through a beta oxidation pathway, an omega oxidation pathway, or a beta oxidation and omega oxidation pathway, towards the production of target product (e.g., sebacic or dodecanedioic acid). In some embodiments, a genetic modification disrupts an activity in the beta oxidation pathway, or disrupts a polynucleotide that encodes a polypeptide that carries out a forward reaction in the beta oxidation pathway, which renders beta oxidation activity undetectable. The term "undetectable" as used herein refers to an amount of an analyte that is below the limits of detection, using detection methods or assays known (e.g., described herein). In certain embodiments, the genetic modification partially reduces beta oxidation activity. The term "partially reduces beta oxidation activity" as used here refers to a level of activity in an engineered organism that is lower than the level of activity found in the host or starting organism.

In some embodiments, a beta-oxidation activity can be modified to alter the catalytic specificity of the chosen activity. In certain embodiments, an acyl-CoA oxidase activity can be altered by modifying a catalytic domain associated with carbon chain length preference and/or specificity. In some embodiments, the altered catalytic specificity can be found by screening naturally occurring variant or mutant populations of a host organism. In certain embodiments, the altered catalytic can be generated by various mutagenesis techniques in conjunction with selection and/or screening for the desired activity. In some embodiments, the altered catalytic activity can be generated by generating chimeric acyl-CoA oxidases using a mix and match approach, followed by selection and/or screening for the desired catalytic activity. Examples of

experiments performed to generate acyl-CoA oxidases with altered catalytic activity are described herein.

An activity within an engineered microorganism provided herein can include one or more (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 or all) of the following activities: 6-oxohexanoic acid dehydrogenase activity; 6-hydroxyhexanoic acid dehydrogenase activity; cytochrome P450 activity; cytochrome P450 reductase activity; fatty alcohol oxidase activity; acyl-CoA ligase activity, acyl-CoA oxidase activity; enoyl-CoA hydratase activity, 3-hydroxyacyl-CoA dehydrogenase activity, fatty acid synthase activity, lipase activity, acetyl-CoA carboxylase activity, acyltransferase activity (diacylglycerol acyl transferase, lecithin-cholesterol acyltransferase, phospholipid:diacylglycerol acyltransferase) and thioesterase activity (e.g., acyl-CoA hydrolase, acyl-CoA thioesterase, acyl-ACP thioesterase, acetyl-CoA C-acyltransferase, beta-ketothiolase, and the like). In certain embodiments, one or more (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or all) of the foregoing activities is altered by way of a genetic modification. In some embodiments, one or more (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 or all) of the foregoing activities is altered by way of (i) adding a heterologous polynucleotide that encodes a polypeptide having the activity, and/or (ii) altering or adding a regulatory sequence that regulates the expression of a polypeptide having the activity. In certain embodiments, one or more of the foregoing activities is altered by way of (i) disrupting an endogenous polynucleotide that encodes a polypeptide having the activity (e.g., insertional mutagenesis), (ii) deleting a regulatory sequence that regulates the expression of a polypeptide having the activity, and/or (iii) deleting the coding sequence that encodes a polypeptide having the activity (e.g., knock out mutagenesis).

The term "omega hydroxyl fatty acid dehydrogenase activity" as used herein refers to conversion of an omega hydroxyl fatty acid to an omega oxo fatty acid. The omega hydroxyl fatty acid dehydrogenase activity can be provided by a polypeptide. In some embodiments, the polypeptide is encoded by a heterologous nucleotide sequence introduced to a host microorganism. In certain embodiments, an endogenous polypeptide having the omega hydroxyl fatty acid dehydrogenase activity is identified in the host microorganism, and the host microorganism is genetically altered to increase the amount of the polypeptide produced (e.g., a heterologous promoter is introduced in operable linkage with a polynucleotide that encodes the polypeptide; the copy number of a polynucleotide that encodes the polypeptide is increased (e.g., by introducing a plasmid that includes the polynucleotide)). Nucleic acid sequences

conferring omega hydroxyl fatty acid dehydrogenase activity can be obtained from a number of sources, including Actinobacter, Norcardia, Pseudomonas and Xanthobacter bacteria. Examples of an amino acid sequence of a polypeptide having omega hydroxyl fatty acid dehydrogenase activity and a nucleotide sequence of a polynucleotide that encodes the polypeptide, are presented herein. Presence, absence or amount of omega hydroxyl fatty acid dehydrogenase activity can be detected by any suitable method known in the art. In some embodiments, omega hydroxyl fatty acid dehydrogenase activity is not altered in a host microorganism, and in certain embodiments, the activity is added or increased in the engineered microorganism relative to the host microorganism.

The term "monooxygenase activity" as used herein refers to inserting one atom of oxygen from $C^{"2}$ into an organic substrate (RH) and reducing the other oxygen atom to water. In some embodiments, monooxygenase activity refers to incorporation of an oxygen atom onto a six-carbon organic substrate. In certain embodiments, monooxygenase activity refers to conversion of hexanoate to 6-hydroxyhexanoic acid. Monooxygenase activity can be provided by any suitable polypeptide, such as a cytochrome P450 polypeptide (hereafter "CYP450") in certain embodiments. Nucleic acid sequences conferring CYP450 activity can be obtained from a number of sources, including Bacillus megaterium and may be induced in organisms including but not limited to Candida tropicalis, Yarrowia lipolytica, Aspergillus nidulans, and Aspergillus parasiticus. Examples of oligonucleotide sequences utilized to isolate a polynucleotide sequence encoding a polypeptide having CYP450 activity (e.g., CYP52A12 polynucleotide, a CYP52A13 polynucleotide, a CYP52A14 polynucleotide, a CYP52A15 polynucleotide, a CYP52A16 polynucleotide, a CYP52A17 polynucleotide, a CYP52A18 polynucleotide, a CYP52A19 polynucleotide, a CYP52A20 polynucleotide, a CYP52D2 polynucleotide, and/or a BM3 polynucleotide) are presented herein. In some embodiments, monooxygenase activity is not altered in a host microorganism, and in certain embodiments, the activity is added or increased in the engineered microorganism relative to the host microorganism. In some embodiments, the altered monooxygenase activity is an endogenous activity, and in certain embodiments, the altered monooxygenase activity is an exogenous activity. In some embodiments, the exogenous activity is a single polypeptide with both monooxygenase and monooxygenase reductase activities (e.g., B. megaterium cytochrome P450:NADPH P450 reductase).

Presence, absence or amount of cytochrome P450 activity can be detected by any suitable method known in the art. For example, detection can be performed by assaying a reaction containing cytochrome P450 (CYP52A family) and NADPH - cytochrome P450 reductase (see Appl. Environ. Microbiol. 69: 5983 and 5992). Briefly, cells are grown under standard conditions and harvested for production of microsomes, which are used to detect CYP activity. Microsomes are prepared by lysing cells in Tris-buffered sucrose (lOmM Tris-HCl pH 7.5, ImM EDTA, 0.25M sucrose). Differential centrifugation is performed first at 25,000xg then at 100,000xg to pellet cell debris then microsomes, respectively. The microsome pellet is resuspended in 0.1M phosphate buffer (pH 7.5), ImM EDTA to a final concentration of approximately lOmg protein/mL. A reaction mixture containing approximately 0.3mg microsomes, O.lmM sodium hexanoate, 0.7mM NADPH, 50mM Tris-HCl pH 7.5 in lmL is initiated by the addition of NADPH and incubated at 37°C for 10 minutes. The reaction is terminated by addition of 0.25mL 5M HC1 and 0.25mL 2.5ug/mL 10-hydroxydecanoic acid is added as an internal standard (3.3 nmol). The mixture is extracted with 4.5mL diethyl ether under NaCl-saturated conditions. The organic phase is transferred to a new tube and evaporated to dryness. The residue is dissolved in acetonitrile containing lOmM 3-bromomethyl-7-methoxy-l,4-benzoxazin-2-one (BrMB) and O.lmL of 15mg/mL 18-crown-6 in acetonitrile saturated with K_2CO_3. The solution is incubated at 40°C for 30 minutes before addition of 0.05mL 2% acetic acid. The fluorescently labeled omega-hydroxy fatty acids are resolved via HPLC with detection at 430nm and excitation at 355nm (Yamada et al., 1991, Anal.Biochem. 199: 132-136). Optionally, specifically induced CYP gene(s) may be detected by Northern blotting and/or quantitative RT-PCR. (Craft et al., 2003, App.Environ.Micro. 69: 5983-5991).

The term "monooxygenase reductase activity" as used herein refers to the transfer of an electron from NAD(P)H, FMN, or FAD by way of an electron transfer chain, reducing the ferric heme iron of cytochrome P450 to the ferrous state. The term "monooxygenase reductase activity" as used herein also can refer to the transfer of a second electron via the electron transport system, reducing a dioxygen adduct to a negatively charged peroxo group. In some embodiments, a monooxygenase activity can donate electrons from the two-electron donor NAD(P)H to the heme of cytochrome P450 (e.g., monooxygenase activity) in a coupled two-step reaction in which NAD(P)H can bind to the NAD(P)H-binding domain of the polypeptide having the monooxygenase reductase activity and electrons are shuttled from NAD(P)H through FAD and FMN to the heme of the monooxygenase activity, thereby regenerating an active

monooxygenase activity (e.g., cytochrome P450). Monooxygenase reductase activity can be provided by any suitable polypeptide, such as a cytochrome P450 reductase polypeptide (hereafter "CPR") in certain embodiments. Nucleic acid sequences conferring CPR activity can be obtained from and/or induced in a number of sources, including but not limited to Bacillus megaterium, Candida tropicalis, Yarrowia lipolytica, Aspergillus nidulans, and Aspergillus parasiticus. Examples of oligonucleotide sequences utilized to isolate a polynucleotide sequence encoding a polypeptide having CPR activity are presented herein. In some embodiments, monooxygenase reductase activity is not altered in a host microorganism, and in certain embodiments, the activity is added or increased in the engineered microorganism relative to the host microorganism. In some embodiments, the altered monooxygenase reductase activity is an endogenous activity, and in certain embodiments, the altered monooxygenase reductase activity is an exogenous activity. In some embodiments, the exogenous activity is a single polypeptide with both monooxygenase and monooxygenase reductase activities (e.g., B. megaterium cytochrome P450:NADPH P450 reductase).

Presence, absence or amount of CPR activity can be detected by any suitable method known in the art. For example, an engineered microorganism having an increased number of genes encoding a CPR activity, relative to the host microorganism, could be detected using quantitative nucleic acid detection methods (e.g., southern blotting, PCR, primer extension, the like and combinations thereof). An engineered microorganism having increased expression of genes encoding a CPR activity, relative to the host microorganism, could be detected using quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof). Alternately, an enzymatic assay can be used to detect Cytochrome P450 reductase activity, where the enzyme activity alters the optical absorbance at 550 nanometers of a substrate solution (Masters, B.S.S., Williams, C.H., Kamin, H. (1967) Methods in Enzymology, X, 565-573).

Acyl-CoA Oxidases

The term "acyl-CoA oxidase activity" as used herein refers to the oxidation of a long chain fatty-acyl-CoA to a trans-2,3-dehydroacyl-CoA fatty alcohol. In some embodiments, the acyl-CoA activity is from a peroxisome. In certain embodiments, the acyl-CoA oxidase activity is a peroxisomal acyl-CoA oxidase (POX) activity, carried out by a POX polypeptide. In some embodiments the acyl-CoA oxidase activity is encoded by the host organism and sometimes

can be altered to generate an engineered organism. Acyl-CoA oxidase activity is encoded by the POX4 and POX5 genes of Candida strain ATCC20336. In certain embodiments, endogenous acyl-CoA oxidase activity can be increased. In some embodiments, acyl-CoA oxidase activity of the POX4 polypeptide or the POX5 polypeptide can be altered independently of each other (e.g., increase activity of POX4 alone, POX5 alone, increase one and disrupt the other, and the like). Increasing the activity of one POX activity, while disrupting the activity of another POX activity, may alter the specific activity of acyl-CoA oxidase with respect to carbon chain length, while maintaining or increasing overall flux through the beta oxidation pathway, in certain embodiments.

In certain embodiments, host acyl-CoA oxidase activity of one of the POX genes can be increased by genetically altering (e.g., increasing) the amount of the polypeptide produced (e.g., a strongly transcribed or constitutively expressed heterologous promoter is introduced in operable linkage with a polynucleotide that encodes the polypeptide; the copy number of a polynucleotide that encodes the polypeptide is increased (e.g., by introducing a plasmid that includes the polynucleotide, integration of additional copies in the host genome)). In some embodiments, the host acyl-CoA oxidase activity can be decreased by disruption (e.g., knockout, insertion mutagenesis, the like and combinations thereof) of an acyl-CoA oxidase gene, or by decreasing the activity of the promoter (e.g., addition of repressor sequences to the promoter or 5'UTR) which transcribes an acyl-CoA oxidase gene.

As noted above, disruption of nucleotide sequences encoding POX4, POX 5, or POX4 and POX5 sometimes can alter pathway efficiency, specificity and/or specific activity with respect to metabolism of carbon chains of different lengths (e.g., carbon chains including fatty alcohols, fatty acids, paraffins, dicarboxylic acids of between about 1 and about 60 carbons in length). In some embodiments, the nucleotide sequence of POX4, POX5, or POX4 and POX5 is disrupted with a URA3 nucleotide sequence encoding a selectable marker, and introduced to a host microorganism, thereby generating an engineered organism deficient in POX4, POX5 or POX4 and POX5 activity. Nucleic acid sequences encoding POX4 and POX5 can be obtained from a number of sources, including Candida tropicalis, for example. Examples of POX4 and POX5 amino acid sequences and nucleotide sequences of polynucleotides that encode the polypeptides, are presented herein. Described in the examples are experiments conducted to amplify the activity encoded by the POX5 gene.

Also as noted above, catalytic specificity of acyl-CoA oxidases (e.g., POX4, POX5) can be altered by a variety of methods. Altering the binding and/or catalytic specificity of acyl-CoA oxidases may prove advantageous for generating novel acyl-CoA oxidases with altered chain length recognition, altered chain length catalytic activity, and/or generation of an acyl-CoA oxidase activity with a narrow or specific chain length specificity, thereby allowing further increases in pathway efficiency, specificity and/or specific activity with respect to metabolism of carbon chains of different lengths or metabolism of carbon chain distributions found in a particular chosen feedstock. In some embodiments the altered acyl-CoA oxidase sequences are identified and/or generated by; (i) screening naturally occurring variant populations; (ii) mutagenesis of endogenous sequences; (iii) introduction of heterologous sequences having a desired specificity; (iv) generation of chimeric sequences having a portion of the coding sequence from one polynucleotide source (e.g., gene, organism) and a portion of the coding sequence from another source and/or (v) intelligent design using nucleotide sequences and three dimensional structure analysis from an acyl-CoA oxidase having a desired specificity to remodel an endogenous acyl-CoA oxidase, thereby generating a novel specificity enzyme. In some embodiments a chimeric acyl-CoA oxidase sequence can have polynucleotide sequence contributions from two or more sources. In some embodiments, a chimeric acyl-CoA oxidase sequence comprises a portion of the coding sequences from an endogenous polynucleotide and a portion of the coding sequence from a heterologous polynucleotide. Described in the examples are methods utilized to identify and/or generate acyl-CoA oxidases with novel catalytic and binding specificities.

Introduction of heterologous Acyl CoA Oxidase sequences having a desired specificity

Thousands of Acyl CoA Oxidases and Acyl CoA-like Oxidases have been cloned, sequenced and isolated from a variety of organisms (SEQ ID NO. 51 through SEQ ID NO. 3673 and SEQ ID NO. 3810 through SEQ ID NO. 3882). Many of these enzymes have reported catalytic activity with selective substrate specificity. For example, some Acyl CoA Oxidases (e.g., Pox5p from a Candida strain) display optimal activity on substrates of 12 to 18 carbons (FIG. 21). In some embodiments, an organism (e.g., a yeast) or a genetically modified organism (e.g., a genetically modified yeast, e.g., a yeast in which β-oxidation activity is blocked) is engineered to express a heterologous Acyl-CoA Oxidase with selective substrate specificity. In some embodiments, an organism (e.g., a yeast) or a genetically modified organism (e.g., a genetically modified yeast, e.g., a yeast in which β-oxidation activity is blocked) is engineered to express

an Acyl-CoA Oxidase or Acyl CoA-like Oxidase selected from SEQ ID NO. 51 to SEQ ID NO. 3673. In some embodiments, an organism (e.g., a yeast) or a genetically modified organism (e.g., a genetically modified yeast, e.g., a yeast in which β-oxidation activity is blocked) is engineered to express an Acyl-CoA Oxidase or Acyl CoA-like Oxidase selected from SEQ ID NO. 3810 through SEQ ID NO. 3882.

Presence, absence or amount of acyl-CoA oxidase activity can be detected by any suitable method known in the art. For example, using enzymatic assays as described in Shimizu et al, 1979, and as described herein in the Examples, POX4, POX5 and other acyl-CoA oxidase activities can be assessed. Alternatively, nucleic acid sequences representing native and/or disrupted POX4 and POX5 sequences also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered organism exhibits decreased RNA and/or polypeptide levels as compared to the host organism.

Genetic Modification of Acyl-CoA Oxidases

A rate-limiting step for β-oxidation is the first step in the pathway carried out by the enzyme acyl CoA oxidase. Different Acyl-CoA oxidases can display different chain-length substrate specificities. Some acyl CoA oxidases display broad chain-length specificity and can accept any fatty acyl CoA (or diacyl-CoA) as a substrate. However, some acyl CoA oxidases can display narrow chain-length specificity.

For example the Pox5 enzyme from Candida strain ATCC20336 displays a decrease in activity on substrates below CIO (FIG. 21) and has low activity on C6 and C8 substrates. In a cell with Pox5 as the only functional acyl CoA oxidase, long chain fatty acyl-CoA or diacyl-CoA substrates can be shortened to about 8 carbons and do not typically enter another cycle of β-oxidation. The shorter substrates (e.g., a C8 fatty dicarboxylic acid) are not typically recognized as a substrate by Pox5, the CoA is removed by peroxisomal thioesterases and the fatty dicarboxylic acid (e.g., an α,ω - dicarboxylic acid) product is secreted from the cell. In this embodiment, the acyl CoA oxidase chain-length substrate specificity effectively controls the chain length of a diacid produced.

In some embodiments, a β-oxidation pathway in a yeast is active and includes a genetically modified acyl CoA oxidase. In some embodiments, an acyl CoA oxidase is genetically modified to prevent complete oxidation of fatty acyl-CoA or diacyl-CoA substrates. Genetic modification of an acyl CoA oxidase can increase the production yield of a desired fatty acid or fatty dicarboxylic acid product. Therefore, in some embodiments, metabolic degradation of a fatty acid of a specified chain length (e.g., the chain length of a desired or target fatty acid or fatty dicarboxylic acid product) is reduced significantly, when an acyl CoA oxidase is genetically modified. In some embodiments, metabolic degradation of a fatty dicarboxylic acid product (e.g., DDDA) by beta-oxidation is reduced significantly, when an acyl CoA oxidase is genetically modified. This can be accomplished by modifying the substrate specificity of an acyl CoA oxidase such that the enzyme has low activity (e.g., enzymatic activity) with chain lengths less than that of a desired product.

In some embodiments, the substrate specificity of an acyl CoA oxidase is modified such that the enzyme has low activity for aliphatic molecules with chain lengths less than C24 (i.e., 24 carbons). In some embodiments, the substrate specificity of an acyl CoA oxidase is modified such that the enzyme has very low activity with chain lengths less than 24, 22, 20, 18, 16, 14, 12, 10, 8, 6 or 4 carbons. In some embodiments, the substrate specificity of an acyl CoA oxidase is modified such that the enzyme has very low activity with chain lengths less than 18, 16, 14, 12, 10 or 8 carbons. In some embodiments, the substrate specificity of an acyl CoA oxidase is modified such that the enzyme has very low activity with chain lengths less than C12. In some embodiments, the substrate specificity of an acyl CoA oxidase is modified such that the enzyme has very low activity with chain lengths less than CIO.

In some embodiments, genes encoding a genetically modified acyl CoA oxidase are engineered and expressed in a suitable organism (e.g., a bacteria (e.g., E.coli) or a yeast) to test the substrate specificity of the modified enzyme in vitro. In some embodiments, genes encoding a genetically modified acyl CoA oxidase are engineered and expressed in a suitable yeast and the substrate specificity is tested. In some embodiments, yeast that express a modified acyl CoA oxidase are tested for production of the desired fatty acid or fatty dicarboxylic acid product. A modified acyl CoA oxidase can be generated in any suitable manner, non limiting examples of which are provided hereafter.

Random Mutagenesis of Acyl-CoA Oxidase

A library of genetically modified acyl CoA oxidases can be generated using several methods known in the art (e.g., site-directed mutagenesis). Genetically modified acyl CoA oxidase genes can then be transformed into a β-oxidation blocked strain of a suitable yeast strain (e.g., Candida spp. (e.g., Candida viswanathii or Candida tropicalis)). In some embodiments, a genetically modified acyl CoA oxidase is expressed under the control of the POX4 promoter or another strong constitutive or inducible promoter in a poχ4Δ/ poχ4Δ poχ5Δ/poχ5Δ (e.g., an organism that lacks endogenous acyl CoA oxidase activity) background. In some embodiments, the genetically modified acyl CoA oxidase is expressed under the control of endogenous promoter. In some embodiments, the genetically modified acyl CoA oxidase is expressed under the control of a heterologous promoter. The transformants can be selected by growth in a fatty acid or methyl-derivate fatty acid containing fatty acids with two more carbons than the diacid product of interest. For example, for adipic acid, the transformants can be grown in caprylic acid or methyl-caprylate. For example, for dodecanedioic acid, the transformants can be grown in tetradecanedioic acid. The group of transformants can then be moved to a medium with a carbon source of a fatty acid of interest (for example dodecanedioic acid) in the presence of an agent that kills growing cells (e.g., Nystatin) and cells that cannot metabolize the carbon source (e.g., dodecanedioic acid in this example) can be selected. The resulting modified strains can then be further characterized for acyl CoA oxidase activity. This method can be used to select for any modified acyl CoA oxidase (e.g., those listed and/or described in the TABLES herein). In addition, this method can be used to select for any heterologous acyl CoA oxidase (e.g., those listed in SEQ ID NO. 5 1 through 3273 and SEQ ID NO. 3728 through 3810) expressed in a suitable organism.

Rational Mutagenesis of Acyl-CoA Oxidase

Structural and sequence information and experimental data can be combined to determine specific mutations to be tested in a acyl-CoA oxidase for altered specificity. For example, primary sequences of acyl-CoA oxidases tested can be compared and correlated with substrate specificity. Based on such an analysis, single amino-acids, small numbers of contiguous amino acids and/or domains can be proposed for providing a desired substrate specificity. Those amino acids positions can be targeted for specific or random mutations for improve specificity.

Acyl CoA oxidase structure also can be modeled against a known tertiary structure using modeling methods known in the art. The models can be used to propose amino acids and regions pertaining to substrate selectivity. For example, biochemical, structure and sequence data suggest that the N-terminus of acyl CoA oxidases often, in part, determines substrate specificity. Mutations or region replacements can be introduced based on such analyses and the specificity of the new acyl CoA oxidase tested as described before. The resulting information can be used to go back to the models to postulate new potential mutations. As for random mutagenesis, any suitable acyl CoA oxidase can be modified to alter substrate specificity (e.g., those listed in SEQ ID NO. 5 1 through 3273 and SEQ ID NO. 3728 through 3810)

The term "acyl CoA oxidase activity" as used herein refers to the enzymatic activity (e.g., catalytic activity) of a acyl CoA oxidase. An acyl CoA oxidase can catalyze the following chemical reaction:

$$\text{acyl-CoA} + 0_2 \leftrightarrow \text{trans-2,3-dehydroacyl-CoA} + H_20_2$$

In some embodiments, acyl CoA oxidase activity refers to oxidation of a long chain fatty-acyl-CoA to a trans-2,3-dehydroacyl-CoA fatty alcohol. In some embodiments, acyl CoA oxidase activity refers to its enzyme activity (or lack thereof) on a selective set of substrates. The activity of an acyl CoA oxidase can be affected by its ability to bind a substrate, oxidize a substrate and/or release a product. In some embodiments, an acyl CoA oxidase is active in one compartment of a cell and not in another compartment of the cell. In some embodiments, the acyl CoA oxidase activity is from a peroxisome. In certain embodiments, the acyl CoA oxidase activity is a peroxisomal acyl CoA oxidase (POX) activity, carried out by a POX polypeptide. In some embodiments the acyl CoA oxidase activity is encoded by the host organism and sometimes can be altered to generate an engineered organism. Acyl-CoA oxidase activity can be encoded by the POX4 and POX5 genes of Candida spp.. In certain embodiments, endogenous acyl CoA oxidase activity can be increased. In some embodiments, acyl CoA oxidases in an organism, containing one or more acyl CoA oxidases, can be independently modified (e.g., one or more acyl CoA oxidases can be modified). In some embodiments, acyl CoA oxidase activity of a POX4 polypeptide or a POX5 polypeptide can be altered independently of each other (e.g., increase activity of POX4 alone, POX5 alone, increase one and disrupt the other, and the like). Increasing the activity of one POX activity, while disrupting the activity of another POX activity, may alter the specific activity of acyl CoA

oxidase with respect to carbon chain length, while maintaining or increasing overall flux through the beta oxidation pathway, in certain embodiments.

In certain embodiments, host activity of one or more acyl CoA oxidase genes can be increased by genetically altering (e.g., increasing) the amount of a polypeptide produced (e.g., a strongly transcribed or constitutively expressed heterologous promoter is introduced in operable linkage with a polynucleotide that encodes a polypeptide; the copy number of a polynucleotide that encodes the polypeptide is increased (e.g., by introducing a plasmid that includes the polynucleotide, integration of additional copies in the host genome)). In some embodiments, host activity of one or more acyl CoA oxidases can be decreased by disruption (e.g., knockout, insertion mutagenesis, the like and combinations thereof) of an acyl CoA oxidase gene, or by decreasing the activity of the promoter (e.g., addition of repressor sequences to the promoter or 5'UTR) which transcribes an acyl CoA oxidase gene.

As noted above, disruption of nucleotide sequences encoding one or more acyl CoA oxidases (e.g., POX4, POX 5, or POX4 and POX5) sometimes can alter pathway efficiency, specificity and/or specific activity with respect to metabolism of carbon chains of different lengths (e.g., carbon chains including fatty alcohols, fatty acids, paraffins, dicarboxylic acids, aliphatic molecules of between about 1 and about 60 carbons in length). In some embodiments, the nucleotide sequence of one or more acyl CoA oxidases (e.g., POX4, POX 5, or POX4 and POX5) is disrupted with a URA3 nucleotide sequence encoding a selectable marker, and introduced to a host microorganism, thereby generating an engineered organism deficient in an acyl CoA oxidase activity.

Site-directed mutagenesis of Candida Pox4 and Pox5for altering substrate specificity

In some embodiments, amino acid residues located in or near the substrate binding pocket (e.g., residues located in helical domain 3 (see TABLE 29) of POX5 are mutated (e.g., deleted or substituted) to alter the activity profile of POX5. In some embodiments amino acid 428 and/or 429 in POX5, or corresponding residues in a related enzyme (e.g., see related enzymes and corresponding residues in TABLE 29) are located in helical domain 3 and independently are mutated (e.g., deleted or substituted) to alter the activity profile of POX5 or a related enzyme. Sometimes amino acid 428 and/or 429 independently are substituted with a suitable amino acid, non-limiting examples of which include positive charged and hydrophilic amino acids (e.g.,

lysine, arginine or histidine), negative charged and hydrophilic amino acids (e.g., aspartic acid, glutamic acid), sulfur containing amino acids (e.g., serine, threonine, cysteine, methionine), amide containing amino acids (e.g., asparagine, glutamine), aliphatic amino acids (e.g., glycine, alanine, valine, leucine, isoleucine), aromatic amino acid (e.g., phenylalanine, tyrosine, tryptophan), proline, ornithine, selenocystein, taurine, the like, or combinations thereof.

Also as noted above, catalytic specificity of acyl CoA oxidases (e.g., POX4, POX5) can be altered by a variety of methods. Altering the binding and/or catalytic specificity of acyl CoA oxidases may prove advantageous for generating novel acyl CoA oxidases with altered chain length recognition, altered chain length catalytic activity, and/or generation of an acyl CoA oxidase activity with a narrow or specific chain length specificity, thereby allowing further increases in pathway efficiency, specificity and/or specific activity with respect to metabolism of carbon chains of different lengths or metabolism of carbon chain distributions found in a particular chosen feedstock. In some embodiments the altered acyl CoA oxidase sequences are identified and/or generated by: (i) screening naturally occurring variant populations; (ii) mutagenesis of endogenous sequences; (iii) introduction of heterologous sequences having a desired specificity (e.g., introduction of one or more unmodified or modified acyl CoA oxidases from another organism into a host organism in which one or more endogenous acyl-CoA oxidases are optionally disrupted); (iv) generation of chimeric sequences having a portion of the coding sequence from one polynucleotide source (e.g., gene, organism) and a portion of the coding sequence from another source and/or (v) intelligent design using nucleotide sequences and three dimensional structure analysis from an acyl CoA oxidase having a desired specificity to remodel an endogenous acyl CoA oxidase, thereby generating a novel specificity enzyme. In some embodiments a chimeric acyl CoA oxidase sequence can have polynucleotide sequence contributions from two or more sources. In some embodiments, a chimeric acyl CoA oxidase sequence comprises a portion of the coding sequences from an endogenous polynucleotide and a portion of the coding sequence from a heterologous polynucleotide. Described in the examples are methods utilized to identify and/or generate acyl CoA oxidases with novel catalytic and binding specificities.

Nucleic acid sequences encoding acyl CoA oxidases (e.g., POX4 and POX5) can be obtained from any suitable source, including any animal (e.g., mammals, fish, reptiles, amphibians, etc.), any plant, fungus, yeast, protozoan, bacteria, virus, phage, and the like). Non-limiting examples of suitable yeast sources include Yarrowia yeast (e.g., Y. lipolytica (formerly

classified as Candida lipolytica)), Candida yeast (e.g., C. revkaufi, C. viswanathii, C. pulcherrima, C. tropicalis, C. utilis), Rhodotorula yeast (e.g., R. glutinus, R. graminis), Rhodosporidium yeast (e.g., R. toruloides), Saccharomyces yeast (e.g., S. cerevisiae, S. bayanus, S. pastorianus, S. carlsbergensis), Cryptococcus yeast, Trichosporon yeast (e.g., T. pullans, T. cutaneum), Pichia yeast (e.g., P. pastoris) and Lipomyces yeast (e.g., L. starkeyii, L. lipoferus). In some embodiments, a suitable yeast is of the genus Arachniotus, Aspergillus, Aureobasidium, Auxarthron, Blastomyces, Candida, Chrysosporuim, Chrysosporuim Debaryomyces, Coccidiodes, Cryptococcus, Gymnoascus, Hansenula, Histoplasma, Issatchenkia, Kluyveromyces, Lipomyces, Lssatchenkia, Microsporum, Myxotrichum, Myxozyma, Oidiodendron, Pachysolen, Penicillium, Pichia, Rhodosporidium, Rhodotorula, Rhodotorula, Saccharomyces , Schizosaccharomyces, Scopulariopsis, Sepedonium, Trichosporon, or Yarrowia. In some embodiments, a suitable yeast is of the species Arachniotus flavoluteus, Aspergillus flavus, Aspergillus fumigatus, Aspergillus niger, Aureobasidium pullulans, Auxarthron thaxteri, Blastomyces dermatitidis, Candida albicans, Candida dubliniensis, Candida famata, Candida glabrata, Candida guilliermondii, Candida kefyr, Candida krusei, Candida lambica, Candida lipolytica, Candida lustitaniae, Candida parapsilosis, Candida pulcherrima, Candida revkaufi, Candida rugosa, Candida tropicalis, Candida utilis, Candida viswanathii, Candida xestobii, Chrysosporuim keratinophilum, Coccidiodes immitis, Cryptococcus albidus var. diffluens, Cryptococcus laurentii, Cryptococcus neofomans, Debaryomyces hansenii, Gymnoascus dugwayensis, Hansenula anomala, Histoplasma capsulatum, Issatchenkia occidentalis, Isstachenkia orientalis, Kluyveromyces lactis, Kluyveromyces marxianus, Kluyveromyces thermotolerans, Kluyveromyces waltii, Lipomyces lipoferus, Lipomyces starkeyii, Microsporum gypseum, Myxotrichum deflexum, Oidiodendron echinulatum, Pachysolen tannophilis, Penicillium notatum, Pichia anomala, Pichia pastoris, Pichia stipitis, Rhodosporidium toruloides, Rhodotorula glutinus, Rhodotorula graminis, Saccharomyces cerevisiae, Saccharomyces kluyveri, Schizosaccharomyces pombe, Scopulariopsis acremonium, Sepedonium chrysospermum, Trichosporon cutaneum, Trichosporon pullans, Yarrowia lipolytica, or Yarrowia lipolytica (formerly classified as Candida lipolytica). In some embodiments, a suitable yeast is a Y. lipolytica strain that includes, but is not limited to, ATCC20362, ATCC8862, ATCC 18944, ATCC20228, ATCC76982 and LGAM S(7)1 strains (Papanikolaou S., and Aggelis G., Bioresour. Technol. 82(1):43-9 (2002)). In certain embodiments, a suitable yeast is a Candida species (i.e., Candida spp.) yeast. Any nucleic acid sequence encoding an acyl CoA oxidase, acyl CoA oxidase-like activity or acyl-CoA dehydrogenase activity can be

used to alter the substrate specificity of a yeast as described herein. Non-limiting examples of acyl CoA oxidase, acyl CoA oxidase-like and acyl CoA dehydrogenase amino acid sequences and nucleotide sequences are provided herein and in SEQ ID NO. 51 through 3810. Described in the examples are experiments conducted to modify and amplify the activity of an acyl CoA oxidase gene (e.g., the POX5 gene). Percent identity at the amino acid and nucleotide level of POX 4 and POX 5 from Candida strain ATCC 20336 and Yarrowia lipolytica is shown in the tables below:

Percent Identity (amino acid)

	POX4	YlAco2	YlAco3	YlAco5
POX5	59.4	41.6	41.6	40.5
POX4		40.4	41.6	39.1
YlAco2			66.8	61
YlAco3				68.8

Percent Identity (nucleotide)

	POX4	YlAco2	YlAco3	YlAco5
POX5	68.6	54.9	53.4	52.4
POX4		55.7	55	52.4
YlAco2			72.1	67.9
YlAco3				72.1

Presence, absence or amount of acyl CoA oxidase activity can be detected by any suitable method known in the art. For example, enzymatic assays as described in Shimizu et al, 1979, and as described herein in the Examples can be used to assess acyl CoA oxidase activity. Nucleic acid sequences representing native and/or disrupted acyl CoA oxidase sequences also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered organism exhibits decreased RNA and/or polypeptide levels as compared to the host organism.

Acyl CoA Dehydrogenase

Acyl-CoA dehydrogenases (ACADs) are a class of enzymes that can function to catalyze the initial step in each cycle of fatty acid β-oxidation in the mitochondria of cells. They can be very similar in structure and function to Acyl CoA oxidases. Their action results in the introduction of a trans double-bond between C2 and C3 of an acyl-CoA thioester substrate. FAD is a required co-factor in the mechanism in order for the enzyme to bind to its appropriate substrate.

Acyl-CoA dehydrogenases can be categorized into four distinct groups based on their specificity for short-, medium-, or long-chain fatty acid, and very long-chain fatty acid acyl-CoA substrates. While different dehydrogenases target fatty acids of varying chain length, all types of acyl-CoA dehydrogenases can be mechanistically similar. Differences in ACADs can occur based on the location of the active site along the amino acid sequence.

The medium chain acyl-CoA dehydrogenase is a homotetramer with each subunit containing roughly 400 amino acids and one equivalent of FAD. The tetramer is classified as a "dimer of dimers".

The interface between the two monomers of a single dimer of an acyl-CoA dehydrogenase contains the FAD binding sites and has extensive bonding interactions. In contrast, the

interface between the two dimers has few interactions. There are a total of 4 active sites within the tetramer, each of which contains a single FAD molecule and an acyl-CoA substrate. This gives a total of four FAD molecules and four acyl-CoA substrates per enzymatic molecule.

FAD is bound between the three domains of the monomer, where only the nucleotide portion is accessible. FAD binding contributes significantly to overall enzyme stability. The acyl-CoA substrate is bound completely within each monomer of the enzyme. In some ACADs, the active site is lined with the residues F252, T255, V259, T96, T99, A100, L103, Y375, Y375, and E376. The area of interest within the substrate can become wedged between Glu 376 and FAD, lining up the molecules into an ideal position for the reaction.

Some ACAD sequences are presented in SEQ ID NO.s 3728 through 3810.

Thioesterase

The term "thioesterase activity" as used herein refers to removal of Coenzyme A from hexanoate. The term "thioesterase activity" as used herein also refers to the removal of Coenzyme A from an activated fatty acid (e.g., fatty-acyl-CoA). A Non-limiting example of an enzyme with thioesterase activity includes acyl-CoA hydrolase (e.g., EC 3.1.2.20; also referred to as acyl coenzyme A thioesterase, acyl-CoA thioesterase, acyl coenzyme A hydrolase, thioesterase B, thioesterase II, lecithinase B, lysophopholipase LI, acyl-CoA thioesterase 1, and acyl-CoA thioesterase). Thioesterases that remove Coenzyme A from fatty-acyl-CoA molecules catalyze the reaction,

acyl-CoA + H20 -→ CoA + a carboxylate,

where the carboxylate often is a fatty acid. The released Coenzyme A can then be reused for other cellular activities.

The thioesterase activity can be provided by a polypeptide. In certain embodiments, the polypeptide is an endogenous nucleotide sequence that is increased in copy number, operably linked to a heterologous and/or endogenous promoter, or increased in copy number and operably linked to a heterologous and/or endogenous promoter. In some embodiments, the polypeptide is encoded by a heterologous nucleotide sequence introduced to a host

microorganism. Nucleic acid sequences conferring thioesterase activity can be obtained from a number of sources, including Cuphea lanceolata, C. tropicalis (e.g., see SEQ ID NOS: 33 and 35), and E. coli (e.g., see SEQ ID NO: 37). Additional organisms that can be used as thioesterase polynucleotide sequence donors are given herein. Examples of such polypeptides include, without limitation, acyl-(ACP) thioesterase type B from Cuphea lanceolata (see SEQ ID NO: 1), acyl-CoA hydrolase (e.g., ACHA and ACHB, see SEQ ID NOS: 34 and 36)) from C. tropicalis, acyl-CoA thioesterase (e.g., TESA, see SEQ ID NO: 38) from E. coli. A non-limiting example of a thioesterase polynucleotide sequences is referenced by accession number CAB60830 at the World Wide Web Uniform Resource Locator (URL) ncbi.nlm.nih.gov of the National Center for Biotechnology Information (NCBI).

Presence, absence or amount of thioesterase activity can be detected by any suitable method known in the art. An example of such a method is described Chemistry and Biology 9: 981-988. In some embodiments, thioesterase activity is not altered in a host microorganism, and in certain embodiments, the activity is added or increased in the engineered microorganism relative to the host microorganism. In some embodiments, a polypeptide having thioesterase activity is linked to another polypeptide (e.g., a hexanoate synthase A or hexanoate synthase B polypeptide). Non-limiting examples of polynucleotide sequences encoding thioesterase activities and polypeptides having thioesterase activity are provided in Example 33.

Reducing Omega Fatty Acid Conversion - General

The term "a genetic modification that reduces omega hydroxyl fatty acid conversion" as used herein refer to genetic alterations of a host microorganism that reduce an endogenous activity that converts an omega hydroxyl fatty acid to another product. In some embodiments, an endogenous omega hydroxyl fatty acid dehydrogenase activity is reduced. Such alterations can advantageously increase the amount of a dicarboxylic acid, which can be purified and further processed.

Reducing Beta Oxidation - General

The term "a genetic modification that reduces beta-oxidation activity" as used herein refers to a genetic alteration of a host microorganism that reduces an endogenous activity that oxidizes a beta carbon of carboxylic acid containing organic molecules. In certain embodiments, the

organic molecule is a ten or twelve carbon molecule, and sometimes contains one or two carboxylic acid moieties located at a terminus of the molecule (e.g., sebacic or dodecanedioic acid). Such alterations can advantageously increase yields of end products, such as a fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid).

Increasing Fatty Acid Synthesis - General

The term "a genetic modification that results in increased fatty acid synthesis" as used herein also refers to a genetic alteration of a host microorganism that reduces an endogenous activity that converts fatty acids into fatty-acyl-CoA intermediates. In some embodiments, an endogenous activity that converts fatty acids into fatty-acyl-CoA intermediates is reduced. In certain embodiments, an acyl-CoA synthetase activity is reduced. Such alterations can advantageously increase yields of end products, such as a fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid).

Acyl-CoA Synthetase

Fatty acids can be converted into fatty-acyl-CoA intermediates by the activity of an acyl-CoA synthetase (e.g., ACS1, ACS2; EC 6.2.1.3; also referred to as acyl-CoA synthetase, acyl-CoA ligase), in many organisms. Acyl-CoA synthetase has six isoforms encoded by ACS1, FAT1, ACS2A, ACS2B, ACS2C and ACS2D, respectively, in Candida spp. (e.g., homologous to FAAl, FAT1, and FAA2 in S. cerevisiae). Acyl-CoA synthetase is a member of the ligase class of enzymes and catalyzes the reaction,

ATP + Fatty Acid + CoA <=> AMP + Pyrophosphate + Fatty-Acyl-CoA.

Fatty acids and Coenzyme A often are utilized in the activation of fatty acids to fatty-acyl-CoA intermediates for entry into various cellular processes. Without being limited by theory, it is believed that reduction in the amount of fatty-acyl-CoA available for various cellular processes can increase the amount of fatty acids available for conversion into a fatty dicarboxylic acid (e.g., a sebacic or dodecanedioic acid) by other engineered pathways in the same host organism (e.g., omega oxidation pathway, beta oxidation pathway, omega oxidation pathway and beta

oxidation pathway). Acyl-CoA synthetase can be inactivated by any suitable means. Described herein are gene knockout methods suitable for use to disrupt the nucleotide sequence that encodes a polypeptide having ACS1 activity. A nucleotide sequence of ACS1 is provided in Example 33, SEQ ID NO: 39. An example of an integration/disruption construct, configured to generate a deletion mutant for ACS1 is also provided in the Examples.

The presence, absence or amount of acyl-CoA synthetase activity can be detected by any suitable method known in the art. Non-limiting examples of suitable detection methods include enzymatic assays (e.g., Lageweg et al "A Fluorometric Assay for Acyl-CoA Synthetase Activity", Analytical Biochemistry, 197(2):384-388 (1991)), PCR based assays (e.g., qPCR, RT-PCR), immunological detection methods (e.g., antibodies specific for acyl-CoA synthetase), the like and combinations thereof.

The term "a genetic modification that results in increased fatty acid synthesis" as used herein also refers to a genetic alteration of a host microorganism that reduces an endogenous activity that converts long chain and very long chain fatty acids into activated fatty-acyl-CoA intermediates. In some embodiments, an endogenous activity that converts long chain and very long chain fatty acids into activated fatty-acyl-CoA intermediates is reduced. In certain embodiments, a long chain acyl-CoA synthetase activity is reduced. Such alterations can advantageously increase yields of end products, such as a fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid).

Long chain fatty acids (e.g., C12-C18 chain lengths) and very long chain fatty acids (e.g., C20-C26) often are activated and/or transported by the thioesterification activity of a long-chain acyl-CoA synthetase (e.g., FAT1; EC 6.2.1.3; also referred to as long-chain fatty acid-CoA ligase, acyl-CoA synthetase; fatty acid thiokinase (long chain); acyl-activating enzyme; palmitoyl-CoA synthase; lignoceroyl-CoA synthase; arachidonyl-CoA synthetase; acyl coenzyme A synthetase; acyl-CoA ligase; palmitoyl coenzyme A synthetase; thiokinase; palmitoyl-CoA ligase; acyl-coenzyme A ligase; fatty acid CoA ligase; long-chain fatty acyl coenzyme A synthetase; oleoyl-CoA synthetase; stearoyl-CoA synthetase; long chain fatty acyl-CoA synthetase; long-chain acyl CoA synthetase; fatty acid elongase (ELO); LCFA synthetase; pristanoyl-CoA synthetase; ACS3; long-chain acyl-CoA synthetase I; long-chain acyl-CoA synthetase II; fatty acyl-coenzyme A synthetase; long-chain acyl-coenzyme A synthetase; and

acid:CoA ligase (AMP-forming)), in some organisms. Fatty acids also can be transported into the host organism from feedstocks by the activity of long chain acyl-CoA synthetase. Long-chain acyl-CoA synthetase catalyzes the reaction,

ATP + a long-chain carboxylic acid + CoA = AMP + diphosphate + an acyl-CoA,

where "an acyl-CoA" refers to a fatty-acyl-CoA molecule. As noted herein, activation of fatty acids is often necessary for entry of fatty acids into various cellular processes (e.g., as an energy source, as a component for membrane formation and/or remodeling, as carbon storage molecules). Deletion mutants of FAT 1 have been shown to accumulate very long chain fatty acids and exhibit decreased activation of these fatty acids. Without being limited by theory, it is believed that reduction in the activity of long-chain acyl-CoA synthetase may reduce the amount of long chain fatty acids converted into fatty-acyl-CoA intermediates, thereby increasing the amount of fatty acids available for conversion into a fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid) by other engineered pathways in the same host organism (e.g., omega oxidation pathway, beta oxidation pathway, omega oxidation pathway and beta oxidation pathway). Long-chain-acyl-CoA synthetase activity can be reduced or inactivated by any suitable means. Described herein are gene knockout methods suitable for disrupting the nucleotide sequence that encodes the polypeptide having FAT1 activity. The nucleotide sequence of FAT 1 is provided in Example 33, SEQ ID NO: 41. DNA vectors suitable for use in constructing "knockout" constructs are described herein.

The presence, absence or amount of long-chain-acyl-CoA synthetase activity can be detected by any suitable method known in the art. Non-limiting examples of suitable detection methods include enzymatic assays, binding assays (e.g., Erland et al, Analytical Biochemistry 295(1):38-44 (2001)), PCR based assays (e.g., qPCR, RTPCR), immunological detection methods (e.g., antibodies specific for long-chain-acyl-CoA synthetase), the like and combinations thereof.

Selective modification of ACS activity

In some embodiments, a beta-oxidation pathway is functional and is modified for selective substrate specificity. In some embodiments a beta-oxidation pathway is selective for only diacyl-CoA thioesters and in some embodiments only on diacyl-CoA's of a chain length greater

than 6, 8, 10, 12, 14, 16, 18 or 20 carbons. Beta-oxidation selectivity can be achieved by: 1) utilizing the difference in transport of acyl-CoA's and diacids across the peroxisomal membrane, 2) selectively knocking out acyl-CoA synthetase (ACS) activity in the cytosolic compartment, 3) knocking out ACS activity in the peroxisomal compartment for isozymes with substrate specificity for short chain substrates, and/or 4) engineering a beta-oxidation pathway that will work only on substrates longer than 6, 8, 10, 12, 14, 16, 18 or 20 carbons.

In S. cerevisiae, cytoplasmic ACS activity is encoded by FAAl, FAA3, FAA4 and FATl, while peroxisomal activity is encoded by FAA2. Homologs for FAAl and FATl were identified in Candida strains however there were no identified homologs for FAA3 or FAA4. As many as five homologs for the S. cerevisiae peroxisomal FAA2 were identified in Candida strains. Two of the five homologs display 95% identity to one another and are most likely alleles of the same gene. Four FAA2 homologs were identified in Candida strain ATCC20336 (e.g., ACS2A through ACS2D).

In some embodiments, one strategy is to control the subcellular location of ACS enzyme activity so that it is only present in the peroxisome. FAAl and FATl mutants, faalA and fatlA were constructed and should have very little ACS activity targeted to the cytoplasm. In these mutant strains, exogenously supplied long-chain free fatty acids accumulate in the cytoplasm since they cannot be transported into the peroxisome unless they are activated to the acyl-CoA thioester. High concentrations of free fatty acid can be toxic, so the cell acts to detoxify itself by oxidizing the free fatty acids to dicarboxylic acids that are much less toxic. Unlike long-chain fatty acids, long-chain dicarboxylic acids are able to diffuse into the peroxisomal compartment where they can then be activated to diacyl-CoA thioesters, which is required for entry into the beta-oxidation pathway. With multiple peroxisomal ACS isozymes it may be that each isozyme has different substrate specificity. In some embodiments, it is desired to retain those peroxisomal ACS enzymes with substrate specificity matching the chain-length of the fatty acid feedstock but without activity (or low activity) on diacids of chain-length <6, 8, 10, 12, 14, 16, 18 or 20 carbons. With this strategy any long-chain dicarboxyl-CoA that is chain-shortened by beta-oxidation to 12 carbons, for example, that is subsequently hydrolyzed to a dicarboxylic acid and free CoA cannot be reactivated to a dicarboxyl-CoA for re-entry into beta-oxidation for further chain shortening. In some embodiments, in combination with controlling the substrate chain-length specificity of the peroxisomal ACS, a peroxisomal thioesterase activity is amplified with maximum activity at the desired chain-length of our

product. This strategy can control the chain-length of the dicarboxylic acid produced by beta-oxidation.

In some embodiments, the flow of fatty acids into the peroxisome is controlled by knocking out the genes PXA1 and PXA2. These genes encode subunits of an ATP binding cassette transporter that is responsible for transporting long-chain fatty acyl-CoA's from the cytoplasm across the peroxisomal membrane into the peroxisomal matrix. Even though, in some embodiments, the genes encoding the cytoplasmic ACS's are knocked out, there may still be some residual ACS activity in the cytoplasm from the peroxisomal ACS's. The ACS isozymes destined for the peroxisome are translated in the cytoplasm and fully folded prior to import into the peroxisome. Therefore the peroxisomal ACS's may contribute to a small amount of cytoplasmic ACS activity. Deletion of the Pxalp/Pxa2p transporter can prevent any of the long-chain fatty acids activated to acyl-CoA thioesters from being transported into the peroxisome for degradation.

Acyl-CoA sterol acyltransferase

The term "a genetic modification that results in increased fatty acid synthesis" as used herein also refers to a genetic alteration of a host microorganism that reduces an endogenous activity that converts fatty acids into cholesterol esters. In some embodiments, an endogenous activity that converts fatty acids into cholesterol esters is reduced. In certain embodiments, an acyl-CoA sterol acyltransferase activity is reduced. Such alterations can advantageously increase yields of end products, such as a fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid).

Fatty acids can be converted into a cholesterol-ester by the activity of acyl-CoA sterol acyltransferase (e.g., ARE1, ARE2, EC 2.3.1.26; also referred to as sterol O-acyltransferase; cholesterol acyltransferase; sterol-ester synthase; sterol-ester synthetase; sterol-ester synthase; acyl coenzyme A-cholesterol-O-acyltransferase; acyl-CoA:cholesterol acyltransferase; ACAT; acylcoenzyme Axholesterol O-acyltransferase; cholesterol ester synthase; cholesterol ester synthetase; and cholesteryl ester synthetase), in many organisms. Without being limited by any theory, cholesterol esterification may be involved in directing fatty acids away from

incorporation into cell membranes and towards storage forms of lipids. Acyl-CoA sterol acyltransferase catalyzes the reaction,

acyl-CoA + cholesterol = CoA + cholesterol ester.

The esterification of cholesterol is believed to limit its solubility in cell membrane lipids and thus promotes accumulation of cholesterol ester in the fat droplets (e.g., a form of carbon storage molecule) within cytoplasm. Therefore, without being limited by any theory esterification of cholesterol may cause the accumulation of lipid storage molecules, and disruption of the activity of acyl-CoA sterol acyltransferase may cause an increase in acyl-CoA levels that can be converted into a fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid) by other engineered pathways in the same host organism (e.g., omega oxidation pathway, beta oxidation pathway, omega oxidation pathway and beta oxidation pathway). Acyl-CoA sterol acyltransferase can be inactivated by any suitable means. Described herein are gene knockout methods suitable for disrupting nucleotide sequences that encode polypeptides having AREl activity, ARE2 activity or AREl activity and ARE2 activity. The nucleotide sequences of AREl and ARE2 are provided in Example 33, SEQ ID NOS: 43 and 45. DNA vectors suitable for use in constructing "knockout" constructs are described herein.

The presence, absence or amount of acyl-CoA sterol acyltransferase activity can be detected by any suitable method known in the art. Non-limiting examples of suitable detection methods include enzymatic assays (e.g., Chen et al, Plant Physiology 145:974-984 (2007)), binding assays, PCR based assays (e.g., qPCR, RT-PCR), immunological detection methods (e.g., antibodies specific for long-chain-acyl-CoA synthetase), the like and combinations thereof.

Diacylglycerol acyltransferase & Acyltransferases

The term "a genetic modification that results in increased fatty acid synthesis" as used herein also refers to a genetic alteration of a host microorganism that reduces an endogenous activity that catalyzes diacylglycerol esterification (e.g., addition of acyl group to a diacylglycerol to form a triacylglycerol). In some embodiments, an endogenous activity that converts diacylglycerol into triacylglycerol is reduced. In certain embodiments, an acyltransferase

activity is reduced. In some embodiments a diacylglycerol acyltransferase activity is reduced. In some embodiments a diacylglycerol acyltransferase (e.g., DGAl, EC 2.3.1.20) activity and an acyltransferase (e.g., LROl) activity are reduced. Such alterations can advantageously increase yields of end products, such as a fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid).

Diacylglycerol can be converted into triacylglycerol by the activity of diacylglycerol acyltransferase (e.g., DGAl; EC 2.3.1.20; also referred to as diglyceride acyltransferase; 1,2-diacylglycerol acyltransferase; diacylglycerol acyltransferase; diglyceride O-acyltransferase; palmitoyl-CoA-sn-l,2-diacylglycerol acyltransferase; acyl-CoA:l,2-diacylglycerol O-acyltransferase and acyl-CoA:l,2-diacyl-sn-glycerol O-acyltransferase), in many organisms. Diacylglycerol acyltransferase catalyzes the reaction,

Acyl-CoA + 1,2-diacyl-sn-glycerol = CoA + triacylglycerol,

and is generally considered the terminal and only committed step in triglyceride synthesis. The product of the DGAl gene in yeast normally is localized to lipid particles.

In addition to the diacylglycerol esterification activity described for DGAl, many organisms also can generate triglycerides by the activity of other acyltransferase activities, non-limiting examples of which include lecithin-cholesterol acyltransferase activity (e.g., LROl; EC 2.3.1.43; also referred to as phosphatidylcholine-sterol O-acyltransferase activity; lecithin-cholesterol acyltransferase activity; phospholipid-cholesterol acyltransferase activity; LCAT (lecithin-cholesterol acyltransferase) activity; lecithinxholesterol acyltransferase activity; and lysolecithin acyltransferase activity) and phospholipid:diacylglycerol acyltransferase (e.g., EC 2.3.1.158; also referred to as PDAT activity and phospholipid: 1,2-diacyl-sn-glycerol O-acyltransferase activity). Acyltransferases of the families EC 2.3.1.43 and EC 2.3.1.58 catalyze the general reaction,

phospholipid + 1,2-diacylglycerol = lysophospholipid + triacylglycerol.

Triacylglycerides often are utilized as carbon (e.g., fatty acid or lipid) storage molecules. Without being limited by any theory, it is believe that reducing the activity of acyltransferase

may reduce the conversion of diacylglycerol to triacylglycerol, which may cause increased accumulation of fatty acid, in conjunction with additional genetic modifications (e.g., lipase to further remove fatty acids from the glycerol backbone) that can be converted into a fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid) by other engineered pathways in the same host organism (e.g., omega oxidation pathway, beta oxidation pathway, omega oxidation pathway and beta oxidation pathway). Acyltransferases can be inactivated by any suitable means. Described herein are gene knockout methods suitable for disrupting nucleotide sequences that encode polypeptides having DGA1 activity, LROl activity or DGA1 activity and LROl activity. The nucleotide sequence of DGA1 is provided in Example 33, SEQ ID NO: 47 The nucleotide sequence of LROl is provided in Example 33, SEQ ID NO: 49. DNA vectors suitable for use in constructing "knockout" constructs are described herein.

The presence, absence or amount of acyltransferase activity can be detected by any suitable method known in the art. Non-limiting examples of suitable detection methods include enzymatic assays (e.g., Geelen, Analytical Biochemistry 322(2):264-268 (2003), Dahlqvist et al, PNAS 97(12):6487-6492 (2000)), binding assays, PCR based assays (e.g., qPCR, RTPCR), immunological detection methods (e.g., antibodies specific for a DGA1 or LROl acyltransferase), the like and combinations thereof.

Polynucleotides and Polypeptides

A nucleic acid (e.g., also referred to herein as nucleic acid reagent, target nucleic acid, target nucleotide sequence, nucleic acid sequence of interest or nucleic acid region of interest) can be from any source or composition, such as DNA, cDNA, gDNA (genomic DNA), RNA, siRNA (short inhibitory RNA), RNAi, tRNA or mRNA, for example, and can be in any form (e.g., linear, circular, supercoiled, single-stranded, double-stranded, and the like). A nucleic acid can also comprise DNA or RNA analogs (e.g., containing base analogs, sugar analogs and/or a non-native backbone and the like). It is understood that the term "nucleic acid" does not refer to or infer a specific length of the polynucleotide chain, thus polynucleotides and oligonucleotides are also included in the definition. Deoxyribonucleotides include deoxyadenosine, deoxycytidine, deoxyguanosine and deoxythymidine. For RNA, the uracil base is uridine.

A nucleic acid sometimes is a plasmid, phage, autonomously replicating sequence (ARS), centromere, artificial chromosome, yeast artificial chromosome (e.g., YAC) or other nucleic acid able to replicate or be replicated in a host cell. In certain embodiments a nucleic acid can be from a library or can be obtained from enzymatically digested, sheared or sonicated genomic DNA (e.g., fragmented) from an organism of interest. In some embodiments, nucleic acid subjected to fragmentation or cleavage may have a nominal, average or mean length of about 5 to about 10,000 base pairs, about 100 to about 1,000 base pairs, about 100 to about 500 base pairs, or about 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000 or 10000 base pairs. Fragments can be generated by any suitable method in the art, and the average, mean or nominal length of nucleic acid fragments can be controlled by selecting an appropriate fragment-generating procedure by the person of ordinary skill. In some embodiments, the fragmented DNA can be size selected to obtain nucleic acid fragments of a particular size range.

Nucleic acid can be fragmented by various methods known to the person of ordinary skill, which include without limitation, physical, chemical and enzymic processes. Examples of such processes are described in U.S. Patent Application Publication No. 200501 12590 (published on May 26, 2005, entitled "Fragmentation-based methods and systems for sequence variation detection and discovery," naming Van Den Boom et al.). Certain processes can be selected by the person of ordinary skill to generate non-specifically cleaved fragments or specifically cleaved fragments. Examples of processes that can generate non-specifically cleaved fragment sample nucleic acid include, without limitation, contacting sample nucleic acid with apparatus that expose nucleic acid to shearing force (e.g., passing nucleic acid through a syringe needle; use of a French press); exposing sample nucleic acid to irradiation (e.g., gamma, x-ray, UV irradiation; fragment sizes can be controlled by irradiation intensity); boiling nucleic acid in water (e.g., yields about 500 base pair fragments) and exposing nucleic acid to an acid and base hydrolysis process.

Nucleic acid may be specifically cleaved by contacting the nucleic acid with one or more specific cleavage agents. The term "specific cleavage agent" as used herein refers to an agent, sometimes a chemical or an enzyme that can cleave a nucleic acid at one or more specific sites. Specific cleavage agents often will cleave specifically according to a particular nucleotide sequence at a particular site. Examples of enzyme specific cleavage agents include without limitation endonucleases (e.g., DNase (e.g., DNase I, II); RNase (e.g., RNase E, F, H, P);

Cleavase™ enzyme; Taq DNA polymerase; E. coli DNA polymerase I and eukaryotic structure-specific endonucleases; murine FEN-1 endonucleases; type I, II or III restriction endonucleases such as Acc I, Afl III, Alu I, Alw44 I, Apa I, Asn I, Ava I, Ava II, BamH I, Ban II, Bel I, Bgl I. Bgl II, Bin I, Bsm I, BssH II, BstE II, Cfo I, Cla I, Dde I, Dpn I, Dra I, EcIX I, EcoR I, EcoR I, EcoR II, EcoR V, Hae II, Hae II, Hind II, Hind III, Hpa I, Hpa II, Kpn I, Ksp I, Mlu I, MIuN I, Msp I, Nci I, Nco I, Nde I, Nde II, Nhe I, Not I, Nru I, Nsi I, Pst I, Pvu I, Pvu II, Rsa I, Sac I, Sal I, Sau3A I, Sea I, ScrF I, Sfi I, Sma I, Spe I, Sph I, Ssp I, Stu I, Sty I, Swa I, Taq I, Xba I, Xho I); glycosylases (e.g., uracil-DNA glycolsylase (UDG), 3-methyladenine DNA glycosylase, 3-methyladenine DNA glycosylase II, pyrimidine hydrate-DNA glycosylase, FaPy-DNA glycosylase, thymine mismatch-DNA glycosylase, hypoxanthine-DNA glycosylase, 5-Hydroxymethyluracil DNA glycosylase (HmUDG), 5-Hydroxymethylcytosine DNA glycosylase, or 1,N6-etheno-adenine DNA glycosylase); exonucleases (e.g., exonuclease III); ribozymes, and DNAzymes. Sample nucleic acid may be treated with a chemical agent, or synthesized using modified nucleotides, and the modified nucleic acid may be cleaved. In non-limiting examples, sample nucleic acid may be treated with (i) alkylating agents such as methylnitrosourea that generate several alkylated bases, including N3-methyladenine and N3-methylguanine, which are recognized and cleaved by alkyl purine DNA-glycosylase; (ii) sodium bisulfite, which causes deamination of cytosine residues in DNA to form uracil residues that can be cleaved by uracil N-glycosylase; and (iii) a chemical agent that converts guanine to its oxidized form, 8-hydroxyguanine, which can be cleaved by formamidopyrimidine DNA N-glycosylase. Examples of chemical cleavage processes include without limitation alkylation, (e.g., alkylation of phosphorothioate-modified nucleic acid); cleavage of acid lability of P3'-N5'-phosphoroamidate-containing nucleic acid; and osmium tetroxide and piperidine treatment of nucleic acid.

As used herein, the term "complementary cleavage reactions" refers to cleavage reactions that are carried out on the same nucleic acid using different cleavage reagents or by altering the cleavage specificity of the same cleavage reagent such that alternate cleavage patterns of the same target or reference nucleic acid or protein are generated. In certain embodiments, nucleic acids of interest may be treated with one or more specific cleavage agents (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more specific cleavage agents) in one or more reaction vessels (e.g., nucleic acid of interest is treated with each specific cleavage agent in a separate vessel).

A nucleic acid suitable for use in the embodiments described herein sometimes is amplified by any amplification process known in the art (e.g., PCR, RT-PCR and the like). Nucleic acid amplification may be particularly beneficial when using organisms that are typically difficult to culture (e.g., slow growing, require specialize culture conditions and the like). The terms "amplify", "amplification", "amplification reaction", or "amplifying" as used herein refer to any in vitro processes for multiplying the copies of a target sequence of nucleic acid. Amplification sometimes refers to an "exponential" increase in target nucleic acid. However, "amplifying" as used herein can also refer to linear increases in the numbers of a select target sequence of nucleic acid, but is different than a one-time, single primer extension step. In some embodiments, a limited amplification reaction, also known as pre-amplification, can be performed. Pre-amplification is a method in which a limited amount of amplification occurs due to a small number of cycles, for example 10 cycles, being performed. Pre-amplification can allow some amplification, but stops amplification prior to the exponential phase, and typically produces about 500 copies of the desired nucleotide sequence(s). Use of pre-amplification may also limit inaccuracies associated with depleted reactants in standard PCR reactions.

In some embodiments, a nucleic acid reagent sometimes is stably integrated into the chromosome of the host organism, or a nucleic acid reagent can be a deletion of a portion of the host chromosome, in certain embodiments (e.g., genetically modified organisms, where alteration of the host genome confers the ability to selectively or preferentially maintain the desired organism carrying the genetic modification). Such nucleic acid reagents (e.g., nucleic acids or genetically modified organisms whose altered genome confers a selectable trait to the organism) can be selected for their ability to guide production of a desired protein or nucleic acid molecule. When desired, the nucleic acid reagent can be altered such that codons encode for (i) the same amino acid, using a different tRNA than that specified in the native sequence, or (ii) a different amino acid than is normal, including unconventional or unnatural amino acids (including detectably labeled amino acids). As described herein, the term "native sequence" refers to an unmodified nucleotide sequence as found in its natural setting (e.g., a nucleotide sequence as found in an organism).

A nucleic acid or nucleic acid reagent can comprise certain elements often selected according to the intended use of the nucleic acid. Any of the following elements can be included in or excluded from a nucleic acid reagent. A nucleic acid reagent, for example, may include one or

more or all of the following nucleotide elements: one or more promoter elements, one or more 5' untranslated regions (5'UTRs), one or more regions into which a target nucleotide sequence may be inserted (an "insertion element"), one or more target nucleotide sequences, one or more 3' untranslated regions (3'UTRs), and one or more selection elements. A nucleic acid reagent can be provided with one or more of such elements and other elements may be inserted into the nucleic acid before the nucleic acid is introduced into the desired organism. In some embodiments, a provided nucleic acid reagent comprises a promoter, 5'UTR, optional 3'UTR and insertion element(s) by which a target nucleotide sequence is inserted (i.e., cloned) into the nucleotide acid reagent. In certain embodiments, a provided nucleic acid reagent comprises a promoter, insertion element(s) and optional 3'UTR, and a 5' UTR/target nucleotide sequence is inserted with an optional 3'UTR. The elements can be arranged in any order suitable for expression in the chosen expression system (e.g., expression in a chosen organism, or expression in a cell free system, for example), and in some embodiments a nucleic acid reagent comprises the following elements in the 5' to 3' direction: (1) promoter element, 5'UTR, and insertion element(s); (2) promoter element, 5'UTR, and target nucleotide sequence; (3) promoter element, 5'UTR, insertion element(s) and 3'UTR; and (4) promoter element, 5'UTR, target nucleotide sequence and 3'UTR.

Promoters

A promoter element typically is required for DNA synthesis and/or RNA synthesis. A promoter element often comprises a region of DNA that can facilitate the transcription of a particular gene, by providing a start site for the synthesis of RNA corresponding to a gene. Promoters generally are located near the genes they regulate, are located upstream of the gene (e.g., 5' of the gene), and are on the same strand of DNA as the sense strand of the gene, in some embodiments. In some embodiments, a promoter element can be isolated from a gene or organism and inserted in functional connection with a polynucleotide sequence to allow altered and/or regulated expression. A non-native promoter (e.g., promoter not normally associated with a given nucleic acid sequence) used for expression of a nucleic acid often is referred to as a heterologous promoter. In certain embodiments, a heterologous promoter and/or a 5'UTR can be inserted in functional connection with a polynucleotide that encodes a polypeptide having a desired activity as described herein. The terms "operably linked" and "in functional connection with" as used herein with respect to promoters, refer to a relationship between a coding sequence and a promoter element. The promoter is operably linked or in functional connection

with the coding sequence when expression from the coding sequence via transcription is regulated, or controlled by, the promoter element. The terms "operably linked" and "in functional connection with" are utilized interchangeably herein with respect to promoter elements.

A promoter often interacts with a RNA polymerase. A polymerase is an enzyme that catalyzes synthesis of nucleic acids using a preexisting nucleic acid reagent. When the template is a DNA template, an RNA molecule is transcribed before protein is synthesized. Enzymes having polymerase activity suitable for use in the present methods include any polymerase that is active in the chosen system with the chosen template to synthesize protein. In some embodiments, a promoter (e.g., a heterologous promoter) also referred to herein as a promoter element, can be operably linked to a nucleotide sequence or an open reading frame (ORF). Transcription from the promoter element can catalyze the synthesis of an RNA corresponding to the nucleotide sequence or ORF sequence operably linked to the promoter, which in turn leads to synthesis of a desired peptide, polypeptide or protein.

Promoter elements sometimes exhibit responsiveness to regulatory control. Promoter elements also sometimes can be regulated by a selective agent. That is, transcription from promoter elements sometimes can be turned on, turned off, up-regulated or down-regulated, in response to a change in environmental, nutritional or internal conditions or signals (e.g., heat inducible promoters, light regulated promoters, feedback regulated promoters, hormone influenced promoters, tissue specific promoters, oxygen and pH influenced promoters, promoters that are responsive to selective agents (e.g., kanamycin) and the like, for example). Promoters influenced by environmental, nutritional or internal signals frequently are influenced by a signal (direct or indirect) that binds at or near the promoter and increases or decreases expression of the target sequence under certain conditions.

Non-limiting examples of selective or regulatory agents that can influence transcription from a promoter element used in embodiments described herein include, without limitation, (1) nucleic acid segments that encode products that provide resistance against otherwise toxic compounds (e.g., antibiotics); (2) nucleic acid segments that encode products that are otherwise lacking in the recipient cell (e.g., essential products, tRNA genes, auxotrophic markers); (3) nucleic acid segments that encode products that suppress the activity of a gene product; (4) nucleic acid segments that encode products that can be readily identified (e.g., phenotypic markers such as

antibiotics (e.g., β-lactamase), β-galactosidase, green fluorescent protein (GFP), yellow fluorescent protein (YFP), red fluorescent protein (RFP), cyan fluorescent protein (CFP), and cell surface proteins); (5) nucleic acid segments that bind products that are otherwise detrimental to cell survival and/or function; (6) nucleic acid segments that otherwise inhibit the activity of any of the nucleic acid segments described in Nos. 1-5 above (e.g., antisense oligonucleotides); (7) nucleic acid segments that bind products that modify a substrate (e.g., restriction endonucleases); (8) nucleic acid segments that can be used to isolate or identify a desired molecule (e.g., specific protein binding sites); (9) nucleic acid segments that encode a specific nucleotide sequence that can be otherwise non-functional (e.g., for PCR amplification of subpopulations of molecules); (10) nucleic acid segments that, when absent, directly or indirectly confer resistance or sensitivity to particular compounds; (11) nucleic acid segments that encode products that either are toxic or convert a relatively non-toxic compound to a toxic compound (e.g., Herpes simplex thymidine kinase, cytosine deaminase) in recipient cells; (12) nucleic acid segments that inhibit replication, partition or heritability of nucleic acid molecules that contain them; and/or (13) nucleic acid segments that encode conditional replication functions, e.g., replication in certain hosts or host cell strains or under certain environmental conditions (e.g., temperature, nutritional conditions, and the like). In some embodiments, the regulatory or selective agent can be added to change the existing growth conditions to which the organism is subjected (e.g., growth in liquid culture, growth in a fermenter, growth on solid nutrient plates and the like for example).

In some embodiments, regulation of a promoter element can be used to alter (e.g., increase, add, decrease or substantially eliminate) the activity of a peptide, polypeptide or protein (e.g., enzyme activity for example). For example, a microorganism can be engineered by genetic modification to express a nucleic acid reagent that can add a novel activity (e.g., an activity not normally found in the host organism) or increase the expression of an existing activity by increasing transcription from a homologous or heterologous promoter operably linked to a nucleotide sequence of interest (e.g., homologous or heterologous nucleotide sequence of interest), in certain embodiments. In some embodiments, a microorganism can be engineered by genetic modification to express a nucleic acid reagent that can decrease expression of an activity by decreasing or substantially eliminating transcription from a homologous or heterologous promoter operably linked to a nucleotide sequence of interest, in certain embodiments.

In some embodiments the activity can be altered using recombinant DNA and genetic techniques known to the artisan. Methods for engineering microorganisms are further described herein. Tables herein provide non-limiting lists of yeast promoters that are up-regulated by oxygen, yeast promoters that are down-regulated by oxygen, yeast transcriptional repressors and their associated genes, DNA binding motifs as determined using the MEME sequence analysis software. Potential regulator binding motifs can be identified using the program MEME to search intergenic regions bound by regulators for overrepresented sequences. For each regulator, the sequences of intergenic regions bound with p-values less than 0.001 were extracted to use as input for motif discovery. The MEME software was run using the following settings: a motif width ranging from 6 to 18 bases, the "zoops" distribution model, a 6th order Markov background model and a discovery limit of 20 motifs. The discovered sequence motifs were scored for significance by two criteria: an E-value calculated by MEME and a specificity score. The motif with the best score using each metric is shown for each regulator. All motifs presented are derived from datasets generated in rich growth conditions with the exception of a previously published dataset for epitope-tagged Gal4 grown in galactose.

In some embodiments, the altered activity can be found by screening the organism under conditions that select for the desired change in activity. For example, certain microorganisms can be adapted to increase or decrease an activity by selecting or screening the organism in question on a media containing substances that are poorly metabolized or even toxic. An increase in the ability of an organism to grow a substance that is normally poorly metabolized may result in an increase in the growth rate on that substance, for example. A decrease in the sensitivity to a toxic substance might be manifested by growth on higher concentrations of the toxic substance, for example. Genetic modifications that are identified in this manner sometimes are referred to as naturally occurring mutations or the organisms that carry them can sometimes be referred to as naturally occurring mutants. Modifications obtained in this manner are not limited to alterations in promoter sequences. That is, screening microorganisms by selective pressure, as described above, can yield genetic alterations that can occur in non-promoter sequences, and sometimes also can occur in sequences that are not in the nucleotide sequence of interest, but in a related nucleotide sequences (e.g., a gene involved in a different step of the same pathway, a transport gene, and the like). Naturally occurring mutants sometimes can be found by isolating naturally occurring variants from unique environments, in some embodiments.

Homology and Identity

In addition to the regulated promoter sequences, regulatory sequences, and coding polynucleotides provided herein, a nucleic acid reagent may include a polynucleotide sequence 80% or more identical to the foregoing (or to the complementary sequences). That is, a nucleotide sequence that is at least 80%> or more, 81% or more, 82% or more, 83% or more, 84%, or more, 85% or more, 86% or more, 87% or more, 88% or more, 89% or more, 90% or more, 91% or more, 92% or more, 93% or more, 94% or more, 95% or more, 96% or more, 97%, or more, 98% or more, or 99% or more identical to a nucleotide sequence described herein can be utilized. The term "identical" as used herein refers to two or more nucleotide sequences having substantially the same nucleotide sequence when compared to each other. One test for determining whether two nucleotide sequences or amino acids sequences are substantially identical is to determine the percent of identical nucleotide sequences or amino acid sequences shared.

Calculations of sequence identity can be performed as follows. Sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). The length of a reference sequence aligned for comparison purposes is sometimes 30% or more, 40% or more, 50% or more, often 60% or more, and more often 70% or more, 80% or more, 90% or more, or 100% of the length of the reference sequence. The nucleotides or amino acids at corresponding nucleotide or polypeptide positions, respectively, are then compared among the two sequences. When a position in the first sequence is occupied by the same nucleotide or amino acid as the corresponding position in the second sequence, the nucleotides or amino acids are deemed to be identical at that position. The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, introduced for optimal alignment of the two sequences.

Comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. Percent identity between two amino acid or nucleotide sequences can be determined using the algorithm of Meyers & Miller, CABIOS 4: 11-17 (1989), which has been incorporated into the ALIGN program (version 2.0), using a

PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4. Also, percent identity between two amino acid sequences can be determined using the Needleman & Wunsch, J. Mol. Biol. 48: 444-453 (1970) algorithm which has been incorporated into the GAP program in the GCG software package (available at the http address www.gcg.com), using either a Blossum 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. Percent identity between two nucleotide sequences can be determined using the GAP program in the GCG software package (available at http address www.gcg.com), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. A set of parameters often used is a Blossum 62 scoring matrix with a gap open penalty of 12, a gap extend penalty of 4, and a frameshift gap penalty of 5.

Sequence identity can also be determined by hybridization assays conducted under stringent conditions. As use herein, the term "stringent conditions" refers to conditions for hybridization and washing. Stringent conditions are known to those skilled in the art and can be found in Current Protocols in Molecular Biology, John Wiley & Sons, N.Y., 6.3.1-6.3.6 (1989). Aqueous and non-aqueous methods are described in that reference and either can be used. An example of stringent hybridization conditions is hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 50°C. Another example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 55°C. A further example of stringent hybridization conditions is hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 60°C. Often, stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 65°C. More often, stringency conditions are 0.5M sodium phosphate, 7% SDS at 65°C, followed by one or more washes at 0.2X SSC, 1% SDS at 65°C.

UTRs

As noted above, nucleic acid reagents may also comprise one or more 5' UTR's, and one or more 3'UTR's. A 5' UTR may comprise one or more elements endogenous to the nucleotide sequence from which it originates, and sometimes includes one or more exogenous elements. A 5' UTR can originate from any suitable nucleic acid, such as genomic DNA, plasmid DNA,

RNA or mRNA, for example, from any suitable organism (e.g., virus, bacterium, yeast, fungi, plant, insect or mammal). The artisan may select appropriate elements for the 5' UTR based upon the chosen expression system (e.g., expression in a chosen organism, or expression in a cell free system, for example). A 5' UTR sometimes comprises one or more of the following elements known to the artisan: enhancer sequences (e.g., transcriptional or translational), transcription initiation site, transcription factor binding site, translation regulation site, translation initiation site, translation factor binding site, accessory protein binding site, feedback regulation agent binding sites, Pribnow box, TATA box, -35 element, E-box (helix-loop-helix binding element), ribosome binding site, replicon, internal ribosome entry site (IRES), silencer element and the like. In some embodiments, a promoter element may be isolated such that all 5' UTR elements necessary for proper conditional regulation are contained in the promoter element fragment, or within a functional subsequence of a promoter element fragment.

A 5 'UTR in the nucleic acid reagent can comprise a translational enhancer nucleotide sequence. A translational enhancer nucleotide sequence often is located between the promoter and the target nucleotide sequence in a nucleic acid reagent. A translational enhancer sequence often binds to a ribosome, sometimes is an 18S rRNA-binding ribonucleotide sequence (i.e., a 40S ribosome binding sequence) and sometimes is an internal ribosome entry sequence (IRES). An IRES generally forms an RNA scaffold with precisely placed RNA tertiary structures that contact a 40S ribosomal subunit via a number of specific intermolecular interactions. Examples of ribosomal enhancer sequences are known and can be identified by the artisan (e.g., Mignone et al, Nucleic Acids Research 33: D141-D146 (2005); Paulous et al, Nucleic Acids Research 31: 722-733 (2003); Akbergenov et al, Nucleic Acids Research 32: 239-247 (2004); Mignone et al, Genome Biology 3(3): reviews0004. 1-000 1.10 (2002); Gallie, Nucleic Acids Research 30: 3401-341 1 (2002); Shaloiko et al, http address www.interscience.wiley.com, DOI: 10.1002/bit.20267; and Gallie et al, Nucleic Acids Research 15: 3257-3273 (1987)).

A translational enhancer sequence sometimes is a eukaryotic sequence, such as a Kozak consensus sequence or other sequence (e.g., hydroid polyp sequence, GenBank accession no. U07128). A translational enhancer sequence sometimes is a prokaryotic sequence, such as a Shine-Dalgarno consensus sequence. In certain embodiments, the translational enhancer sequence is a viral nucleotide sequence. A translational enhancer sequence sometimes is from a 5' UTR of a plant virus, such as Tobacco Mosaic Virus (TMV), Alfalfa Mosaic Virus (AMV); Tobacco Etch Virus (ETV); Potato Virus Y (PVY); Turnip Mosaic (poty) Virus and Pea Seed

Borne Mosaic Virus, for example. In certain embodiments, an omega sequence about 67 bases in length from TMV is included in the nucleic acid reagent as a translational enhancer sequence (e.g., devoid of guanosine nucleotides and includes a 25 nucleotide long poly (CAA) central region).

A 3' UTR may comprise one or more elements endogenous to the nucleotide sequence from which it originates and sometimes includes one or more exogenous elements. A 3' UTR may originate from any suitable nucleic acid, such as genomic DNA, plasmid DNA, RNA or mRNA, for example, from any suitable organism (e.g., a virus, bacterium, yeast, fungi, plant, insect or mammal). The artisan can select appropriate elements for the 3' UTR based upon the chosen expression system (e.g., expression in a chosen organism, for example). A 3' UTR sometimes comprises one or more of the following elements known to the artisan: transcription regulation site, transcription initiation site, transcription termination site, transcription factor binding site, translation regulation site, translation termination site, translation initiation site, translation factor binding site, ribosome binding site, replicon, enhancer element, silencer element and polyadenosine tail. A 3' UTR often includes a polyadenosine tail and sometimes does not, and if a polyadenosine tail is present, one or more adenosine moieties may be added or deleted from it (e.g., about 5, about 10, about 15, about 20, about 25, about 30, about 35, about 40, about 45 or about 50 adenosine moieties may be added or subtracted).

In some embodiments, modification of a 5' UTR and/or a 3' UTR can be used to alter (e.g., increase, add, decrease or substantially eliminate) the activity of a promoter. Alteration of the promoter activity can in turn alter the activity of a peptide, polypeptide or protein (e.g., enzyme activity for example), by a change in transcription of the nucleotide sequence(s) of interest from an operably linked promoter element comprising the modified 5' or 3' UTR. For example, a microorganism can be engineered by genetic modification to express a nucleic acid reagent comprising a modified 5' or 3' UTR that can add a novel activity (e.g., an activity not normally found in the host organism) or increase the expression of an existing activity by increasing transcription from a homologous or heterologous promoter operably linked to a nucleotide sequence of interest (e.g., homologous or heterologous nucleotide sequence of interest), in certain embodiments. In some embodiments, a microorganism can be engineered by genetic modification to express a nucleic acid reagent comprising a modified 5' or 3' UTR that can decrease the expression of an activity by decreasing or substantially eliminating transcription

from a homologous or heterologous promoter operably linked to a nucleotide sequence of interest, in certain embodiments.

Target Nucleotide Sequence

A nucleotide reagent sometimes can comprise a target nucleotide sequence. A "target nucleotide sequence" as used herein encodes a nucleic acid, peptide, polypeptide or protein of interest, and may be a ribonucleotide sequence or a deoxyribonucleotide sequence. A target nucleic acid sometimes is an untranslated ribonucleic acid and sometimes is a translated ribonucleic acid. An untranslated ribonucleic acid may include, but is not limited to, a small interfering ribonucleic acid (siRNA), a short hairpin ribonucleic acid (shRNA), other ribonucleic acid capable of RNA interference (RNAi), an antisense ribonucleic acid, or a ribozyme. A translatable target nucleotide sequence (e.g., a target ribonucleotide sequence) sometimes encodes a peptide, polypeptide or protein, which are sometimes referred to herein as "target peptides," "target polypeptides" or "target proteins."

Any peptides, polypeptides or proteins, or an activity catalyzed by one or more peptides, polypeptides or proteins may be encoded by a target nucleotide sequence and may be selected by a user. Representative proteins include enzymes (e.g., acetyl-CoA carboxylase, acyl-CoA oxidase, thioesterase, monooxygenase, monooxygenase reductase, fatty alcohol oxidase, acyltransferase and the like, for example), antibodies, serum proteins (e.g., albumin), membrane bound proteins, hormones (e.g., growth hormone, erythropoietin, insulin, etc.), cytokines, etc., and include both naturally occurring and exogenously expressed polypeptides. Representative activities (e.g., enzymes or combinations of enzymes which are functionally associated to provide an activity) include thioesterase activity, monooxygenase activity, monooxygenase reductase activity, acyltransferase activity, omega hydroxyl fatty acid dehydrogenase activity, beta-oxidation activity, omega-oxidation activity and the like, for example. The term "enzyme" as used herein refers to a protein which can act as a catalyst to induce a chemical change in other compounds, thereby producing one or more products from one or more substrates.

Specific polypeptides (e.g., enzymes) useful for embodiments described herein are listed herein. The term "protein" as used herein refers to a molecule having a sequence of amino acids linked by peptide bonds. This term includes fusion proteins, oligopeptides, peptides, cyclic peptides, polypeptides and polypeptide derivatives, whether native or recombinant, and also includes

fragments, derivatives, homologs, and variants thereof. A protein or polypeptide sometimes is of intracellular origin (e.g., located in the nucleus, cytosol, or interstitial space of host cells in vivo) and sometimes is a cell membrane protein in vivo. In some embodiments (described above, and in further detail hereafter in Engineering and Alteration Methods), a genetic modification can result in a modification (e.g., increase, substantially increase, decrease or substantially decrease) of a target activity.

A translatable nucleotide sequence generally is located between a start codon (AUG in ribonucleic acids and ATG in deoxyribonucleic acids) and a stop codon (e.g., UAA (ochre), UAG (amber) or UGA (opal) in ribonucleic acids and TAA, TAG or TGA in deoxyribonucleic acids), and sometimes is referred to herein as an "open reading frame" (ORF). A translatable nucleotide sequence (e.g., ORF) sometimes is encoded differently in one organism (e.g., most organisms encode CTG as leucine) than in another organism (e.g., C. tropicalis encodes CTG as serine). In some embodiments, a translatable nucleotide sequence is altered to correct alternate genetic code (e.g., codon usage) differences between a nucleotide donor organism and an nucleotide recipient organism (e.g., engineered organism). In certain embodiments, a translatable nucleotide sequence is altered to improve; (i) codon usage, (ii) transcriptional efficiency, (iii) translational efficiency, (iv) the like, and combinations thereof.

Nucleic Acid Reagents & Tools

A nucleic acid reagent sometimes comprises one or more ORFs. An ORF may be from any suitable source, sometimes from genomic DNA, mRNA, reverse transcribed RNA or complementary DNA (cDNA) or a nucleic acid library comprising one or more of the foregoing, and is from any organism species that contains a nucleic acid sequence of interest, protein of interest, or activity of interest. Non-limiting examples of organisms from which an ORF can be obtained include bacteria, yeast, fungi, human, insect, nematode, bovine, equine, canine, feline, rat or mouse, for example.

A nucleic acid reagent sometimes comprises a nucleotide sequence adjacent to an ORF that is translated in conjunction with the ORF and encodes an amino acid tag. The tag-encoding nucleotide sequence is located 3' and/or 5' of an ORF in the nucleic acid reagent, thereby encoding a tag at the C-terminus or N-terminus of the protein or peptide encoded by the ORF. Any tag that does not abrogate in vitro transcription and/or translation may be utilized and may

be appropriately selected by the artisan. Tags may facilitate isolation and/or purification of the desired ORF product from culture or fermentation media.

A tag sometimes specifically binds a molecule or moiety of a solid phase or a detectable label, for example, thereby having utility for isolating, purifying and/or detecting a protein or peptide encoded by the ORF. In some embodiments, a tag comprises one or more of the following elements: FLAG (e.g., DYKDDDDKG), V5 (e.g., GKPIPNPLLGLDST), c-MYC (e.g., EQKLISEEDL), HSV (e.g., QPELAPEDPED), influenza hemaglutinin, HA (e.g., YPYDVPDYA), VSV-G (e.g., YTDIEMNRLGK), bacterial glutathione-S-transferase, maltose binding protein, a streptavidin- or avidin-binding tag (e.g., pcDNA™6 BioEase™ Gateway® Biotinylation System (Invitrogen)), thioredoxin, β-galactosidase, VSV-glycoprotein, a fluorescent protein (e.g., green fluorescent protein or one of its many color variants (e.g., yellow, red, blue)), a polylysine or polyarginine sequence, a polyhistidine sequence (e.g., His6) or other sequence that chelates a metal (e.g., cobalt, zinc, copper), and/or a cysteine-rich sequence that binds to an arsenic-containing molecule. In certain embodiments, a cysteine-rich tag comprises the amino acid sequence CC-Xn-CC, wherein X is any amino acid and n is 1 to 3, and the cysteine-rich sequence sometimes is CCPGCC. In certain embodiments, the tag comprises a cysteine-rich element and a polyhistidine element (e.g., CCPGCC and His6).

A tag often conveniently binds to a binding partner. For example, some tags bind to an antibody (e.g., FLAG) and sometimes specifically bind to a small molecule. For example, a polyhistidine tag specifically chelates a bivalent metal, such as copper, zinc and cobalt; a polylysine or polyarginine tag specifically binds to a zinc finger; a glutathione S-transferase tag binds to glutathione; and a cysteine-rich tag specifically binds to an arsenic-containing molecule. Arsenic-containing molecules include LUMIO™ agents (Invitrogen, California), such as FlAsH™ (EDT2[4',5*-bis(l,3,2-dithioarsolan-2-yl)fluorescein-(l,2-ethanedithiol)2]) and ReAsH reagents (e.g., U.S. Patent 5,932,474 to Tsien et al, entitled "Target Sequences for Synthetic Molecules;" U.S. Patent 6,054,271 to Tsien et al, entitled "Methods of Using Synthetic Molecules and Target Sequences;" U.S. Patents 6,451,569 and 6,008,378; published U.S. Patent Application 2003/0083373, and published PCT Patent Application WO 99/21013, all to Tsien et al. and all entitled "Synthetic Molecules that Specifically React with Target Sequences"). Such antibodies and small molecules sometimes are linked to a solid phase for convenient isolation of the target protein or target peptide.

A tag sometimes comprises a sequence that localizes a translated protein or peptide to a component in a system, which is referred to as a "signal sequence" or "localization signal sequence" herein. A signal sequence often is incorporated at the N-terminus of a target protein or target peptide, and sometimes is incorporated at the C-terminus. Examples of signal sequences are known to the artisan, are readily incorporated into a nucleic acid reagent, and often are selected according to the organism in which expression of the nucleic acid reagent is performed. A signal sequence in some embodiments localizes a translated protein or peptide to a cell membrane. Examples of signal sequences include, but are not limited to, a nucleus targeting signal (e.g., steroid receptor sequence and N-terminal sequence of SV40 virus large T antigen); mitochondrial targeting signal (e.g., amino acid sequence that forms an amphipathic helix); peroxisome targeting signal (e.g., C-terminal sequence in YFG from S.cerevisiae); and a secretion signal (e.g., N-terminal sequences from invertase, mating factor alpha, PH05 and SUC2 in S.cerevisiae; multiple N-terminal sequences of B. subtilis proteins (e.g., Tjalsma et al, Microbiol.Molec. Biol. Rev. 64: 515-547 (2000)); alpha amylase signal sequence (e.g., U.S. Patent No. 6,288,302); pectate lyase signal sequence (e.g., U.S. Patent No. 5,846,818); precollagen signal sequence (e.g., U.S. Patent No. 5,712,1 14); OmpA signal sequence (e.g., U.S. Patent No. 5,470,719); lam beta signal sequence (e.g., U.S. Patent No. 5,389,529); B. brevis signal sequence (e.g., U.S. Patent No. 5,232,841); and P. pastoris signal sequence (e.g., U.S. Patent No. 5,268,273)).

A tag sometimes is directly adjacent to the amino acid sequence encoded by an ORF (i.e., there is no intervening sequence) and sometimes a tag is substantially adjacent to an ORF encoded amino acid sequence (e.g., an intervening sequence is present). An intervening sequence sometimes includes a recognition site for a protease, which is useful for cleaving a tag from a target protein or peptide. In some embodiments, the intervening sequence is cleaved by Factor Xa (e.g., recognition site I (E/D)GR), thrombin (e.g., recognition site LVPRGS), enterokinase (e.g., recognition site DDDDK), TEV protease (e.g., recognition site ENLYFQG) or PreScission™ protease (e.g., recognition site LEVLFQGP), for example.

An intervening sequence sometimes is referred to herein as a "linker sequence," and may be of any suitable length selected by the artisan. A linker sequence sometimes is about 1 to about 20 amino acids in length, and sometimes about 5 to about 10 amino acids in length. The artisan may select the linker length to substantially preserve target protein or peptide function (e.g., a tag may reduce target protein or peptide function unless separated by a linker), to enhance

disassociation of a tag from a target protein or peptide when a protease cleavage site is present (e.g., cleavage may be enhanced when a linker is present), and to enhance interaction of a tag/target protein product with a solid phase. A linker can be of any suitable amino acid content, and often comprises a higher proportion of amino acids having relatively short side chains (e.g., glycine, alanine, serine and threonine).

A nucleic acid reagent sometimes includes a stop codon between a tag element and an insertion element or ORF, which can be useful for translating an ORF with or without the tag. Mutant tRNA molecules that recognize stop codons (described above) suppress translation termination and thereby are designated "suppressor tRNAs." Suppressor tRNAs can result in the insertion of amino acids and continuation of translation past stop codons (e.g., U.S. Patent Application No. 60/587,583, filed July 14, 2004, entitled "Production of Fusion Proteins by Cell-Free Protein Synthesis,"; Eggertsson, et al, (1988) Microbiological Review 52(3):354-374, and Engleerg-Kukla, et al. (1996) in Escherichia coli and Salmonella Cellular and Molecular Biology, Chapter 60, pps 909-921, Neidhardt, et al. eds., ASM Press, Washington, DC). A number of suppressor tRNAs are known, including but not limited to, supE, supP, supD, supF and supZ suppressors, which suppress the termination of translation of the amber stop codon; supB, glT, supL, supN, supC and supM suppressors, which suppress the function of the ochre stop codon and glyT, trpT and Su-9 suppressors, which suppress the function of the opal stop codon. In general, suppressor tRNAs contain one or more mutations in the anti-codon loop of the tRNA that allows the tRNA to base pair with a codon that ordinarily functions as a stop codon. The mutant tRNA is charged with its cognate amino acid residue and the cognate amino acid residue is inserted into the translating polypeptide when the stop codon is encountered. Mutations that enhance the efficiency of termination suppressors (i.e., increase stop codon read-through) have been identified. These include, but are not limited to, mutations in the uar gene (also known as the prfA gene), mutations in the ups gene, mutations in the sueA, sueB and sueC genes, mutations in the rpsD (ramA) and rpsE (spcA) genes and mutations in the rplL gene.

Thus, a nucleic acid reagent comprising a stop codon located between an ORF and a tag can yield a translated ORF alone when no suppressor tRNA is present in the translation system, and can yield a translated ORF-tag fusion when a suppressor tRNA is present in the system. Suppressor tRNA can be generated in cells transfected with a nucleic acid encoding the tRNA (e.g., a replication incompetent adenovirus containing the human tRNA-Ser suppressor gene can be transfected into cells, or a YAC containing a yeast or bacterial tRNA suppressor gene

can be transfected into yeast cells, for example). Vectors for synthesizing suppressor tRNA and for translating ORFs with or without a tag are available to the artisan (e.g., Tag-On-Demand™ kit (Invitrogen Corporation, California); Tag-On-Demand™ Suppressor Supernatant Instruction Manual, Version B, 6 June 2003, at http address www.invitrogen.com/content/sfs/ manuals/tagondemand _supernatant_man.pdf; Tag-On-Demand™ Gateway® Vector Instruction Manual, Version B, 20 June, 2003 at http address www.invitrogen.com/content/sfs/ manuals/tagondemand _vectors_man.pdf; and Capone et al, Amber, ochre and opal suppressor tRNA genes derived from a human serine tRNA gene. EMBO J. 4:213, 1985).

Any convenient cloning strategy known in the art may be utilized to incorporate an element, such as an ORF, into a nucleic acid reagent. Known methods can be utilized to insert an element into the template independent of an insertion element, such as (1) cleaving the template at one or more existing restriction enzyme sites and ligating an element of interest and (2) adding restriction enzyme sites to the template by hybridizing oligonucleotide primers that include one or more suitable restriction enzyme sites and amplifying by polymerase chain reaction (described in greater detail herein). Other cloning strategies take advantage of one or more insertion sites present or inserted into the nucleic acid reagent, such as an oligonucleotide primer hybridization site for PCR, for example, and others described herein. In some embodiments, a cloning strategy can be combined with genetic manipulation such as recombination (e.g., recombination of a nucleic acid reagent with a nucleic acid sequence of interest into the genome of the organism to be modified, as described further herein). In some embodiments, the cloned ORF(s) can produce (directly or indirectly) a fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid), by engineering a microorganism with one or more ORFs of interest, which microorganism comprises one or more altered activities selected from the group consisting of omega hydroxyl fatty acid dehydrogenase activity, acyl-CoA oxidase activity, acyltransferase activity, thioesterase activity, monooxygenase activity and monooxygenase reductase activity.

In some embodiments, the nucleic acid reagent includes one or more recombinase insertion sites. A recombinase insertion site is a recognition sequence on a nucleic acid molecule that participates in an integration/recombination reaction by recombination proteins. For example, the recombination site for Cre recombinase is loxP, which is a 34 base pair sequence comprised of two 13 base pair inverted repeats (serving as the recombinase binding sites) flanking an 8

base pair core sequence (e.g., Figure 1 of Sauer, B., Curr. Opin. Biotech. 5:521-527 (1994)). Other examples of recombination sites include attB, attP, attL, and attR sequences, and mutants, fragments, variants and derivatives thereof, which are recognized by the recombination protein λ Int and by the auxiliary proteins integration host factor (IHF), FIS and excisionase (Xis) (e.g., U.S. Patent Nos. 5,888,732; 6,143,557; 6,171,861; 6,270,969; 6,277,608; and 6,720,140; U.S. Patent Appln. Nos. 09/517,466, filed March 2, 2000, and 09/732,914, filed August 14, 2003, and in U.S. patent publication no. 2002-000705 1-Al; Landy, Curr. Opin. Biotech. 3:699-707 (1993)).

Examples of recombinase cloning nucleic acids are in Gateway® systems (Invitrogen, California), which include at least one recombination site for cloning a desired nucleic acid molecules in vivo or in vitro. In some embodiments, the system utilizes vectors that contain at least two different site-specific recombination sites, often based on the bacteriophage lambda system (e.g., attl and att2), and are mutated from the wild-type (attO) sites. Each mutated site has a unique specificity for its cognate partner att site (i.e., its binding partner recombination site) of the same type (for example attBl with attPl, or attLl with attRl) and will not cross-react with recombination sites of the other mutant type or with the wild-type attO site. Different site specificities allow directional cloning or linkage of desired molecules thus providing desired orientation of the cloned molecules. Nucleic acid fragments flanked by recombination sites are cloned and subcloned using the Gateway® system by replacing a selectable marker (for example, ccdB) flanked by att sites on the recipient plasmid molecule, sometimes termed the Destination Vector. Desired clones are then selected by transformation of a ccdB sensitive host strain and positive selection for a marker on the recipient molecule. Similar strategies for negative selection (e.g., use of toxic genes) can be used in other organisms such as thymidine kinase (TK) in mammals and insects.

A recombination system useful for engineering yeast is outlined briefly. The system makes use of the URA3 gene (e.g., for S. cerevisiae and C. albicans, for example) or URA4 and URA5 genes (e.g., for S. pombe, for example) and toxicity of the nucleotide analogue 5-Fluoroorotic acid (5-FOA). The URA3 or URA4 and URA5 genes encode orotine-5' -monophosphate (OMP) dicarboxylase. Yeast with an active URA3 or URA4 and URA5 gene (phenotypically Ura+) convert 5-FOA to fluorodeoxyuridine, which is toxic to yeast cells. Yeast carrying a mutation in the appropriate gene(s) or having a knock out of the appropriate gene(s) can grow in the presence of 5-FOA, if the media is also supplemented with uracil.

A nucleic acid engineering construct can be made which may comprise the URA3 gene or cassette (for S. cerevisiae), flanked on either side by the same nucleotide sequence in the same orientation. The URA3 cassette comprises a promoter, the URA3 gene and a functional transcription terminator. Target sequences which direct the construct to a particular nucleic acid region of interest in the organism to be engineered are added such that the target sequences are adjacent to and abut the flanking sequences on either side of the URA3 cassette. Yeast can be transformed with the engineering construct and plated on minimal media without uracil. Colonies can be screened by PCR to determine those transformants that have the engineering construct inserted in the proper location in the genome. Checking insertion location prior to selecting for recombination of the ura3 cassette may reduce the number of incorrect clones carried through to later stages of the procedure. Correctly inserted transformants can then be replica plated on minimal media containing 5-FOA to select for recombination of the URA3 cassette out of the construct, leaving a disrupted gene and an identifiable footprint (e.g., nucleic acid sequence) that can be use to verify the presence of the disrupted gene. The technique described is useful for disrupting or "knocking out" gene function, but also can be used to insert genes or constructs into a host organisms genome in a targeted, sequence specific manner.

In certain embodiments, a nucleic acid reagent includes one or more topoisomerase insertion sites. A topoisomerase insertion site is a defined nucleotide sequence recognized and bound by a site-specific topoisomerase. For example, the nucleotide sequence 5'-(C/T)CCTT-3' is a topoisomerase recognition site bound specifically by most poxvirus topoisomerases, including vaccinia virus DNA topoisomerase I. After binding to the recognition sequence, the topoisomerase cleaves the strand at the 3'-most thymidine of the recognition site to produce a nucleotide sequence comprising 5'-(C/T)CCTT-P04-TOPO, a complex of the topoisomerase covalently bound to the 3' phosphate via a tyrosine in the topoisomerase (e.g., Shuman, J. Biol. Chem. 266:1 1372-1 1379, 1991; Sekiguchi and Shuman, Nucl. Acids Res. 22:5360-5365, 1994; U.S. Pat. No. 5,766,891; PCT/US95/16099; and PCT/US98/12372). In comparison, the nucleotide sequence 5'-GCAACTT-3' is a topoisomerase recognition site for type IA E. coli topoisomerase III. An element to be inserted often is combined with topoisomerase-reacted template and thereby incorporated into the nucleic acid reagent (e.g., World Wide Web URL invitrogen.com/downloads/F-13512_Topo_Flyer.pdf; World Wide Web URL invitrogen. com/content/sfs/brochures/

710_021849%20_B_TOPOCloning_bro.pdf; TOPO TA Cloning® Kit and Zero Blunt® TOPO® Cloning Kit product information).

A nucleic acid reagent sometimes contains one or more origin of replication (ORI) elements. In some embodiments, a template comprises two or more ORIs, where one functions efficiently in one organism (e.g., a bacterium) and another functions efficiently in another organism (e.g., a eukaryote, like yeast for example). In some embodiments, an ORI may function efficiently in one species (e.g., S. cerevisiae, for example) and another ORI may function efficiently in a different species (e.g., S. pombe, for example). A nucleic acid reagent also sometimes includes one or more transcription regulation sites.

A nucleic acid reagent can include one or more selection elements (e.g., elements for selection of the presence of the nucleic acid reagent, and not for activation of a promoter element which can be selectively regulated). Selection elements often are utilized using known processes to determine whether a nucleic acid reagent is included in a cell. In some embodiments, a nucleic acid reagent includes two or more selection elements, where one functions efficiently in one organism and another functions efficiently in another organism. Examples of selection elements include, but are not limited to, (1) nucleic acid segments that encode products that provide resistance against otherwise toxic compounds (e.g., antibiotics); (2) nucleic acid segments that encode products that are otherwise lacking in the recipient cell (e.g., essential products, tRNA genes, auxotrophic markers); (3) nucleic acid segments that encode products that suppress the activity of a gene product; (4) nucleic acid segments that encode products that can be readily identified (e.g., phenotypic markers such as antibiotics (e.g., β-lactamase), β-galactosidase, green fluorescent protein (GFP), yellow fluorescent protein (YFP), red fluorescent protein (RFP), cyan fluorescent protein (CFP), and cell surface proteins); (5) nucleic acid segments that bind products that are otherwise detrimental to cell survival and/or function; (6) nucleic acid segments that otherwise inhibit the activity of any of the nucleic acid segments described inNos. 1-5 above (e.g., antisense oligonucleotides); (7) nucleic acid segments that bind products that modify a substrate (e.g., restriction endonucleases); (8) nucleic acid segments that can be used to isolate or identify a desired molecule (e.g., specific protein binding sites); (9) nucleic acid segments that encode a specific nucleotide sequence that can be otherwise non-functional (e.g., for PCR amplification of subpopulations of molecules); (10) nucleic acid segments that, when absent, directly or indirectly confer resistance or sensitivity to particular compounds; (11) nucleic acid segments that encode products that either are toxic or

convert a relatively non-toxic compound to a toxic compound (e.g., Herpes simplex thymidine kinase, cytosine deaminase) in recipient cells; (12) nucleic acid segments that inhibit replication, partition or heritability of nucleic acid molecules that contain them; and/or (13) nucleic acid segments that encode conditional replication functions, e.g., replication in certain hosts or host cell strains or under certain environmental conditions (e.g., temperature, nutritional conditions, and the like).

A nucleic acid reagent is of any form useful for in vivo transcription and/or translation. A nucleic acid sometimes is a plasmid, such as a supercoiled plasmid, sometimes is a yeast artificial chromosome (e.g., YAC), sometimes is a linear nucleic acid (e.g., a linear nucleic acid produced by PCR or by restriction digest), sometimes is single-stranded and sometimes is double-stranded. A nucleic acid reagent sometimes is prepared by an amplification process, such as a polymerase chain reaction (PCR) process or transcription-mediated amplification process (TMA). In TMA, two enzymes are used in an isothermal reaction to produce amplification products detected by light emission (see, e.g., Biochemistry 1996 Jun 25;35(25):8429-38 and http address www.devicelink.com/ivdt/archive/00/l l/007.html). Standard PCR processes are known (e.g., U. S. Patent Nos. 4,683,202; 4,683,195; 4,965,188; and 5,656,493), and generally are performed in cycles. Each cycle includes heat denaturation, in which hybrid nucleic acids dissociate; cooling, in which primer oligonucleotides hybridize; and extension of the oligonucleotides by a polymerase (i.e., Taq polymerase). An example of a PCR cyclical process is treating the sample at 95□C for 5 minutes; repeating forty-five cycles of 95□C for 1 minute, 59□C for 1 minute, 10 seconds, and 72□C for 1 minute 30 seconds; and then treating the sample at 72□C for 5 minutes. Multiple cycles frequently are performed using a commercially available thermal cycler. PCR amplification products sometimes are stored for a time at a lower temperature (e.g., at 4□C) and sometimes are frozen (e.g., at -20 □C) before analysis.

In some embodiments, a nucleic acid reagent, protein reagent, protein fragment reagent or other reagent described herein is isolated or purified. The term "isolated" as used herein refers to material removed from its original environment (e.g., the natural environment if it is naturally occurring, or a host cell if expressed exogenously), and thus is altered "by the hand of man" from its original environment. The term "purified" as used herein with reference to molecules does not refer to absolute purity. Rather, "purified" refers to a substance in a composition that contains fewer substance species in the same class (e.g., nucleic acid or protein species) other

than the substance of interest in comparison to the sample from which it originated. "Purified," if a nucleic acid or protein for example, refers to a substance in a composition that contains fewer nucleic acid species or protein species other than the nucleic acid or protein of interest in comparison to the sample from which it originated. Sometimes, a protein or nucleic acid is "substantially pure," indicating that the protein or nucleic acid represents at least 50% of protein or nucleic acid on a mass basis of the composition. Often, a substantially pure protein or nucleic acid is at least 75% on a mass basis of the composition, and sometimes at least 95% on a mass basis of the composition.

Engineering and Alteration Methods

Methods and compositions (e.g., nucleic acid reagents) described herein can be used to generate engineered microorganisms. As noted above, the term "engineered microorganism" as used herein refers to a modified organism that includes one or more activities distinct from an activity present in a microorganism utilized as a starting point for modification (e.g., host microorganism or unmodified organism). Engineered microorganisms typically arise as a result of a genetic modification, usually introduced or selected for, by one of skill in the art using readily available techniques. Non-limiting examples of methods useful for generating an altered activity include, introducing a heterologous polynucleotide (e.g., nucleic acid or gene integration, also referred to as "knock in"), removing an endogenous polynucleotide, altering the sequence of an existing endogenous nucleic acid sequence (e.g., site-directed mutagenesis), disruption of an existing endogenous nucleic acid sequence (e.g., knock outs and transposon or insertion element mediated mutagenesis), selection for an altered activity where the selection causes a change in a naturally occurring activity that can be stably inherited (e.g., causes a change in a nucleic acid sequence in the genome of the organism or in an epigenetic nucleic acid that is replicated and passed on to daughter cells), PCR-based mutagenesis, and the like. The term "mutagenesis" as used herein refers to any modification to a nucleic acid (e.g., nucleic acid reagent, or host chromosome, for example) that is subsequently used to generate a product in a host or modified organism. Non-limiting examples of mutagenesis include, deletion, insertion, substitution, rearrangement, point mutations, suppressor mutations and the like. Mutagenesis methods are known in the art and are readily available to the artisan. Non-limiting examples of mutagenesis methods are described herein and can also be found in Maniatis, T., E. F. Fritsch and J. Sambrook (1982) *Molecular Cloning: a Laboratory Manual;* Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. Another non-limiting example of mutagenesis

can be conducted using a Stratagene (San Diego, CA) "QuickChange" kit according to the manufacturer's instructions.

The term "genetic modification" as used herein refers to any suitable nucleic acid addition, removal or alteration that facilitates production of a target fatty dicarboxylic acid product (e.g., sebacic or dodecanedioic acid) in an engineered microorganism. Genetic modifications include, without limitation, insertion of one or more nucleotides in a native nucleic acid of a host organism in one or more locations, deletion of one or more nucleotides in a native nucleic acid of a host organism in one or more locations, modification or substitution of one or more nucleotides in a native nucleic acid of a host organism in one or more locations, insertion of a non-native nucleic acid into a host organism (e.g., insertion of an autonomously replicating vector), and removal of a non-native nucleic acid in a host organism (e.g., removal of a vector).

The term "heterologous polynucleotide" as used herein refers to a nucleotide sequence not present in a host microorganism in some embodiments. In certain embodiments, a heterologous polynucleotide is present in a different amount (e.g., different copy number) than in a host microorganism, which can be accomplished, for example, by introducing more copies of a particular nucleotide sequence to a host microorganism (e.g., the particular nucleotide sequence may be in a nucleic acid autonomous of the host chromosome or may be inserted into a chromosome). A heterologous polynucleotide is from a different organism in some embodiments, and in certain embodiments, is from the same type of organism but from an outside source (e.g., a recombinant source).

In some embodiments, an organism engineered using the methods and nucleic acid reagents described herein can produce a fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid). In certain embodiments, an engineered microorganism described herein that produces a fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid) may comprise one ore more altered activities selected from the group consisting of omega oxo fatty acid dehydrogenase activity, omega hydroxyl fatty acid dehydrogenase activity, fatty acid synthase activity, acetyl CoA carboxylase activity, acyl-CoA oxidase activity, monooxygenase activity and monooxygenase reductase activity. In some embodiments, an engineered microorganism as described herein may comprise a genetic modification that adds or increases

the omega oxo fatty acid dehydrogenase activity, omega hydroxyl fatty acid dehydrogenase activity, fatty acid synthase activity, acetyl CoA carboxylase activity, acyl-CoA oxidase activity, monooxygenase activity and monooxygenase reductase activity.

In certain embodiments, an engineered microorganism described herein can comprise an altered thioesterase activity. In some embodiments, the engineered microorganism may comprise a genetic alteration that adds or increases a thioesterase activity. In some embodiments, the engineered microorganism comprising a genetic alteration that adds or increases a thioesterase activity, may further comprise a heterologous polynucleotide encoding a polypeptide having thioesterase activity.

The term "altered activity" as used herein refers to an activity in an engineered microorganism that is added or modified relative to the host microorganism (e.g., added, increased, reduced, inhibited or removed activity). An activity can be altered by introducing a genetic modification to a host microorganism that yields an engineered microorganism having added, increased, reduced, inhibited or removed activity.

An added activity often is an activity not detectable in a host microorganism. An increased activity generally is an activity detectable in a host microorganism that has been increased in an engineered microorganism. An activity can be increased to any suitable level for production of a target fatty dicarboxylic acid product (e.g., sebacic or dodecanedioic acid), including but not limited to less than 2-fold (e.g., about 10% increase to about 99% increase; about 20%>, 30%>, 40%, 50%, 60%, 70%, 80%, 90% increase), 2-fold, 3-fold, 4-fold, 5-fold, 6-fold, 7-fold, 8-fold, 9-fold, of 10-fold increase, or greater than about 10-fold increase. A reduced or inhibited activity generally is an activity detectable in a host microorganism that has been reduced or inhibited in an engineered microorganism. An activity can be reduced to undetectable levels in some embodiments, or detectable levels in certain embodiments. An activity can be decreased to any suitable level for production of a target fatty dicarboxylic acid product (e.g., sebacic or dodecanedioic acid), including but not limited to less than 2-fold (e.g., about 10%> decrease to about 99% decrease; about 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% decrease), 2-fold, 3-fold, 4-fold, 5-fold, 6-fold, 7-fold, 8-fold, 9-fold, of 10-fold decrease, or greater than about 10-fold decrease.

The term "level", as used herein, often refers to a level or amount (e.g., a quantitative or relative amount) of a protein or nucleic acid (e.g. an RNA (e.g. an mRNA) or DNA).

An altered activity sometimes is an activity not detectable in a host organism and is added to an engineered organism. An altered activity also may be an activity detectable in a host organism and is increased in an engineered organism. An activity may be added or increased by increasing the number of copies of a polynucleotide that encodes a polypeptide having a target activity, in some embodiments. In some embodiments, the activity of a native polypeptide can be increased by increasing in the modified organism the number of copies of a polynucleotide that encodes the polypeptide (e.g., introducing 1 to about 100 additional copies of the polynucleotide (e.g., introducing 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 22, 24, 26, 28, 30 or more additional copies of the polynucleotide). In certain embodiments an activity can be added or increased by inserting into a host microorganism a polynucleotide that encodes a heterologous polypeptide having the added activity or encodes a modified endogenous polypeptide. In such embodiments, 1 to about 100 copies of the polynucleotide can be introduced (e.g., introducing 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 22, 24, 26, 28, 30 copies). A "modified endogenous polypeptide" often has an activity different than an activity of a native polypeptide counterpart (e.g., different catalytic activity and/or different substrate specificity), and often is active (e.g., an activity (e.g., substrate turnover) is detectable). In certain embodiments, an activity can be added or increased by inserting into a host microorganism a heterologous polynucleotide that is (i) operably linked to another polynucleotide that encodes a polypeptide having the added activity, and (ii) up regulates production of the polynucleotide. Thus, an activity can be added or increased by inserting or modifying a regulatory polynucleotide operably linked to another polynucleotide that encodes a polypeptide having the target activity. In certain embodiments, an activity can be added or increased by subjecting a host microorganism to a selective environment and screening for microorganisms that have a detectable level of the target activity. Examples of a selective environment include, without limitation, a medium containing a substrate that a host organism can process and a medium lacking a substrate that a host organism can process.

An altered activity sometimes is an activity detectable in a host organism and is reduced, inhibited or removed (i.e., not detectable) in an engineered organism. An activity may be reduced or removed by decreasing the number of copies of a polynucleotide that encodes a polypeptide having a target activity, in some embodiments. In some embodiments, an activity

can be reduced or removed by (i) inserting a polynucleotide within a polynucleotide that encodes a polypeptide having the target activity (disruptive insertion), and/or (ii) removing a portion of or all of a polynucleotide that encodes a polypeptide having the target activity (deletion or knock out, respectively). In certain embodiments, an activity can be reduced or removed by inserting into a host microorganism a heterologous polynucleotide that is (i) operably linked to another polynucleotide that encodes a polypeptide having the target activity, and (ii) down regulates production of the polynucleotide. Thus, an activity can be reduced or removed by inserting or modifying a regulatory polynucleotide operably linked to another polynucleotide that encodes a polypeptide having the target activity.

An activity also can be reduced or removed by (i) inhibiting a polynucleotide that encodes a polypeptide having the activity or (ii) inhibiting a polynucleotide operably linked to another polynucleotide that encodes a polypeptide having the activity. A polynucleotide can be inhibited by a suitable technique known in the art, such as by contacting an RNA encoded by the polynucleotide with a specific inhibitory RNA (e.g., RNAi, siRNA, ribozyme). An activity also can be reduced or removed by contacting a polypeptide having the activity with a molecule that specifically inhibits the activity (e.g., enzyme inhibitor, antibody). In certain embodiments, an activity can be reduced or removed by subjecting a host microorganism to a selective environment and screening for microorganisms that have a reduced level or removal of the target activity.

In some embodiments, an untranslated ribonucleic acid, or a cDNA can be used to reduce the expression of a particular activity or enzyme. For example, a microorganism can be engineered by genetic modification to express a nucleic acid reagent that reduces the expression of an activity by producing an RNA molecule that is partially or substantially homologous to a nucleic acid sequence of interest which encodes the activity of interest. The RNA molecule can bind to the nucleic acid sequence of interest and inhibit the nucleic acid sequence from performing its natural function, in certain embodiments. In some embodiments, the RNA may alter the nucleic acid sequence of interest which encodes the activity of interest in a manner that the nucleic acid sequence of interest is no longer capable of performing its natural function (e.g., the action of a ribozyme for example).

In certain embodiments, nucleotide sequences sometimes are added to, modified or removed from one or more of the nucleic acid reagent elements, such as the promoter, 5'UTR, target

sequence, or 3'UTR elements, to enhance, potentially enhance, reduce, or potentially reduce transcription and/or translation before or after such elements are incorporated in a nucleic acid reagent. In some embodiments, one or more of the following sequences may be modified or removed if they are present in a 5'UTR: a sequence that forms a stable secondary structure (e.g., quadruplex structure or stem loop stem structure (e.g., EMBL sequences X12949, AF274954, AF139980, AF152961, S95936, U194144, AF1 16649 or substantially identical sequences that form such stem loop stem structures)); a translation initiation codon upstream of the target nucleotide sequence start codon; a stop codon upstream of the target nucleotide sequence translation initiation codon; an ORF upstream of the target nucleotide sequence translation initiation codon; an iron responsive element (IRE) or like sequence; and a 5' terminal oligopyrimidine tract (TOP, e.g., consisting of 5-15 pyrimidines adjacent to the cap). A translational enhancer sequence and/or an internal ribosome entry site (IRES) sometimes is inserted into a 5'UTR (e.g., EMBL nucleotide sequences J04513, X87949, M95825, M12783, AF025841, AF013263, AF006822, M17169, M13440, M22427, D14838 and M17446 and substantially identical nucleotide sequences).

An AU-rich element (ARE, e.g., AUUUA repeats) and/or splicing junction that follows a non-sense codon sometimes is removed from or modified in a 3'UTR. A polyadenosine tail sometimes is inserted into a 3'UTR if none is present, sometimes is removed if it is present, and adenosine moieties sometimes are added to or removed from a polyadenosine tail present in a 3'UTR. Thus, some embodiments are directed to a process comprising: determining whether any nucleotide sequences that increase, potentially increase, reduce or potentially reduce translation efficiency are present in the elements, and adding, removing or modifying one or more of such sequences if they are identified. Certain embodiments are directed to a process comprising: determining whether any nucleotide sequences that increase or potentially increase translation efficiency are not present in the elements, and incorporating such sequences into the nucleic acid reagent.

In some embodiments, an activity can be altered by modifying the nucleotide sequence of an ORF. An ORF sometimes is mutated or modified (for example, by point mutation, deletion mutation, insertion mutation, PCR based mutagenesis and the like) to alter, enhance or increase, reduce, substantially reduce or eliminate the activity of the encoded protein or peptide. The protein or peptide encoded by a modified ORF sometimes is produced in a lower amount or may not be produced at detectable levels, and in other embodiments, the product or protein

encoded by the modified ORF is produced at a higher level (e.g., codons sometimes are modified so they are compatible with tRNA's preferentially used in the host organism or engineered organism). To determine the relative activity, the activity from the product of the mutated ORF (or cell containing it) can be compared to the activity of the product or protein encoded by the unmodified ORF (or cell containing it).

In some embodiments, an ORF nucleotide sequence sometimes is mutated or modified to alter the triplet nucleotide sequences used to encode amino acids (e.g., amino acid codon triplets, for example). Modification of the nucleotide sequence of an ORF to alter codon triplets sometimes is used to change the codon found in the original sequence to better match the preferred codon usage of the organism in which the ORF or nucleic acid reagent will be expressed. The codon usage, and therefore the codon triplets encoded by a nucleic acid sequence, in bacteria may be different from the preferred codon usage in eukaryotes, like yeast or plants for example. Preferred codon usage also may be different between bacterial species. In certain embodiments an ORF nucleotide sequences sometimes is modified to eliminate codon pairs and/or eliminate mRNA secondary structures that can cause pauses during translation of the mRNA encoded by the ORF nucleotide sequence. Translational pausing sometimes occurs when nucleic acid secondary structures exist in an mRNA, and sometimes occurs due to the presence of codon pairs that slow the rate of translation by causing ribosomes to pause. In some embodiments, the use of lower abundance codon triplets can reduce translational pausing due to a decrease in the pause time needed to load a charged tRNA into the ribosome translation machinery. Therefore, to increase transcriptional and translational efficiency in bacteria (e.g., where transcription and translation are concurrent, for example) or to increase translational efficiency in eukaryotes (e.g., where transcription and translation are functionally separated), the nucleotide sequence of a nucleotide sequence of interest can be altered to better suit the transcription and/or translational machinery of the host and/or genetically modified microorganism. In certain embodiments, slowing the rate of translation by the use of lower abundance codons, which slow or pause the ribosome, can lead to higher yields of the desired product due to an increase in correctly folded proteins and a reduction in the formation of inclusion bodies.

Codons can be altered and optimized according to the preferred usage by a given organism by determining the codon distribution of the nucleotide sequence donor organism and comparing the distribution of codons to the distribution of codons in the recipient or host organism. Techniques described herein (e.g., site directed mutagenesis and the like) can then be used to

alter the codons accordingly. Comparisons of codon usage can be done by hand, or using nucleic acid analysis software commercially available to the artisan.

Modification of the nucleotide sequence of an ORF also can be used to correct codon triplet sequences that have diverged in different organisms. For example, certain yeast (e.g., C. tropicalis and C. maltosa) use the amino acid triplet CUG (e.g., CTG in the DNA sequence) to encode serine. CUG typically encodes leucine in most organisms. In order to maintain the correct amino acid in the resultant polypeptide or protein, the CUG codon must be altered to reflect the organism in which the nucleic acid reagent will be expressed. Thus, if an ORF from a bacterial donor is to be expressed in either Candida yeast strain mentioned above, the heterologous nucleotide sequence must first be altered or modified to the appropriate leucine codon. Therefore, in some embodiments, the nucleotide sequence of an ORF sometimes is altered or modified to correct for differences that have occurred in the evolution of the amino acid codon triplets between different organisms. In some embodiments, the nucleotide sequence can be left unchanged at a particular amino acid codon, if the amino acid encoded is a conservative or neutral change in amino acid when compared to the originally encoded amino acid.

In some embodiments, an activity can be altered by modifying translational regulation signals, like a stop codon for example. A stop codon at the end of an ORF sometimes is modified to another stop codon, such as an amber stop codon described above. In some embodiments, a stop codon is introduced within an ORF, sometimes by insertion or mutation of an existing codon. An ORF comprising a modified terminal stop codon and/or internal stop codon often is translated in a system comprising a suppressor tRNA that recognizes the stop codon. An ORF comprising a stop codon sometimes is translated in a system comprising a suppressor tRNA that incorporates an unnatural amino acid during translation of the target protein or target peptide. Methods for incorporating unnatural amino acids into a target protein or peptide are known, which include, for example, processes utilizing a heterologous tRNA/synthetase pair, where the tRNA recognizes an amber stop codon and is loaded with an unnatural amino acid (e.g., World Wide Web URL iupac.org/news/prize/2003/wang.pdf).

Depending on the portion of a nucleic acid reagent (e.g., Promoter, 5' or 3' UTR, ORI, ORF, and the like) chosen for alteration (e.g., by mutagenesis, introduction or deletion, for example) the modifications described above can alter a given activity by (i) increasing or decreasing

feedback inhibition mechanisms, (ii) increasing or decreasing promoter initiation, (iii) increasing or decreasing translation initiation, (iv) increasing or decreasing translational efficiency, (v) modifying localization of peptides or products expressed from nucleic acid reagents described herein, or (vi) increasing or decreasing the copy number of a nucleotide sequence of interest, (vii) expression of an anti-sense RNA, RNAi, siRNA, ribozyme and the like. In some embodiments, alteration of a nucleic acid reagent or nucleotide sequence can alter a region involved in feedback inhibition (e.g., 5' UTR, promoter and the like). A modification sometimes is made that can add or enhance binding of a feedback regulator and sometimes a modification is made that can reduce, inhibit or eliminate binding of a feedback regulator.

In certain embodiments, alteration of a nucleic acid reagent or nucleotide sequence can alter sequences involved in transcription initiation (e.g., promoters, 5' UTR, and the like). A modification sometimes can be made that can enhance or increase initiation from an endogenous or heterologous promoter element. A modification sometimes can be made that removes or disrupts sequences that increase or enhance transcription initiation, resulting in a decrease or elimination of transcription from an endogenous or heterologous promoter element.

In some embodiments, alteration of a nucleic acid reagent or nucleotide sequence can alter sequences involved in translational initiation or translational efficiency (e.g., 5' UTR, 3' UTR, codon triplets of higher or lower abundance, translational terminator sequences and the like, for example). A modification sometimes can be made that can increase or decrease translational initiation, modifying a ribosome binding site for example. A modification sometimes can be made that can increase or decrease translational efficiency. Removing or adding sequences that form hairpins and changing codon triplets to a more or less preferred codon are non-limiting examples of genetic modifications that can be made to alter translation initiation and translation efficiency.

In certain embodiments, alteration of a nucleic acid reagent or nucleotide sequence can alter sequences involved in localization of peptides, proteins or other desired products (e.g., a sebacic acid or dodecanedioic acid, for example). A modification sometimes can be made that can alter, add or remove sequences responsible for targeting a polypeptide, protein or product to an intracellular organelle, the periplasm, cellular membranes, or extracellularly. Transport of a heterologous product to a different intracellular space or extracellularly sometimes can reduce

or eliminate the formation of inclusion bodies (e.g., insoluble aggregates of the desired product).

In some embodiments, alteration of a nucleic acid reagent or nucleotide sequence can alter sequences involved in increasing or decreasing the copy number of a nucleotide sequence of interest. A modification sometimes can be made that increases or decreases the number of copies of an ORF stably integrated into the genome of an organism or on an epigenetic nucleic acid reagent. Non-limiting examples of alterations that can increase the number of copies of a sequence of interest include, adding copies of the sequence of interest by duplication of regions in the genome (e.g., adding additional copies by recombination or by causing gene amplification of the host genome, for example), cloning additional copies of a sequence onto a nucleic acid reagent, or altering an ORI to increase the number of copies of an epigenetic nucleic acid reagent. Non-limiting examples of alterations that can decrease the number of copies of a sequence of interest include, removing copies of the sequence of interest by deletion or disruption of regions in the genome, removing additional copies of the sequence from epigenetic nucleic acid reagents, or altering an ORI to decrease the number of copies of an epigenetic nucleic acid reagent.

In certain embodiments, increasing or decreasing the expression of a nucleotide sequence of interest can also be accomplished by altering, adding or removing sequences involved in the expression of an anti-sense RNA, RNAi, siRNA, ribozyme and the like. The methods described above can be used to modify expression of anti-sense RNA, RNAi, siRNA, ribozyme and the like.

The methods and nucleic acid reagents described herein can be used to generate genetically modified microorganisms with altered activities in cellular processes involved in a fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid) synthesis. In some embodiments, an engineered microorganism described herein may comprise an increased number of copies of an endogenous polynucleotide encoding a polypeptide having omega oxo fatty acid dehydrogenase activity. In certain embodiments, an engineered microorganism described herein may comprise an increased number of copies of an endogenous polynucleotide encoding a polypeptide having omega hydroxyl fatty acid dehydrogenase activity. In some embodiments, an engineered microorganism described herein may comprise a heterologous

polynucleotide encoding a polypeptide having omega oxo fatty acid dehydrogenase activity. In some embodiments, an engineered microorganism described herein may comprise a heterologous polynucleotide encoding a polypeptide having omega hydroxyl fatty acid dehydrogenase activity. In some embodiments, the heterologous polynucleotide can be from a bacterium. In some embodiments, the bacterium can be an Acinetobacter, Nocardia, Pseudomonas or Xanthobacter bacterium.

In some embodiments, an engineered microorganism described herein may comprise a heterologous polynucleotide encoding a polypeptide having monooxygenase activity. In certain embodiments, the heterologous polynucleotide can be from a bacterium. In some embodiments, the bacterium can be a Bacillus bacterium. In certain embodiments, the Bacillus bacterium is B. megaterium.

In certain embodiments, an engineered microorganism described herein may comprise a genetic modification that reduces omega hydroxyl fatty acid conversion. In some embodiments, the genetic modification can reduce omega hydroxyl fatty acid dehydrogenase activity. In certain embodiments, an engineered microorganism described herein may comprise a genetic modification that reduces beta-oxidation activity. In some embodiments, the genetic modification can reduce a target activity described herein.

Engineered microorganisms that produce a fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid), as described herein, can comprise an altered monooxygenase activity, in certain embodiments. In some embodiments, the engineered microorganism described herein may comprise a genetic modification that alters the monooxygenase activity. In certain embodiments, the engineered microorganism described herein can comprise an increase number of copies of an endogenous polynucleotide encoding a polypeptide having monooxygenase activity. In some embodiments, the engineered microorganism described herein can comprise a heterologous polynucleotide encoding a polypeptide having monooxygenase activity. In certain embodiments, the heterologous polynucleotide can be from a bacterium. In some embodiments, the bacterium can be a Bacillus bacterium. In certain embodiments, the Bacillus bacterium is B. megaterium. In some embodiments, the genetic modification can reduce a polyketide synthase activity.

Engineered microorganisms that produce a fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid), as described herein, can comprise an altered thioesterase activity, in certain embodiments. In some embodiments, the engineered microorganism may comprise a genetic modification that adds or increases the thioesterase activity. In certain embodiments, the engineered microorganism may comprise a heterologous polynucleotide encoding a polypeptide having thioesterase activity.

In some embodiments, the engineered microorganism with an altered thioesterase activity may comprise an altered omega oxo fatty acid dehydrogenase activity. In certain embodiments, the engineered microorganism with an altered thioesterase activity may comprise a genetic modification that adds or increases omega oxo fatty acid dehydrogenase activity. In some embodiments, the engineered microorganism may comprise a heterologous polynucleotide encoding a polypeptide having altered omega oxo fatty acid dehydrogenase activity. In certain embodiments, the heterologous polynucleotide can be from a bacterium. In some embodiments, the bacterium can be an Acinetobacter, Nocardia, Pseudomonas or Xanthobacter bacterium.

Engineered microorganisms that produce a fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid), as described herein, can comprise an altered omega hydroxyl fatty acid dehydrogenase activity. In certain embodiments, the engineered microorganism may comprise a genetic modification that adds or increases the omega hydroxyl fatty acid dehydrogenase activity. In certain embodiments, the engineered microorganism may comprise a heterologous polynucleotide encoding a polypeptide having altered omega hydroxyl fatty acid dehydrogenase activity. In some embodiments, the heterologous polynucleotide is from a bacterium. In certain embodiments, the bacterium can be an Acinetobacter, Nocardia, Pseudomonas or Xanthobacter bacterium. In some embodiments, the engineered microorganism can be a eukaryote. In certain embodiments, the eukaryote can be a yeast. In some embodiments, the eukaryote may be a fungus. In certain embodiments, the yeast can be a Candida yeast. In some embodiments, the Candida yeast may be C. troplicalis. In certain embodiments, the fungus can be a Yarrowia fungus. In some embodiments the Yarrowia fungus may be Y. lipolytica. In certain embodiments, the fungus can be an Aspergillus fungus. In some embodiments, the Aspergillus fungus may be A. parasiticus or A. nidulans. In some

embodiments, an engineered microorganism as described above may comprise a genetic modification that reduces omega hydroxyl fatty acid conversion. In certain embodiments, the genetic modification can reduce omega hydroxyl fatty acid dehydrogenase activity. In some embodiments the genetic may reduce beta-oxidation activity. In certain embodiments, the genetic modification may reduce a target activity described herein.

Engineered microorganisms can be prepared by altering, introducing or removing nucleotide sequences in the host genome or in stably maintained epigenetic nucleic acid reagents, as noted above. The nucleic acid reagents use to alter, introduce or remove nucleotide sequences in the host genome or epigenetic nucleic acids can be prepared using the methods described herein or available to the artisan.

Nucleic acid sequences having a desired activity can be isolated from cells of a suitable organism using lysis and nucleic acid purification procedures described in a known reference manual (e.g., Maniatis, T., E. F. Fritsch and J. Sambrook (1982) *Molecular Cloning: a Laboratory Manual;* Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.) or using commercially available cell lysis and DNA purification reagents and kits. In some embodiments, nucleic acids used to engineer microorganisms can be provided for conducting methods described herein after processing of the organism containing the nucleic acid. For example, the nucleic acid of interest may be extracted, isolated, purified or amplified from a sample (e.g., from an organism of interest or culture containing a plurality of organisms of interest, like yeast or bacteria for example). The term "isolated" as used herein refers to nucleic acid removed from its original environment (e.g., the natural environment if it is naturally occurring, or a host cell if expressed exogenously), and thus is altered "by the hand of man" from its original environment. An isolated nucleic acid generally is provided with fewer non-nucleic acid components (e.g., protein, lipid) than the amount of components present in a source sample. A composition comprising isolated sample nucleic acid can be substantially isolated (e.g., about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or greater than 99% free of non-nucleic acid components). The term "purified" as used herein refers to sample nucleic acid provided that contains fewer nucleic acid species than in the sample source from which the sample nucleic acid is derived. A composition comprising sample nucleic acid may be substantially purified (e.g., about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or greater than 99% free of other nucleic acid species). The term "amplified" as used herein refers to subjecting nucleic acid of a cell, organism or sample to a process that linearly or

exponentially generates amplicon nucleic acids having the same or substantially the same nucleotide sequence as the nucleotide sequence of the nucleic acid in the sample, or portion thereof. As noted above, the nucleic acids used to prepare nucleic acid reagents as described herein can be subjected to fragmentation or cleavage.

Amplification of nucleic acids is sometimes necessary when dealing with organisms that are difficult to culture. Where amplification may be desired, any suitable amplification technique can be utilized. Non-limiting examples of methods for amplification of polynucleotides include, polymerase chain reaction (PCR); ligation amplification (or ligase chain reaction (LCR)); amplification methods based on the use of Q-beta replicase or template-dependent polymerase (see US Patent Publication Number US20050287592); helicase-dependent isothermal amplification (Vincent et al., "Helicase-dependent isothermal DNA amplification". EMBO reports 5 (8): 795-800 (2004)); strand displacement amplification (SDA); thermophilic SDA nucleic acid sequence based amplification (3SR or NASBA) and transcription-associated amplification (TAA). Non-limiting examples of PCR amplification methods include standard PCR, AFLP-PCR, Allele-specific PCR, Alu-PCR, Asymmetric PCR, Colony PCR, Hot start PCR, Inverse PCR (IPCR), In situ PCR (ISH), Intersequence-specific PCR (ISSR-PCR), Long PCR, Multiplex PCR, Nested PCR, Quantitative PCR, Reverse Transcriptase PCR (RT-PCR), Real Time PCR, Single cell PCR, Solid phase PCR, combinations thereof, and the like. Reagents and hardware for conducting PCR are commercially available.

Protocols for conducting the various type of PCR listed above are readily available to the artisan. PCR conditions can be dependent upon primer sequences, target abundance, and the desired amount of amplification, and therefore, one of skill in the art may choose from a number of PCR protocols available (see, e.g., U.S. Pat. Nos. 4,683,195 and 4,683,202; and PCR Protocols: A Guide to Methods and Applications, Innis et al, eds, 1990. PCR often is carried out as an automated process with a thermostable enzyme. In this process, the temperature of the reaction mixture is cycled through a denaturing region, a primer-annealing region, and an extension reaction region automatically. Machines specifically adapted for this purpose are commercially available. A non-limiting example of a PCR protocol that may be suitable for embodiments described herein is, treating the sample at 95°C for 5 minutes; repeating forty-five cycles of 95°C for 1 minute, 59°C for 1 minute, 10 seconds, and 72°C for 1 minute 30 seconds; and then treating the sample at 72°C for 5 minutes. Additional PCR protocols are described in the example section. Multiple cycles frequently are performed using a commercially available

thermal cycler. Suitable isothermal amplification processes known and selected by the person of ordinary skill in the art also may be applied, in certain embodiments. In some embodiments, nucleic acids encoding polypeptides with a desired activity can be isolated by amplifying the desired sequence from an organism having the desired activity using oligonucleotides or primers designed based on sequences described herein.

Amplified, isolated and/or purified nucleic acids can be cloned into the recombinant DNA vectors described in Figures herein or into suitable commercially available recombinant DNA vectors. Cloning of nucleic acid sequences of interest into recombinant DNA vectors can facilitate further manipulations of the nucleic acids for preparation of nucleic acid reagents, (e.g., alteration of nucleotide sequences by mutagenesis, homologous recombination, amplification and the like, for example). Standard cloning procedures (e.g., enzymatic digestion, ligation, and the like) are known (e.g., described in Maniatis, T., E. F. Fritsch and J. Sambrook (1982) *Molecular Cloning: a Laboratory Manual;* Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.).

In some embodiments, nucleic acid sequences prepared by isolation or amplification can be used, without any further modification, to add an activity to a microorganism and thereby create a genetically modified or engineered microorganism. In certain embodiments, nucleic acid sequences prepared by isolation or amplification can be genetically modified to alter (e.g., increase or decrease, for example) a desired activity. In some embodiments, nucleic acids, used to add an activity to an organism, sometimes are genetically modified to optimize the heterologous polynucleotide sequence encoding the desired activity (e.g., polypeptide or protein, for example). The term "optimize" as used herein can refer to alteration to increase or enhance expression by preferred codon usage. The term optimize can also refer to modifications to the amino acid sequence to increase the activity of a polypeptide or protein, such that the activity exhibits a higher catalytic activity as compared to the "natural" version of the polypeptide or protein.

Nucleic acid sequences of interest can be genetically modified using methods known in the art. Mutagenesis techniques are particularly useful for small scale (e.g., 1, 2, 5, 10 or more nucleotides) or large scale (e.g., 50, 100, 150, 200, 500, or more nucleotides) genetic modification. Mutagenesis allows the artisan to alter the genetic information of an organism in a stable manner, either naturally (e.g., isolation using selection and screening) or experimentally

by the use of chemicals, radiation or inaccurate DNA replication (e.g., PCR mutagenesis). In some embodiments, genetic modification can be performed by whole scale synthetic synthesis of nucleic acids, using a native nucleotide sequence as the reference sequence, and modifying nucleotides that can result in the desired alteration of activity. Mutagenesis methods sometimes are specific or targeted to specific regions or nucleotides (e.g., site-directed mutagenesis, PCR-based site-directed mutagenesis, and in vitro mutagenesis techniques such as transplacement and in vivo oligonucleotide site-directed mutagenesis, for example). Mutagenesis methods sometimes are non-specific or random with respect to the placement of genetic modifications (e.g., chemical mutagenesis, insertion element (e.g., insertion or transposon elements) and inaccurate PCR based methods, for example).

Site directed mutagenesis is a procedure in which a specific nucleotide or specific nucleotides in a DNA molecule are mutated or altered. Site directed mutagenesis typically is performed using a nucleic acid sequence of interest cloned into a circular plasmid vector. Site-directed mutagenesis requires that the wild type sequence be known and used a platform for the genetic alteration. Site-directed mutagenesis sometimes is referred to as oligonucleotide-directed mutagenesis because the technique can be performed using oligonucleotides which have the desired genetic modification incorporated into the complement a nucleotide sequence of interest. The wild type sequence and the altered nucleotide are allowed to hybridize and the hybridized nucleic acids are extended and replicated using a DNA polymerase. The double stranded nucleic acids are introduced into a host (e.g., E. coli, for example) and further rounds of replication are carried out in vivo. The transformed cells carrying the mutated nucleic acid sequence are then selected and/or screened for those cells carrying the correctly mutagenized sequence. Cassette mutagenesis and PCR-based site-directed mutagenesis are further modifications of the site-directed mutagenesis technique. Site-directed mutagenesis can also be performed in vivo (e.g., transplacement "pop-in pop-out", In vivo site-directed mutagenesis with synthetic oligonucleotides and the like, for example).

PCR-based mutagenesis can be performed using PCR with oligonucleotide primers that contain the desired mutation or mutations. The technique functions in a manner similar to standard site-directed mutagenesis, with the exception that a thermocycler and PCR conditions are used to replace replication and selection of the clones in a microorganism host. As PCR-based mutagenesis also uses a circular plasmid vector, the amplified fragment (e.g., linear nucleic acid molecule) containing the incorporated genetic modifications can be separated from the plasmid

containing the template sequence after a sufficient number of rounds of thermocycler amplification, using standard electrophoretic procedures. A modification of this method uses linear amplification methods and a pair of mutagenic primers that amplify the entire plasmid. The procedure takes advantage of the E. coli Dam methylase system which causes DNA replicated in vivo to be sensitive to the restriction endonucleases Dpnl. PCR synthesized DNA is not methylated and is therefore resistant to Dpnl. This approach allows the template plasmid to be digested, leaving the genetically modified, PCR synthesized plasmids to be isolated and transformed into a host bacteria for DNA repair and replication, thereby facilitating subsequent cloning and identification steps. A certain amount of randomness can be added to PCR-based sited directed mutagenesis by using partially degenerate primers.

Recombination sometimes can be used as a tool for mutagenesis. Homologous recombination allows the artisan to specifically target regions of known sequence for insertion of heterologous nucleotide sequences using the host organisms natural DNA replication and repair enzymes. Homologous recombination methods sometimes are referred to as "pop in pop out" mutagenesis, transplacement, knock out mutagenesis or knock in mutagenesis. Integration of a nucleic acid sequence into a host genome is a single cross over event, which inserts the entire nucleic acid reagent (e.g., pop in). A second cross over event excises all but a portion of the nucleic acid reagent, leaving behind a heterologous sequence, often referred to as a "footprint" (e.g., pop out). Mutagenesis by insertion (e.g., knock in) or by double recombination leaving behind a disrupting heterologous nucleic acid (e.g., knock out) both server to disrupt or "knock out" the function of the gene or nucleic acid sequence in which insertion occurs. By combining selectable markers and/or auxotrophic markers with nucleic acid reagents designed to provide the appropriate nucleic acid target sequences, the artisan can target a selectable nucleic acid reagent to a specific region, and then select for recombination events that "pop out" a portion of the inserted (e.g., "pop in") nucleic acid reagent.

Such methods take advantage of nucleic acid reagents that have been specifically designed with known target nucleic acid sequences at or near a nucleic acid or genomic region of interest. Popping out typically leaves a "foot print" of left over sequences that remain after the recombination event. The left over sequence can disrupt a gene and thereby reduce or eliminate expression of that gene. In some embodiments, the method can be used to insert sequences, upstream or downstream of genes that can result in an enhancement or reduction in expression of the gene. In certain embodiments, new genes can be introduced into the genome of a host

organism using similar recombination or "pop in" methods. An example of a yeast recombination system using the ura3 gene and 5-FOA were described briefly above and further detail is presented herein.

A method for modification is described in Alani et al., "A method for gene disruption that allows repeated use of URA3 selection in the construction of multiply disrupted yeast strains", Genetics 116(4):541-545 August 1987. The original method uses a Ura3 cassette with 1000 base pairs (bp) of the same nucleotide sequence cloned in the same orientation on either side of the URA3 cassette. Targeting sequences of about 50 bp are added to each side of the construct. The double stranded targeting sequences are complementary to sequences in the genome of the host organism. The targeting sequences allow site-specific recombination in a region of interest. The modification of the original technique replaces the two 1000 bp sequence direct repeats with two 200 bp direct repeats. The modified method also uses 50 bp targeting sequences. The modification reduces or eliminates recombination of a second knock out into the 1000 bp repeat left behind in a first mutagenesis, therefore allowing multiply knocked out yeast. Additionally, the 200 bp sequences used herein are uniquely designed, self-assembling sequences that leave behind identifiable footprints. The technique used to design the sequences incorporate design features such as low identity to the yeast genome, and low identity to each other. Therefore a library of the self-assembling sequences can be generated to allow multiple knockouts in the same organism, while reducing or eliminating the potential for integration into a previous knockout.

As noted above, the URA3 cassette makes use of the toxicity of 5-FOA in yeast carrying a functional URA3 gene. Uracil synthesis deficient yeast are transformed with the modified URA3 cassette, using standard yeast transformation protocols, and the transformed cells are plated on minimal media minus uracil. In some embodiments, PCR can be used to verify correct insertion into the region of interest in the host genome, and certain embodiments the PCR step can be omitted. Inclusion of the PCR step can reduce the number of transformants that need to be counter selected to "pop out" the URA3 cassette. The transformants (e.g., all or the ones determined to be correct by PCR, for example) can then be counter-selected on media containing 5-FOA, which will select for recombination out (e.g., popping out) of the URA3 cassette, thus rendering the yeast ura3 deficient again, and resistant to 5-FOA toxicity. Targeting sequences used to direct recombination events to specific regions are presented herein. A modification of the method described above can be used to integrate genes in to the

chromosome, where after recombination a functional gene is left in the chromosome next to the 200bp footprint.

In some embodiments, other auxotrophic or dominant selection markers can be used in place of URA3 (e.g., an auxotrophic selectable marker), with the appropriate change in selection media and selection agents. Auxotrophic selectable markers are used in strains deficient for synthesis of a required biological molecule (e.g., amino acid or nucleoside, for example). Non-limiting examples of additional auxotrophic markers include; HIS3, TRP1, LEU2, LEU2-d, and LYS2. Certain auxotrophic markers (e.g., URA3 and LYS2) allow counter selection to select for the second recombination event that pops out all but one of the direct repeats of the recombination construct. HIS3 encodes an activity involved in histidine synthesis. TRP1 encodes an activity involved in tryptophan synthesis. LEU2 encodes an activity involved in leucine synthesis. LEU2-d is a low expression version of LEU2 that selects for increased copy number (e.g., gene or plasmid copy number, for example) to allow survival on minimal media without leucine. LYS2 encodes an activity involved in lysine synthesis, and allows counter selection for recombination out of the LYS2 gene using alpha-aminoadipate (a-aminoadipate).

Dominant selectable markers are useful because they also allow industrial and/or prototrophic strains to be used for genetic manipulations. Additionally, dominant selectable markers provide the advantage that rich medium can be used for plating and culture growth, and thus growth rates are markedly increased. Non-limiting examples of dominant selectable markers include; Tn903 kan^r, Cm^r, Hyg^r, CUP1, and DHFR. Tn903 kan^r encodes an activity involved in kanamycin antibiotic resistance (e.g., typically neomycin phosphotransferase II or NPTII, for example). Cm^r encodes an activity involved in chloramphenicol antibiotic resistance (e.g., typically chloramphenicol acetyl transferase or CAT, for example). Hyg^r encodes an activity involved in hygromycin resistance by phosphorylation of hygromycin B (e.g., hygromycin phosphotransferase, or HPT). CUP1 encodes an activity involved in resistance to heavy metal (e.g., copper, for example) toxicity. DHFR encodes a dihydrofolate reductase activity which confers resistance to methotrexate and sulfanilamde compounds.

In contrast to site-directed or specific mutagenesis, random mutagenesis does not require any sequence information and can be accomplished by a number of widely different methods. Random mutagenesis often is used to generate mutant libraries that can be used to screen for the desired genotype or phenotype. Non-limiting examples of random mutagenesis include;

chemical mutagenesis, UV-induced mutagenesis, insertion element or transposon-mediated mutagenesis, DNA shuffling, error-prone PCR mutagenesis, and the like.

Chemical mutagenesis often involves chemicals like ethyl methanesulfonate (EMS), nitrous acid, mitomycin C, N-methyl-N-nitrosourea (MNU), diepoxybutane (DEB), 1, 2, 7, 8-diepoxyoctane (DEO), methyl methane sulfonate (MMS), N-methyl- N'-nitro-N-nitrosoguanidine (MNNG), 4-nitroquinoline 1-oxide (4-NQO), 2-methyloxy-6-chloro-9(3-[ethyl-2-chloroethyl]-aminopropylamino)-acridinedihydrochloride (ICR-170), 2-amino purine (2AP), and hydroxylamine (HA), provided herein as non-limiting examples. These chemicals can cause base-pair substitutions, frameshift mutations, deletions, transversion mutations, transition mutations, incorrect replication, and the like. In some embodiments, the mutagenesis can be carried out in vivo. Sometimes the mutagenic process involves the use of the host organisms DNA replication and repair mechanisms to incorporate and replicate the mutagenized base or bases.

Another type of chemical mutagenesis involves the use of base-analogs. The use of base-analogs cause incorrect base pairing which in the following round of replication is corrected to a mismatched nucleotide when compared to the starting sequence. Base analog mutagenesis introduces a small amount of non-randomness to random mutagenesis, because specific base analogs can be chose which can be incorporated at certain nucleotides in the starting sequence. Correction of the mispairing typically yields a known substitution. For example, Bromo-deoxyuridine (BrdU) can be incorporated into DNA and replaces T in the sequence. The host DNA repair and replication machinery can sometime correct the defect, but sometimes will mispair the BrdU with a G. The next round of replication then causes a G-C transversion from the original A-T in the native sequence.

Ultra violet (UV) induced mutagenesis is caused by the formation of thymidine dimers when UV light irradiates chemical bonds between two adjacent thymine residues. Excision repair mechanism of the host organism correct the lesion in the DNA, but occasionally the lesion is incorrectly repaired typically resulting in a C to T transition.

Insertion element or transposon-mediated mutagenesis makes use of naturally occurring or modified naturally occurring mobile genetic elements. Transposons often encode accessory activities in addition to the activities necessary for transposition (e.g., movement using a

transposase activity, for example). In many examples, transposon accessory activities are antibiotic resistance markers (e.g., see Tn903 kan^r described above, for example). Insertion elements typically only encode the activities necessary for movement of the nucleic acid sequence. Insertion element and transposon mediated mutagenesis often can occur randomly, however specific target sequences are known for some transposons. Mobile genetic elements like IS elements or Transposons (Tn) often have inverted repeats, direct repeats or both inverted and direct repeats flanking the region coding for the transposition genes. Recombination events catalyzed by the transposase cause the element to remove itself from the genome and move to a new location, leaving behind a portion of an inverted or direct repeat. Classic examples of transposons are the "mobile genetic elements" discovered in maize. Transposon mutagenesis kits are commercially available which are designed to leave behind a 5 codon insert (e.g., Mutation Generation System kit, Finnzymes, World Wide Web URL fmnzymes.us, for example). This allows the artisan to identify the insertion site, without fully disrupting the function of most genes.

DNA shuffling is a method which uses DNA fragments from members of a mutant library and reshuffles the fragments randomly to generate new mutant sequence combinations. The fragments are typically generated using DNasel, followed by random annealing and re-joining using self priming PCR. The DNA overhanging ends, from annealing of random fragments, provide "primer" sequences for the PCR process. Shuffling can be applied to libraries generated by any of the above mutagenesis methods.

Error prone PCR and its derivative rolling circle error prone PCR uses increased magnesium and manganese concentrations in conjunction with limiting amounts of one or two nucleotides to reduce the fidelity of the Taq polymerase. The error rate can be as high as 2% under appropriate conditions, when the resultant mutant sequence is compared to the wild type starting sequence. After amplification, the library of mutant coding sequences must be cloned into a suitable plasmid. Although point mutations are the most common types of mutation in error prone PCR, deletions and frameshift mutations are also possible. There are a number of commercial error-prone PCR kits available, including those from Stratagene and Clontech (e.g., World Wide Web URL strategene.com and World Wide Web URL clontech.com, respectively, for example). Rolling circle error-prone PCR is a variant of error-prone PCR in which wild-type sequence is first cloned into a plasmid, then the whole plasmid is amplified under error-prone conditions.

As noted above, organisms with altered activities can also be isolated using genetic selection and screening of organisms challenged on selective media or by identifying naturally occurring variants from unique environments. For example, 2-Deoxy-D-glucose is a toxic glucose analog. Growth of yeast on this substance yields mutants that are glucose-deregulated. A number of mutants have been isolated using 2-Deoxy-D-glucose including transport mutants, and mutants that ferment glucose and galactose simultaneously instead of glucose first then galactose when glucose is depleted. Similar techniques have been used to isolate mutant microorganisms that can metabolize plastics (e.g., from landfills), petrochemicals (e.g., from oil spills), and the like, either in a laboratory setting or from unique environments.

Similar methods can be used to isolate naturally occurring mutations in a desired activity when the activity exists at a relatively low or nearly undetectable level in the organism of choice, in some embodiments. The method generally consists of growing the organism to a specific density in liquid culture, concentrating the cells, and plating the cells on various concentrations of the substance to which an increase in metabolic activity is desired. The cells are incubated at a moderate growth temperature, for 5 to 10 days. To enhance the selection process, the plates can be stored for another 5 to 10 days at a low temperature. The low temperature sometimes can allow strains that have gained or increased an activity to continue growing while other strains are inhibited for growth at the low temperature. Following the initial selection and secondary growth at low temperature, the plates can be replica plated on higher or lower concentrations of the selection substance to further select for the desired activity.

A native, heterologous or mutagenized polynucleotide can be introduced into a nucleic acid reagent for introduction into a host organism, thereby generating an engineered microorganism. Standard recombinant DNA techniques (restriction enzyme digests, ligation, and the like) can be used by the artisan to combine the mutagenized nucleic acid of interest into a suitable nucleic acid reagent capable of (i) being stably maintained by selection in the host organism, or (ii) being integrating into the genome of the host organism. As noted above, sometimes nucleic acid reagents comprise two replication origins to allow the same nucleic acid reagent to be manipulated in bacterial before final introduction of the final product into the host organism (e.g., yeast or fungus for example). Standard molecular biology and recombinant DNA methods are known (e.g., described in Maniatis, T., E. F. Fritsch and J. Sambrook (1982)

Molecular Cloning: a Laboratory Manual; Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.).

Nucleic acid reagents can be introduced into microorganisms using various techniques. Non-limiting examples of methods used to introduce heterologous nucleic acids into various organisms include; transformation, transfection, transduction, electroporation, ultrasound-mediated transformation, particle bombardment and the like. In some instances the addition of carrier molecules (e.g., bis-benzimdazolyl compounds, for example, see US Patent 5595899) can increase the uptake of DNA in cells typically though to be difficult to transform by conventional methods. Conventional methods of transformation are known (e.g., described in Maniatis, T., E. F. Fritsch and J. Sambrook (1982) *Molecular Cloning: a Laboratory Manual;* Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.).

Modified Activities

Certain activities in a genetically modified organism can be modified by techniques known in the art. An acyl-CoA oxidase activity or acyl-CoA dehydrogenase activity, or acyl-CoA oxidase activity and acyl-CoA dehydrogenase activity, can be modified in an organism in certain embodiments. In some embodiments, a modified endogenous acyl-CoA oxidase polypeptide, modified endogenous acyl-CoA dehydrogenase polypeptide, modified heterologous acyl-CoA oxidase polypeptide, and/or modified heterologous acyl-CoA dehydrogenase polypeptide may be introduced into an organism. A modified polypeptide can be expressed by a host organism that includes a modified polynucleotide encoding the modified polypeptide. Modified polypeptides often have an activity different than the activity of an unmodified counterpart. A modified activity sometimes is a different catalytic activity or a different substrate specificity, or a different catalytic activity and a different substrate specificity. A different activity sometimes is an activity that is higher or lower than the activity of an unmodified counterpart polypeptide. In some embodiments, the catalytic activity of a modified polypeptide is higher or lower than the catalytic activity of the unmodified counterpart for a particular substrate. In certain embodiments, the substrate specificity of a modified polypeptide is higher or lower than the substrate specificity of the unmodified counterpart for a particular substrate. A modified polypeptide often is active and an activity of a modified polypeptide often can be detected (e.g., substrate turnover can be detected). A desired activity for a particular polypeptide sometimes is referred to as a "target activity."

In some embodiments a genetic modification in a genetically modified organism alters a substrate specificity of an acyl-CoA oxidase polypeptide produced in the organism. Sometimes the substrate specificity is reduced for a substrate having a particular chain length. In some embodiments, a modified acyl-CoA oxidase substrate specificity is reduced for a C8, CIO, C12, C14, C16, C18, C20 substrate or combination thereof. In certain embodiments, a modified acyl-CoA oxidase substrate specificity is reduced for a CIO, C12, or C18 substrate.

In some embodiments a genetic modification in a genetically modified organism alters a substrate specificity of an acyl-CoA dehydrogenase polypeptide produced in the organism. Sometimes a co-factor specificity is modified, and in some embodiments the modified polypeptide prefers to utilize oxygen as a co-factor.

One or more particular modifications can be selected to generate a modified polypeptide having a target activity. Modifications often are amino acid modifications (e.g., deletion, insertion of one or more amino acids). Amino acid modifications sometimes are amino acid substitutions. Amino acid substitutions sometimes are conservative, non-limiting examples of which include substitution of an amino acid containing an acidic moiety to another amino acid containing an acidic moiety (e.g., D, E), substitution of an amino acid containing a basic moiety to another amino acid containing a basic moiety (e.g., H, K, R), substitution of an amino acid containing an aliphatic chain moiety to another amino acid containing an aliphatic chain moiety (V, L, I, A), substitution of an amino acid containing a cyclic moiety to another amino acid containing a cyclic moiety (e.g., W, F, Y), and substitution of an amino acid containing a polar moiety to another amino acid containing a polar moiety (e.g., S, T). Amino acid substitutions sometimes are non-conservative, non-limiting examples of which include substitution of an amino acid containing an acidic moiety to an amino acid containing a basic moiety, substitution of an amino acid containing a basic moiety to an amino acid containing an acidic moiety, substitution of an amino acid containing relatively small moiety (e.g., G, A) to another amino acid containing a relatively large moiety (e.g., Y, W, F, I, L), and substitution of an amino acid containing a relatively large moiety to another amino acid containing an relatively small moiety.

Particular modifications can be selected using any suitable method known in the art. In certain embodiments, a reference structure is known for a related polypeptide with a known activity,

and modifications to a target polypeptide can be guided by alignment of the target polypeptide structure to the reference structure. A reference structure sometimes is a primary structure (e.g., polynucleotide or polypeptide sequence) and the primary structure of a target can be aligned to the reference structure using an alignment method known in the art. Particular amino acids in the target that align with (e.g., are identical to or homologous to) or do not align with (e.g., are not identical to or not homologous to) particular amino acids in the reference can be selected for modification. Selections can be made by inspection of an alignment or by software known in the art that identifies, scores and/or ranks amino acids for modification based on an alignment.

A reference structure sometimes is a secondary structure, tertiary structure or quaternary structure, each of which are three dimensional structures pertaining to a polypeptide. A primary structure of a target polypeptide can be modeled to a secondary, tertiary or quaternary reference structure using three-dimensional modeling software known in the art. A secondary, tertiary or quaternary structure of a target polypeptide can be compared to a secondary, tertiary or quaternary reference structure using three-dimensional comparative software known in the art. Particular structures (e.g., a particular individual amino acid; a particular group of contiguous or non-contiguous amino acids) in the target that align with or map to, or do not align with or map to, particular structures in the reference can be selected for modification. Also, particular structures in the target that are in proximity to a substrate or co-factor can be selected for modification. Selections can be made by inspection of an alignment or map or by software known in the art that identifies, scores and/or ranks amino acids and/or structures for modification based on an alignment and map.

After particular amino acids and/or structures are selected for modification in a first polypeptide, amino acids and structures in a second polypeptide that align with the selected amino acids and structures in the first polypeptide may be identified. In a non-limiting example, particular amino acid substitutions and structural modifications (e.g., loop amino acid deletion/insertion) for Candida spp. POX4 and POX5 polypeptides are disclosed herein. A primary structure of another acyl-CoA oxidase polypeptide can be aligned with the amino acid sequence or modeled structure of a POX4 or POX5 polypeptide and some or all amino acids of the other polypeptide that align with those selected for modification in the POX4 or POX5 polypeptide also can be selected for modification. Certain criteria for selecting acyl-CoA dehydrogenase modifications also are described herein.

One or more activities of a modified polypeptide can be characterized using any suitable assay known in the art. A modified polypeptide can be expressed in an organism other than a target organism in which a target product will be produced, for assaying activity. For example, a modified polypeptide can be expressed in a bacterium (e.g., E. coli), assayed and then introduced into a yeast (e.g., Candida spp. yeast) for production of a target diacid.

Feedstocks, Media, Supplements & Additives

Engineered microorganisms often are cultured under conditions that optimize yield of a fatty dicarboxylic acid (e.g., an eight to eighteen-carbon fatty dicarboxylic acid). Non-limiting examples of fatty dicarboxylic acids include suberic acid (i.e., octanedioic acid, 1,8-octanedioic acid, octanedioic acid, octane-l,8-dioic acid, 1,6-hexanedicarboxylic acid, capryllic diacids), sebacic acid (i.e., 1,10-decanedioic acid, decanedioic acid, decane-l,10-dioic acid, 1,8-octanedicarboxylic acid, capric diacid), dodecanedioic acid (i.e., DDDA, 1,12-dodecanedioic acid, dodecanedioic acid, dodecane-l,12-dioic acid, 1,10-decanedicarboxylic acid, decamethylenedicaboxylic acid, 1,10-dicarboxydecane, lauric diacid), tetradecanedioic acid (i.e., TDDA, 1,14-tetradecanedioic acid, tetradecanedioic acid, tetradecane-l,14-dioic acid, 1,12-dodecanedicarboxylic acid, myristic diacid), thapsic acid (i.e., hexadecanedioic acid, 1,16-hexadecanedioic acid, hexadecanedioic acid, hexadecane-l,16-dioic acid, 1,14-tetradecanedicarboxylic acid, palmitic diacid), cis-9-hexadecenedioic acid (i.e., palmitoleic diacids), octanedioic acid (i.e., 1,18-octadecanedioic acid, octadecanedioic acid, octadecane-1,18-dioic acid, 1,16-hexadecanedicarboxylic acid, stearic diacid), cis-9-octadecenedioic acid (i.e., oleic diacids), cis-9,12-octadecenedioic acid (i.e., linoleic diacids), cis-9, 12,15-octadecenedioic acid (i.e., linolenic diacids), arachidic diacid (i.e., eicosanoic diacid, icosanoic diacid), 11-eicosenoic diacid (i.e., cis-1 1-eicosenedioic acid), 13-eicosenoic diacids (i.e., cis-13-eicosenedioic acid), arachidonic diacid (i.e., cis-5,8,1 1,14-eicosatetraenedioic acid). Culture conditions often optimize activity of one or more of the following activities: omega oxo fatty acid dehydrogenase activity, omega hydroxyl fatty acid dehydrogenase activity, acetyl CoA carboxylase activity, monooxygenase activity, monooxygenase reductase activity, fatty alcohol oxidase, acyl-CoA ligase, acyl-CoA oxidase, enoyl-CoA hydratase, 3-hydroxyacyl-CoA dehydrogenase, and/or acyltransferase (e.g., acetyl-CoA C-acyltransferase) activities. In general, non-limiting examples of conditions that may be optimized include the type and amount of carbon source, the type and amount of nitrogen source, the carbon-to-nitrogen ratio,

the oxygen level, growth temperature, pH, length of the biomass production phase, length of target product accumulation phase, and time of cell harvest.

Culture media generally contain a suitable carbon source. Carbon sources useful for culturing microorganisms and/or fermentation processes sometimes are referred to as feedstocks. The term "feedstock" as used herein refers to a composition containing a carbon source that is provided to an organism, which is used by the organism to produce energy and metabolic products useful for growth. A feedstock may be a natural substance, a "man-made substance," a purified or isolated substance, a mixture of purified substances, a mixture of unpurified substances or combinations thereof. A feedstock often is prepared by and/or provided to an organism by a person, and a feedstock often is formulated prior to administration to the organism. A carbon source may comprise, but is not limited to including, one or more of the following substances: alkanes, alkenes, mono-carboxylic acids, di-carboxylic acids, monosaccharides (e.g., also referred to as "saccharides," which include 6-carbon sugars (e.g., glucose, fructose), 5-carbon sugars (e.g., xylose and other pentoses) and the like), disaccharides (e.g., lactose, sucrose), oligosaccharides (e.g., glycans, homopolymers of a monosaccharide), polysaccharides (e.g., starch, cellulose, heteropolymers of monosaccharides or mixtures thereof), sugar alcohols (e.g., glycerol), and renewable feedstocks (e.g., cheese whey permeate, cornsteep liquor, sugar beet molasses, barley malt).

Carbon sources also can be selected from one or more of the following non-limiting examples: paraffin (e.g., saturated paraffin, unsaturated paraffin, substituted paraffin, linear paraffin, branched paraffin, or combinations thereof); alkanes (e.g., dodecane), alkenes or alkynes, each of which may be linear, branched, saturated, unsaturated, substituted or combinations thereof (described in greater detail below); linear or branched alcohols (e.g., dodecanol); fatty acids (e.g., about 1 carbon to about 60 carbons, including free fatty acids, soap stock, for example); esters of fatty acids; monoglycerides; diglycerides; triglycerides, phospholipids. Non-limiting commercial sources of products for preparing feedstocks include plants, plant oils or plant products (e.g., vegetable oils (e.g., almond oil, canola oil, cocoa butter, coconut oil, corn oil, cottonseed oil, flaxseed oil, grape seed oil, illipe, olive oil, palm oil, palm olein, palm kernel oil, safflower oil, peanut oil, soybean oil, sesame oil, shea nut oil, sunflower oil walnut oil, the like and combinations thereof) and animal fats (e.g., beef tallow, butterfat, lard, cod liver oil). A carbon source may include a petroleum product and/or a petroleum distillate (e.g., diesel, fuel oils, gasoline, kerosene, paraffin wax, paraffin oil, petrochemicals). In some embodiments, a

feedstock comprises petroleum distillate. A carbon source can be a fatty acid distillate (e.g., a palm oil distillate or corn oil distillate). Fatty acid distillates can be by-products from the refining of crude plant oils. In some embodiments, a feedstock comprises a fatty acid distillate.

In some embodiments, a feedstock comprises a soapstock (i.e. soap stock). A widely practiced method for purifying crude vegetable oils for edible use is the alkali or caustic refining method. This process employs a dilute aqueous solution of caustic soda to react with the free fatty acids present which results in the formation of soaps. The soaps together with hydrated phosphatides, gums and prooxidant metals are typically separated from the refined oil as the heavy phase discharge from the refining centrifuge and are typically known as soapstock.

A carbon source also may include a metabolic product that can be used directly as a metabolic substrate in an engineered pathway described herein, or indirectly via conversion to a different molecule using engineered or native biosynthetic pathways in an engineered microorganism. In certain embodiments, metabolic pathways can be preferentially biased towards production of a desired product by increasing the levels of one or more activities in one or more metabolic pathways having and/or generating at least one common metabolic and/or synthetic substrate. In some embodiments, a metabolic byproduct (e.g., fatty acid) of an engineered activity (e.g., omega oxidation activity) can be used in one or more metabolic pathways selected from gluconeogenesis, pentose phosphate pathway, glycolysis, fatty acid synthesis, beta oxidation, and omega oxidation, to generate a carbon source that can be converted to a fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid).

The term "paraffin" as used herein refers to the common name for alkane hydrocarbons, independent of the source (e.g., plant derived, petroleum derived, chemically synthesized, fermented by a microorganism), or carbon chain length. A carbon source sometimes comprises a paraffin, and in some embodiments, a paraffin is predominant in a carbon source (e.g., about 75%, 80%, 85%, 90% or 95% paraffin). A paraffin sometimes is saturated (e.g., fully saturated), sometimes includes one or more unsaturations (e.g., about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 unsaturations) and sometimes is substituted with one or more non-hydrogen substituents. Non-limiting examples of non-hydrogen substituents include halo, acetyl, =0, =N-CN, =N-OR, =NR, OR, NR_2, SR, $S0_2R$, $S0_2NR_2$, $NRS0_2R$, $NRCONR_2$, NRCOOR, NRCOR, CN, COOR, $CONR_2$, OOCR, COR, and $N0_2$, where each R is independently H, C1-C8 alkyl, C2-C8

heteroalkyl, C1-C8 acyl, C2-C8 heteroacyl, C2-C8 alkenyl, C2-C8 heteroalkenyl, C2-C8 alkynyl, C2-C8 heteroalkynyl, C6-C10 aryl, or C5-C10 heteroaryl, and each R is optionally substituted with halo, =0, =N-CN, =N-OR', =NR', OR', NR'_2, SR', SO_2R', $SO_2NR'_2$, NR'SO_2R', NR'CONR'_2, NR'COOR', NR'COR', CN, COOR', $CONR'_2$, OOCR', COR', and NO_2, where each R' is independently H, C1-C8 alkyl, C2-C8 heteroalkyl, C1-C8 acyl, C2-C8 heteroacyl, C6-C10 aryl or C5-C10 heteroaryl.

In some embodiments a feedstock is selected according to the genotype and/or phenotype of the engineered microorganism to be cultured. For example, a feedstock rich in 12-carbon fatty acids, 12-carbon dicarboxylic acids or 12-carbon paraffins, or a mixture of 10, 12 and 14-carbon compounds can be useful for culturing yeast strains harboring an alteration that partially blocks beta oxidation by disrupting POX4 activity, as described herein. Non-limiting examples of carbon sources having 10 to 14 carbons include fats (e.g., coconut oil, palm kernel oil), paraffins (e.g., alkanes, alkenes, or alkynes) having 10 to 14 carbons, (e.g., dodecane (also referred to as adakanel2, bihexyl, dihexyl and duodecane); tetradecane), alkene and alkyne derivatives), fatty acids (dodecanoic acid, tetradecanoic acid), fatty alcohols (dodecanol, tetradecanol), the like, non-toxic substituted derivatives or combinations thereof.

A carbon source sometimes comprises an alkyl, alkenyl or alkynyl compound or molecule (e.g., a compound that includes an alkyl, alkenyl or alkynyl moiety (e.g., alkane, alkene, alkyne)). In certain embodiments, an alkyl, alkenyl or alkynyl molecule, or combination thereof, is predominant in a carbon source (e.g., about 75%, 80%, 85%>, 90%> or 95% of such molecules). As used herein, the terms "alkyl," "alkenyl" and "alkynyl" include straight-chain (referred to herein as "linear"), branched-chain (referred to herein as "non-linear"), cyclic monovalent hydrocarbyl radicals, and combinations of these, which contain only C and H atoms when they are unsubstituted. Non-limiting examples of alkyl moieties include methyl, ethyl, isobutyl, cyclohexyl, cyclopentylethyl, 2-propenyl, 3-butynyl, and the like. An alkyl that contains only C and H atoms and is unsubstituted sometimes is referred to as "saturated." An alkenyl or alkynyl generally is "unsaturated" as it contains one or more double bonds or triple bonds, respectively. An alkenyl can include any number of double bonds, such as 1, 2, 3, 4 or 5 double bonds, for example. An alkynyl can include any number of triple bonds, such as 1, 2, 3, 4 or 5 triple bonds, for example. Alkyl, alkenyl and alkynyl molecules sometimes contain between about 2 to about 60 carbon atoms (C). For example, an alkyl, alkenyl and alkynyl molecule can include about 1 carbon atom, about 2 carbon atoms, about 3 carbon atoms, about 4 carbon atoms, about

5 carbon atoms, about 6 carbon atoms, about 7 carbon atoms, about 8 carbon atoms, about 9 carbon atoms, about 10 carbon atoms, about 12 carbon atoms, about 14 carbon atoms, about 16 carbon atoms, about 18 carbon atoms, about 20 carbon atoms, about 22 carbon atoms, about 24 carbon atoms, about 26 carbon atoms, about 28 carbon atoms, about 30 carbon atoms, about 32 carbon atoms, about 34 carbon atoms, about 36 carbon atoms, about 38 carbon atoms, about 40 carbon atoms, about 42 carbon atoms, about 44 carbon atoms, about 46 carbon atoms, about 48 carbon atoms, about 50 carbon atoms, about 52 carbon atoms, about 54 carbon atoms, about 56 carbon atoms, about 58 carbon atoms or about 60 carbon atoms. In some embodiments, paraffins can have a mean number of carbon atoms of between about 8 to about 18 carbon atoms (e.g., about 8 carbon atoms, about 9 carbon atoms, about 10 carbon atoms, about 11 carbon atoms, about 12 carbon atoms, about 13 carbon atoms, about 14 carbon atoms, about 15 carbon atoms, about 16 carbon atoms, about 17 carbon atoms and about 18 carbon atoms). A single group can include more than one type of multiple bond, or more than one multiple bond. Such groups are included within the definition of the term "alkenyl" when they contain at least one carbon-carbon double bond, and are included within the term "alkynyl" when they contain at least one carbon-carbon triple bond. Alkyl, alkenyl and alkynyl molecules include molecules that comprise an alkyl, alkenyl and/or alkynyl moiety, and include molecules that consist of an alkyl, alkenyl or alkynyl moiety (i.e., alkane, alkene and alkyne molecules).

Alkyl, alkenyl and alkynyl substituents sometimes contain 1-20C (alkyl) or 2-20C (alkenyl or alkynyl). They can contain about 8-20C or about 10-20C in some embodiments. A single group can include more than one type of multiple bond, or more than one multiple bond. Such groups are included within the definition of the term "alkenyl" when they contain at least one carbon-carbon double bond, and are included within the term "alkynyl" when they contain at least one carbon-carbon triple bond.

Alkyl, alkenyl and alkynyl groups or compounds sometimes are substituted to the extent that such substitution can be synthesized and can exist. Typical substituents include, but are not limited to, halo, acetyl, =0, =N-CN, =N-OR, =NR, OR, NR_2, SR, $S0_2R$, $S0_2NR_2$, $NRS0_2R$, $NRCONR_2$, NRCOOR, NRCOR, CN, COOR, $CONR_2$, OOCR, COR, and$N0_2$, where each R is independently H, C1-C8 alkyl, C2-C8 heteroalkyl, C1-C8 acyl, C2-C8 heteroacyl, C2-C8 alkenyl, C2-C8 heteroalkenyl, C2-C8 alkynyl, C2-C8 heteroalkynyl, C6-C1 1 aryl, or C5-C1 1 heteroaryl, and each R is optionally substituted with halo, =0, =N-CN, =N-OR', =NR', OR', NR'_2, SR', $S0_2R'$, $S0_2NR'_2$, $NR'S0_2R'$, $NR'CONR'_2$, NR'COOR', NR'COR', CN, COOR',

CONR'2, OOCR', COR', andN0 $_2$, where each R' is independently H, C1-C8 alkyl, C2-C8 heteroalkyl, C1-C8 acyl, C2-C8 heteroacyl, C6-C10 aryl or C5-C10 heteroaryl. Alkyl, alkenyl and alkynyl groups can also be substituted by C1-C8 acyl, C2-C8 heteroacyl, C6-C10 aryl or C5-C10 heteroaryl, each of which can be substituted by the substituents that are appropriate for the particular group.

"Acetylene" or "acetyl" substituents are 2-IOC alkynyl groups that are optionally substituted, and are of the formula -C≡C-Ri, where Ri is H or C1-C8 alkyl, C2-C8 heteroalkyl, C2-C8 alkenyl, C2-C8 heteroalkenyl, C2-C8 alkynyl, C2-C8 heteroalkynyl, C1-C8 acyl, C2-C8 heteroacyl, C6-C10 aryl, C5-C10 heteroaryl, C7-C12 arylalkyl, or C6-C12 heteroarylalkyl, and each Ri group is optionally substituted with one or more substituents selected from halo, =0, =N-CN, =N-OR', =NR', OR', NR'2, SR', S0 $_2$R', S0 $_2$NR' $_2$, NR'S0 $_2$R', NR'CONR' $_2$, NR'COOR', NR'COR', CN, COOR', CONR' $_2$, OOCR', COR', and N0 $_2$, where each R' is independently H, C1-C6 alkyl, C2-C6 heteroalkyl, C1-C6 acyl, C2-C6 heteroacyl, C6-C10 aryl, C5-C10 heteroaryl, C7-12 arylalkyl, or C6-12 heteroarylalkyl, each of which is optionally substituted with one or more groups selected from halo, C1-C4 alkyl, C1-C4 heteroalkyl, Cl-C6 acyl, C1-C6 heteroacyl, hydroxy, amino, and =0; and where two R' can be linked to form a 3-7 membered ring optionally containing up to three heteroatoms selected from N, O and S. In some embodiments, Ri of -C≡C-Ri is H or Me.

A carbon source sometimes comprises a heteroalkyl, heteroalkenyl and/or heteroalkynyl molecule or compound (e.g., comprises heteroalkyl, heteroalkenyl and/or heteroalkynyl moiety (e.g., heteroalkane, heteroalkene or heteroalkyne)). "Heteroalkyl", "heteroalkenyl", and "heteroalkynyl" and the like are defined similarly to the corresponding hydrocarbyl (alkyl, alkenyl and alkynyl) groups, but the 'hetero' terms refer to groups that contain one to three O, S or N heteroatoms or combinations thereof within the backbone; thus at least one carbon atom of a corresponding alkyl, alkenyl, or alkynyl group is replaced by one of the specified heteroatoms to form a heteroalkyl, heteroalkenyl, or heteroalkynyl group. The typical and sizes for heteroforms of alkyl, alkenyl and alkynyl groups are generally the same as for the corresponding hydrocarbyl groups, and the substituents that may be present on the heteroforms are the same as those described above for the hydrocarbyl groups. For reasons of chemical stability, it is also understood that, unless otherwise specified, such groups do not include more than two contiguous heteroatoms except where an oxo group is present on N or S as in a nitro or sulfonyl group.

The term "alkyl" as used herein includes cycloalkyl and cycloalkylalkyl groups and compounds, the term "cycloalkyl" may be used herein to describe a carbocyclic non-aromatic compound or group that is connected via a ring carbon atom, and "cycloalkylalkyl" may be used to describe a carbocyclic non-aromatic compound or group that is connected to a molecule through an alkyl linker. Similarly, "heterocyclyl" may be used to describe a non-aromatic cyclic group that contains at least one heteroatom as a ring member and that is connected to the molecule via a ring atom, which may be C or N; and "heterocyclylalkyl" may be used to describe such a group that is connected to another molecule through a linker. The sizes and substituents that are suitable for the cycloalkyl, cycloalkylalkyl, heterocyclyl, and heterocyclylalkyl groups are the same as those described above for alkyl groups. As used herein, these terms also include rings that contain a double bond or two, as long as the ring is not aromatic.

A carbon source sometimes comprises an acyl compound or moiety (e.g., compound comprising an acyl moiety). As used herein, "acyl" encompasses groups comprising an alkyl, alkenyl, alkynyl, aryl or arylalkyl radical attached at one of the two available valence positions of a carbonyl carbon atom, and heteroacyl refers to the corresponding groups where at least one carbon other than the carbonyl carbon has been replaced by a heteroatom chosen from N, O and S. Thus heteroacyl includes, for example, -C(=0)OR and -C(=0)NR_2 as well as -C(=0)-heteroaryl.

Acyl and heteroacyl groups are bonded to any group or molecule to which they are attached through the open valence of the carbonyl carbon atom. Typically, they are C1-C8 acyl groups, which include formyl, acetyl, pivaloyl, and benzoyl, and C2-C8 heteroacyl groups, which include methoxyacetyl, ethoxycarbonyl, and 4-pyridinoyl. The hydrocarbyl groups, aryl groups, and heteroforms of such groups that comprise an acyl or heteroacyl group can be substituted with the substituents described herein as generally suitable substituents for each of the corresponding component of the acyl or heteroacyl group.

A carbon source sometimes comprises one or more aromatic moieties and/or heteroaromatic moieties. "Aromatic" moiety or "aryl" moiety refers to a monocyclic or fused bicyclic moiety having the well-known characteristics of aromaticity; examples include phenyl and naphthyl. Similarly, "heteroaromatic" and "heteroaryl" refer to such monocyclic or fused bicyclic ring

systems which contain as ring members one or more heteroatoms selected from O, S and N. The inclusion of a heteroatom permits aromaticity in 5 membered rings as well as 6 membered rings. Typical heteroaromatic systems include monocyclic C5-C6 aromatic groups such as pyridyl, pyrimidyl, pyrazinyl, thienyl, furanyl, pyrrolyl, pyrazolyl, thiazolyl, oxazolyl, and imidazolyl and the fused bicyclic moieties formed by fusing one of these monocyclic groups with a phenyl ring or with any of the heteroaromatic monocyclic groups to form a C8-C10 bicyclic group such as indolyl, benzimidazolyl, indazolyl, benzotriazolyl, isoquinolyl, quinolyl, benzothiazolyl, benzofuranyl, pyrazolopyridyl, quinazolinyl, quinoxalinyl, cinnolinyl, and the like. Any monocyclic or fused ring bicyclic system which has the characteristics of aromaticity in terms of electron distribution throughout the ring system is included in this definition. It also includes bicyclic groups where at least the ring which is directly attached to the remainder of the molecule has the characteristics of aromaticity. Typically, the ring systems contain 5-12 ring member atoms. The monocyclic heteroaryls sometimes contain 5-6 ring members, and the bicyclic heteroaryls sometimes contain 8-10 ring members.

Aryl and heteroaryl moieties may be substituted with a variety of substituents including C1-C8 alkyl, C2-C8 alkenyl, C2-C8 alkynyl, C5-C12 aryl, C1-C8 acyl, and heteroforms of these, each of which can itself be further substituted; other substituents for aryl and heteroaryl moieties include halo, OR, NR_2, SR, $S0_2R$, $S0_2NR_2$, $NRS0_2R$, $NRCONR_2$, NRCOOR, NRCOR, CN, COOR, $CONR_2$, OOCR, COR, and$N0_2$, where each R is independently H, C1-C8 alkyl, C2-C8 heteroalkyl, C2-C8 alkenyl, C2-C8 heteroalkenyl, C2-C8 alkynyl, C2-C8 heteroalkynyl, C6-C10 aryl, C5-C10 heteroaryl, C7-C12 arylalkyl, or C6-C12 heteroarylalkyl, and each R is optionally substituted as described above for alkyl groups. The substituent groups on an aryl or heteroaryl group may be further substituted with the groups described herein as suitable for each type of such substituents or for each component of the substituent. Thus, for example, an arylalkyl substituent may be substituted on the aryl portion with substituents typical for aryl groups, and it may be further substituted on the alkyl portion with substituents as typical or suitable for alkyl groups.

Similarly, "arylalkyl" and "heteroarylalkyl" refer to aromatic and heteroaromatic ring systems, which are stand-alone molecules (e.g., benzene or substituted benzene, pyridine or substituted pyridine), or which are bonded to an attachment point through a linking group such as an alkylene, including substituted or unsubstituted, saturated or unsaturated, cyclic or acyclic linkers. A linker often is C1-C8 alkyl or a hetero form thereof. These linkers also may include

a carbonyl group, thus making them able to provide substituents as an acyl or heteroacyl moiety. An aryl or heteroaryl ring in an arylalkyl or heteroarylalkyl group may be substituted with the same substituents described above for aryl groups. An arylalkyl group sometimes includes a phenyl ring optionally substituted with the groups defined above for aryl groups and a C1-C4 alkylene that is unsubstituted or is substituted with one or two C1-C4 alkyl groups or heteroalkyl groups, where the alkyl or heteroalkyl groups can optionally cyclize to form a ring such as cyclopropane, dioxolane, or oxacyclopentane. Similarly, a heteroarylalkyl group often includes a C5-C6 monocyclic heteroaryl group optionally substituted with one or more of the groups described above as substituents typical on aryl groups and a C1-C4 alkylene that is unsubstituted. A heteroarylalkyl group sometimes is substituted with one or two C1-C4 alkyl groups or heteroalkyl groups, or includes an optionally substituted phenyl ring or C5-C6 monocyclic heteroaryl and a C1-C4 heteroalkylene that is unsubstituted or is substituted with one or two C1-C4 alkyl or heteroalkyl groups, where the alkyl or heteroalkyl groups can optionally cyclize to form a ring such as cyclopropane, dioxolane, or oxacyclopentane.

Where an arylalkyl or heteroarylalkyl group is described as optionally substituted, the substituents may be on the alkyl or heteroalkyl portion or on the aryl or heteroaryl portion of the group. The substituents optionally present on the alkyl or heteroalkyl portion sometimes are the same as those described above for alkyl groups, and the substituents optionally present on the aryl or heteroaryl portion often are the same as those described above for aryl groups generally.

"Arylalkyl" groups as used herein are hydrocarbyl groups if they are unsubstituted, and are described by the total number of carbon atoms in the ring and alkylene or similar linker. Thus a benzyl group is a C7-arylalkyl group, and phenylethyl is a C8-arylalkyl.

"Heteroarylalkyl" as described above refers to a moiety comprising an aryl group that is attached through a linking group, and differs from "arylalkyl" in that at least one ring atom of the aryl moiety or one atom in the linking group is a heteroatom selected from N, O and S. The heteroarylalkyl groups are described herein according to the total number of atoms in the ring and linker combined, and they include aryl groups linked through a heteroalkyl linker; heteroaryl groups linked through a hydrocarbyl linker such as an alkylene; and heteroaryl groups linked through a heteroalkyl linker. Thus, for example, C7-heteroarylalkyl includes pyridylmethyl, phenoxy, and N-pyrrolylmethoxy.

"Alkylene" as used herein refers to a divalent hydrocarbyl group. Because an alkylene is divalent, it can link two other groups together. An alkylene often is referred to as $-(CH_2)_n-$ where n can be 1-20, 1-10, 1-8, or 1-4, though where specified, an alkylene can also be substituted by other groups, and can be of other lengths, and the open valences need not be at opposite ends of a chain. Thus -CH(Me)- and $-C(Me)_2-$ may also be referred to as alkylenes, as can a cyclic group such as cyclopropan-l,l-diyl. Where an alkylene group is substituted, the substituents include those typically present on alkyl groups as described herein.

In some embodiments, a feedstock includes a mixture of carbon sources, where each carbon source in the feedstock is selected based on the genotype of the engineered microorganism. In certain embodiments, a mixed carbon source feedstock includes one or more carbon sources selected from sugars, cellulose, alkanes, fatty acids, triacylglycerides, paraffins, the like and combinations thereof.

Nitrogen may be supplied from an inorganic (e.g., (NH.sub.4).sub.2SO.sub.4) or organic source (e.g., urea or glutamate). In addition to appropriate carbon and nitrogen sources, culture media also can contain suitable minerals, salts, cofactors, buffers, vitamins, metal ions (e.g., Mn.sup.+2, Co.sup.+2, Zn.sup.+2, Mg.sup.+2) and other components suitable for culture of microorganisms .

Engineered microorganisms sometimes are cultured in complex media (e.g., yeast extract-peptone-dextrose broth (YPD)). In some embodiments, engineered microorganisms are cultured in a defined minimal media that lacks a component necessary for growth and thereby forces selection of a desired expression cassette (e.g., Yeast Nitrogen Base (DIFCO Laboratories, Detroit, Mich.)).

Culture media in some embodiments are common commercially prepared media, such as Yeast Nitrogen Base (DIFCO Laboratories, Detroit, Mich.). Other defined or synthetic growth media may also be used and the appropriate medium for growth of the particular microorganism are known. A variety of host organisms can be selected for the production of engineered microorganisms. Non-limiting examples include yeast (e.g., Candida tropicalis (e.g., ATCC20336, ATCC20913, ATCC20962), Yarrowia lipolytica (e.g., ATCC20228)) and filamentous fungi (e.g., Aspergillus nidulans (e.g., ATCC38164) and Aspergillus parasiticus

(e.g., ATCC 24690)). In specific embodiments, yeast are cultured in YPD media (10 g/L Bacto Yeast Extract, 20 g/L Bacto Peptone, and 20 g/L Dextrose). Filamentous fungi, in particular embodiments, are grown in CM (Complete Medium) containing 10 g/L Dextrose, 2 g/L Bacto Peptone, lg/L Bacto Yeast Extract, 1 g/L Casamino acids, 50 mL /L 20X Nitrate Salts (120 g/L $NaN0_3$, 10.4 g/L KC1, 10.4 g/L $MgS0_4 \cdot 7H_20$), 1 mL/L 1000X Trace Elements (22 g/L $ZnS0_4 \cdot 7H_20$, 11 g/L H_3BO_3, 5 g/L $MnCl_2 \cdot 7H_20$, 5 g/L $FeS0_4 \cdot 7H_20$, 1.7 g/L $CoCl_2 \cdot 6H_20$, 1.6 g/L $CuS0_4 \cdot 5H_20$, 1.5 g/L $Na_2Mo0_4 \cdot 2H_20$, and 50 g/L Na_4EDTA), and 1 mL/L Vitamin Solution (100 mg each of Biotin, pyridoxine, thiamine, riboflavin, p-aminobenzoic acid, and nicotinic acid in 100 mL water).

Growth Conditions & Fermentation

A suitable pH range for the fermentation often is between about pH 4.0 to about pH 8.0, where a pH in the range of about pH 5.5 to about pH 7.0 sometimes is utilized for initial culture conditions. Depending on the host organism, culturing may be conducted under aerobic or anaerobic conditions, where microaerobic conditions sometimes are maintained. A two-stage process may be utilized, where one stage promotes microorganism proliferation and another state promotes production of target molecule. In a two-stage process, the first stage may be conducted under aerobic conditions (e.g., introduction of air and/or oxygen) and the second stage may be conducted under anaerobic conditions (e.g., air or oxygen are not introduced to the culture conditions). In some embodiments, the first stage may be conducted under anaerobic conditions and the second stage may be conducted under aerobic conditions. In certain embodiments, a two-stage process may include two more organisms, where one organism generates an intermediate product in one stage and another organism processes the intermediate product into a target fatty dicarboxylic acid product (e.g., sebacic or dodecanedioic acid) in another stage, for example.

A variety of fermentation processes may be applied for commercial biological production of a target fatty dicarboxylic acid product. In some embodiments, commercial production of a target fatty dicarboxylic acid product from a recombinant microbial host is conducted using a batch, fed-batch or continuous fermentation process, for example.

A batch fermentation process often is a closed system where the media composition is fixed at the beginning of the process and not subject to further additions beyond those required for

maintenance of pH and oxygen level during the process. At the beginning of the culturing process the media is inoculated with the desired organism and growth or metabolic activity is permitted to occur without adding additional sources (i.e., carbon and nitrogen sources) to the medium. In batch processes the metabolite and biomass compositions of the system change constantly up to the time the culture is terminated. In a typical batch process, cells proceed through a static lag phase to a high-growth log phase and finally to a stationary phase, wherein the growth rate is diminished or halted. Left untreated, cells in the stationary phase will eventually die.

A variation of the standard batch process is the fed-batch process, where the carbon source is continually added to the fermenter over the course of the fermentation process. Fed-batch processes are useful when catabolite repression is apt to inhibit the metabolism of the cells or where it is desirable to have limited amounts of carbon source in the media at any one time. Measurement of the carbon source concentration in fed-batch systems may be estimated on the basis of the changes of measurable factors such as pH, dissolved oxygen and the partial pressure of waste gases (e.g., CO. sub.2).

Batch and fed-batch culturing methods are known in the art. Examples of such methods may be found in Thomas D. Brock in Biotechnology: A Textbook of Industrial Microbiology, 2.sup.nd ed., (1989) Sinauer Associates Sunderland, Mass. and Deshpande, Mukund V., Appl. Biochem. BiotechnoL, 36:227 (1992).

In continuous fermentation process a defined media often is continuously added to a bioreactor while an equal amount of culture volume is removed simultaneously for product recovery. Continuous cultures generally maintain cells in the log phase of growth at a constant cell density. Continuous or semi-continuous culture methods permit the modulation of one factor or any number of factors that affect cell growth or end product concentration. For example, an approach may limit the carbon source and allow all other parameters to moderate metabolism. In some systems, a number of factors affecting growth may be altered continuously while the cell concentration, measured by media turbidity, is kept constant. Continuous systems often maintain steady state growth and thus the cell growth rate often is balanced against cell loss due to media being drawn off the culture. Methods of modulating nutrients and growth factors for continuous culture processes, as well as techniques for maximizing the rate of product formation, are known and a variety of methods are detailed by Brock, supra.

In some embodiments involving fermentation, the fermentation can be carried out using two or more microorganisms (e.g., host microorganism, engineered microorganism, isolated naturally occurring microorganism, the like and combinations thereof), where a feedstock is partially or completely utilized by one or more organisms in the fermentation (e.g., mixed fermentation), and the products of cellular respiration or metabolism of one or more organisms can be further metabolized by one or more other organisms to produce a desired target product (e.g., sebacic acid, dodecanedioic acid, hexanoic acid). In certain embodiments, each organism can be fermented independently and the products of cellular respiration or metabolism purified and contacted with another organism to produce a desired target product. In some embodiments, one or more organisms are partially or completely blocked in a metabolic pathway (e.g., beta oxidation, omega oxidation, the like or combinations thereof), thereby producing a desired product that can be used as a feedstock for one or more other organisms. Any suitable combination of microorganisms can be utilized to carry out mixed fermentation or sequential fermentation.

Target Product Production, Isolation and Yield

In various embodiments a fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid) is isolated or purified from the culture media or extracted from the engineered microorganisms. In some embodiments, fermentation of feedstocks by methods described herein can produce a target fatty dicarboxylic acid product (e.g., sebacic or dodecanedioic acid) at a level of about 10% to about 100% of theoretical yield (e.g., about 15%, about 20%, about 25% or more of theoretical yield (e.g., 25% or more, 26% or more, 27% or more, 28% or more, 29% or more, 30% or more, 31% or more, 32% or more, 33% or more, 34% or more, 35% or more, 36% or more, 37% or more, 38% or more, 39% or more, 40% or more, 41% or more, 42% or more, 43% or more, 44% or more, 45% or more, 46% or more, 47% or more, 48% or more, 49% or more, 50% or more, 51% or more, 52% or more, 53% or more, 54% or more, 55% or more, 56% or more, 57% or more, 58% or more, 59% or more, 60% or more, 61% or more, 62% or more, 63% or more, 64% or more, 65% or more, 66% or more, 67% or more, 68% or more, 69% or more, 70% or more, 71% or more, 72% or more, 73% or more, 74% or more, 75% or more, 76% or more, 77% or more, 78% or more, 79% or

more, 80% or more, 81% or more, 82% or more, 83% or more, 84% or more, 85% or more, 86% or more, 87% or more, 88% or more, 89% or more, 90% or more, 91% or more, 92% or more, 93% or more, 94% or more, 95% or more, 96% or more, 97% or more, 98% or more, or 99% or more of theoretical yield). The term "theoretical yield" as used herein refers to the amount of product that could be made from a starting material if the reaction is 100% complete. Theoretical yield is based on the stoichiometry of a reaction and ideal conditions in which starting material is completely consumed, undesired side reactions do not occur, the reverse reaction does not occur, and there are no losses in the work-up procedure. Culture media may be tested for target product (e.g., sebacic or dodecanedioic acid) concentration and drawn off when the concentration reaches a predetermined level. Detection methods are known in the art, including but not limited to chromatographic methods (e.g., gas chromatography) or combined chromatographic/mass spectrometry (e.g., GC-MS) methods. Target product (e.g., sebacic or dodecanedioic acid) may be present at a range of levels as described herein.

A target fatty dicarboxylic acid product sometimes is retained within an engineered microorganism after a culture process is completed, and in certain embodiments, the target product is secreted out of the microorganism into the culture medium. For the latter embodiments, (i) culture media may be drawn from the culture system and fresh medium may be supplemented, and/or (ii) target product may be extracted from the culture media during or after the culture process is completed. Engineered microorganisms may be cultured on or in solid, semi-solid or liquid media. In some embodiments media is drained from cells adhering to a plate. In certain embodiments, a liquid-cell mixture is centrifuged at a speed sufficient to pellet the cells but not disrupt the cells and allow extraction of the media, as known in the art. The cells may then be resuspended in fresh media. Target product may be purified from culture media according to known methods know in the art.

Provided herein are non-limiting examples of methods useful for recovering target product from fermentation broth and/or isolating/partially purifying a target fatty dicarboxylic acid product from non-target products when utilizing mixed chain length feedstocks. Recovery of a fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid) from fermentation broth can be accomplished using a variety of methods. Optionally, one can first employ a centrifugation step to separate cell mass and a fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid,

octadecanedioic acid, eicosanedioic acid) from the aqueous phase. A fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid) has limited solubility in water under fermentation conditions, and has a density similar to that of water. Upon centrifugation, the majority of fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid) will be pulled away from the water stream, and be concentrated in the cell mass stream. The concentrated fatty dicarboxylic acid stream will then be further concentrated via filtration steps (e.g., solid dodecanedioic acid will be retained on a filter, allowing water to pass through, concentrating the product). Once the fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid) is concentrated to the desired level, the temperature will be increased to above its melting point of 130 C. After the fatty dicarboxylic acid is melted, the remaining impurities are removed via filtration; the final product is recovered by decreasing the temperature, allowing the fatty dicarboxylic acid to solidify, and collecting the solid product. Alternatively, a fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid) can be recovered from fermentation broth by first extracting the broth with an organic solvent in which a fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid) is soluble (e.g., ethanol). The organic solvent phase can then be filtered through various membranes to further purify the fatty dicarboxylic acid. Subsequent extractions with the same or a different organic solvent can then be performed and each round of extraction can be followed by membrane filtration to further concentrate the fatty dicarboxylic acid. The organic solvent can be evaporated, leaving the fatty dicarboxylic acid behind as a residue and the residue can be dried to provide the fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid) in solid form.

In certain embodiments, target product is extracted from the cultured engineered microorganisms. The microorganism cells may be concentrated through centrifugation at a speed sufficient to shear the cell membranes. In some embodiments, the cells may be physically disrupted (e.g., shear force, sonication) or chemically disrupted (e.g., contacted with

detergent or other lysing agent). The phases may be separated by centrifugation or other method known in the art and target product may be isolated according to known methods.

Commercial grade target product sometimes is provided in substantially pure form (e.g., 90% pure or greater, 95% pure or greater, 99% pure or greater or 99.5% pure or greater). In some embodiments, target product may be modified into any one of a number of downstream products. For example, a fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid) may be polycondensed with hexamethylenediamine to produce nylon. Nylon may be further processed into fibers for applications in carpeting, automobile tire cord and clothing. A fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid) is also used for manufacturing plasticizers, lubricant components and polyester polyols for polyurethane systems. Various esters of food grade fatty dicarboxylic acids (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid) are used as components in fragrance manufacture, gelling aids, flavorings, acidulant, leavening and buffering agent. A fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid) has two carboxylic acid (-COOH) groups, which can yield two kinds of salts. Its derivatives, acyl halides, anhydrides, esters, amides and nitriles, are used in making a variety of downstream products through further reactions of substitution, catalytic reduction, metal hydride reduction, diborane reduction, keto formation with organometallic reagents, electrophile bonding at oxygen, and condensation.

Target product may be provided within cultured microbes containing target product, and cultured microbes may be supplied fresh or frozen in a liquid media or dried. Fresh or frozen microbes may be contained in appropriate moisture-proof containers that may also be temperature controlled as necessary. Target product sometimes is provided in culture medium that is substantially cell-free. In some embodiments target product or modified target product purified from microbes is provided, and target product sometimes is provided in substantially pure form. In certain embodiments crystallized or powdered target product is provided. Dodecanedioic acid (1,12 dodecanedioic acid; DDDA) is a white powder or crystal with a melting point of between 260°F and 266°F. Sebacic acid (1,8 ocatanedicarboxylic acid) is also a white powder or crystal with a melting point of between 268°F and 274°F. A crystallized or

powdered fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid) may be transported in a variety of containers including one ton cartons, drums, 50 pound bags and the like.

In certain embodiments, a fatty dicarboxylic acid target product (e.g., dodecanedioic acid or sebacic acid) is produced with a yield of about 0.50 grams of target product per gram of feedstock added, or greater; 0.51 grams of target product per gram of feedstock added, or greater; 0.52 grams of target product per gram of feedstock added, or greater; 0.53 grams of target product per gram of feedstock added, or greater; 0.54 grams of target product per gram of feedstock added, or greater; 0.55 grams of target product per gram of feedstock added, or greater; 0.56 grams of target product per gram of feedstock added, or greater; 0.57 grams of target product per gram of feedstock added, or greater; 0.58 grams of target product per gram of feedstock added, or greater; 0.59 grams of target product per gram of feedstock added, or greater; 0.60 grams of target product per gram of feedstock added, or greater; 0.61 grams of target product per gram of feedstock added, or greater; 0.62 grams of target product per gram of feedstock added, or greater; 0.63 grams of target product per gram of feedstock added, or greater; 0.64 grams of target product per gram of feedstock added, or greater; 0.65 grams of target product per gram of feedstock added, or greater; 0.66 grams of target product per gram of feedstock added, or greater; 0.67 grams of target product per gram of feedstock added, or greater; 0.68 grams of target product per gram of feedstock added, or greater; 0.69 grams of target product per gram of feedstock added, or greater; 0.70 grams of target product per gram of feedstock added or greater; 0.71 grams of target product per gram of feedstock added, or greater; 0.72 grams of target product per gram of feedstock added, or greater; 0.73 grams of target product per gram of feedstock added, or greater; 0.74 grams of target product per gram of feedstock added, or greater; 0.75 grams of target product per gram of feedstock added, or greater; 0.76 grams of target product per gram of feedstock added, or greater; 0.77 grams of target product per gram of feedstock added, or greater; 0.78 grams of target product per gram of feedstock added, or greater; 0.79 grams of target product per gram of feedstock added, or greater; 0.80 grams of target product per gram of feedstock added, or greater; 0.81 grams of target product per gram of feedstock added, or greater; 0.82 grams of target product per gram of feedstock added, or greater; 0.83 grams of target product per gram of feedstock added, or greater; 0.84 grams of target product per gram of feedstock added, or greater; 0.85 grams of target product per gram of feedstock added, or greater; 0.86 grams of target product per gram of

feedstock added, or greater; 0.87 grams of target product per gram of feedstock added, or greater; 0.88 grams of target product per gram of feedstock added, or greater; 0.89 grams of target product per gram of feedstock added, or greater; 0.90 grams of target product per gram of feedstock added, or greater; 0.91 grams of target product per gram of feedstock added, or greater; 0.92 grams of target product per gram of feedstock added, or greater; 0.93 grams of target product per gram of feedstock added, or greater; 0.94 grams of target product per gram of feedstock added, or greater; 0.95 grams of target product per gram of feedstock added, or greater; 0.96 grams of target product per gram of feedstock added, or greater; 0.97 grams of target product per gram of feedstock added, or greater; 0.98 grams of target product per gram of feedstock added, or greater; 0.99 grams of target product per gram of feedstock added, or greater; 1.0 grams of target product per gram of feedstock added, or greater; 1.1 grams of target product per gram of feedstock added, or greater; 1.2 grams of target product per gram of feedstock added, or greater; 1.3 grams of target product per gram of feedstock added, or greater; 1.4 grams of target product per gram of feedstock added, or greater; or about 1.5 grams of target product per gram of feedstock added, or greater.

In certain embodiments, a fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid) is produced with a yield of greater than about 0.15 grams per gram of the feedstock (e.g., dodecane, mixed chain length alkanes, lauric acid, mixed chain length fatty acids, oil, the like or combinations of the foregoing). In some embodiments, a fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid) is produced at between about 10% and about 100% of maximum theoretical yield of any introduced feedstock ((e.g., about 15%, about 20%>, about 25%> or more of theoretical yield (e.g., 25%> or more, 26%> or more, 27%> or more, 28%> or more, 29%> or more, 30%> or more, 31%> or more, 32%> or more, 33%o or more, 34%> or more, 35%> or more, 36%> or more, 37%> or more, 38%> or more, 39%> or more, 40% or more, 41% or more, 42% or more, 43% or more, 44% or more, 45% or more, 46%o or more, 47%> or more, 48%> or more, 49%> or more, 50%> or more, 51%> or more, 52%> or more, 53%> or more, 54%> or more, 55%> or more, 56%> or more, 57%> or more, 58%> or more, 59%o or more, 60%> or more, 61%> or more, 62%> or more, 63%>or more, 64%> or more, 65%>or more, 66%> or more, 67%> or more, 68%> or more, 69%> or more, 70%> or more, 71%> or more, 72%o or more, 73%> or more, 74%> or more, 75%> or more, 76%> or more, 77% or more, 78% or more, 79% or more, 80%> or more, 81% or more, 82% or more, 83% or more, 84% or more,

85% or more, 86% or more, 87% or more, 88% or more, 89% or more, 90% or more, 91% or more, 92% or more, 93% or more, 94% or more, 95% or more, 96% or more, 97% or more, 98% or more, or 99% or more of theoretical maximum yield). In certain embodiments, a fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid) is produced in a concentration range of between about 50 g/L to about lOOOg/L of culture media (e.g., about 50 g/L, about 55 g/L, about 60 g/L, about 65 g/L, about 70 g/L, about 75 g/L, about 80 g/L, about 85 g/L, about 90 g/L, about 95 g/L, about 100 g/L, about 110 g/L, about 120 g/L, about 130 g/L, about 140 g/L, about 150 g/L, about 160 g/L, about 170 g/L, about 180 g/L, about 190 g/L, about 200 g/L, about 225 g/L, about 250 g/L, about 275 g/L, about 300 g/L, about 325 g/L, about 350 g/L, about 375 g/L, about 400 g/L, about 425 g/L, about 450 g/L, about 475 g/L, about 500 g/L, about 550 g/L, about 600 g/L, about 650 g/L, about 700 g/L, about 750 g/L, about 800 g/L, about 850 g/L, about 900 g/L, about 950 g/L, or about 1000 g/L).

In some embodiments, a fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid) is produced at a rate of between about 0.5 g/L/hour to about 5 g/L/hour (e.g., about 0.5 g/L/hour, about 0.6 g/L/hour, about 0.7 g/L/hour, about 0.8 g/L/hour, about 0.9 g/L/hour, about 1.0 g/L/hour, about 1.1 g/L/hour, about 1.2 g/L/hour, about 1.3 g/L/hour, about 1.4 g/L/hour, about 1.5 g/L/hour, about 1.6 g/L/hour, about 1.7 g/L/hour, about 1.8 g/L/hour, about 1.9 g/L/hour, about 2.0 g/L/hour, about 2.25 g/L/hour, about 2.5 g/L/hour, about 2.75 g/L/hour, about 3.0 g/L/hour, about 3.25 g/L/hour, about 3.5 g/L/hour, about 3.75 g/L/hour, about 4.0 g/L/hour, about 4.25 g/L/hour, about 4.5 g/L/hour, about 4.75 g/L/hour, or about 5.0 g/L/hour.) In certain, embodiments, the engineered organism comprises between about a 5-fold to about a 500-fold increase in a fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid) production when compared to wild-type or partially engineered organisms of the same strain, under identical fermentation conditions (e.g., about a 5-fold increase, about a 10-fold increase, about a 15-fold increase, about a 20-fold increase, about a 25-fold increase, about a 30-fold increase, about a 35-fold increase, about a 40-fold increase, about a 45-fold increase, about a 50-fold increase, about a 55-fold increase, about a 60-fold increase, about a 65-fold increase, about a 70-fold increase, about a 75-fold increase, about a 80-fold increase, about a 85-fold increase, about a 90-fold increase, about a 95-fold increase, about a 100-fold increase, about a 125-fold increase, about a 150-fold increase, about a 175-fold increase, about

a 200-fold increase, about a 250-fold increase, about a 300-fold increase, about a 350-fold increase, about a 400-fold increase, about a 450-fold increase, or about a 500-fold increase).

In certain embodiments, the maximum theoretical yield (Y_{max}) of dodecanedioic acid in a fully beta-oxidation blocked engineered microorganism is about 1.15 grams of dodecanedioic acid produced per gram of lauric acid added. In some embodiments, the maximum theoretical yield (Y_{max}) of dodecanedioic acid in a fully beta-oxidation blocked engineered microorganism is about 1.07 grams of dodecanedioic acid produced per gram of methyl laurate added. In certain embodiments, the maximum theoretical yield (Y_{max}) of dodecanedioic acid in a partially beta-oxidation blocked engineered microorganism is about 0.82 grams of dodecanedioic acid produced per gram of oleic acid added. In some embodiments, the maximum theoretical yield (Y_{max}) of dodecanedioic acid in a partially beta-oxidation blocked engineered microorganism is about 0.95 grams of dodecanedioic acid produced per gram of coconut oil added. The percentage of Y_{max} for the engineered microorganism under conditions in which dodecanedioic acid is produced is calculated as ($\%Y_{max}$) = $Y_{p/S}$ / Y_{max} * 100, where ($Y_{p/S}$) = [dodecanedioic acid (g/L)] * final volume of culture in flask (L)] / [feedstock added to flask (g)]. In some embodiments, the engineered microorganism produces dodecanedioic acid at about 10% to about 100% of maximum theoretical yield.

In certain embodiments, the maximum theoretical yield (Y_{max}) of sebacic acid in a fully beta-oxidation blocked engineered microorganism is about 1.42 grams of sebacic acid produced per gram of decane added. In some embodiments, the maximum theoretical yield (Y_{max}) of sebacic acid in a fully beta-oxidation blocked engineered microorganism is about 1.17 grams of sebacic acid produced per gram of capric acid added. In certain embodiments, the maximum theoretical yield (Y_{max}) of sebacic acid in a partially beta-oxidation blocked engineered microorganism is about 0.83 grams of sebacic acid produced per gram of coconut oil added. In some embodiments, the maximum theoretical yield (Y_{max}) of sebacic acid in a partially beta-oxidation blocked engineered microorganism is about 0.72 grams of sebacic acid produced per gram of oleic acid added. The percentage of Y_{max} for the engineered microorganism under conditions in which sebacic acid is produced is calculated as ($\%>Y_{max}$) = $Y_{p/S}$ / Y_{max} * 100, where ($Y_{p/S}$) = [sebacic acid (g/L)] * final volume of culture in flask (L)] / [feedstock added to flask (g)]. In some embodiments, the engineered microorganism produces sebacic acid at about 10%> to about 100% of maximum theoretical yield.

Examples

The examples set forth below illustrate certain embodiments and do not limit the technology. Certain examples set forth below utilize standard recombinant DNA and other biotechnology protocols known in the art. Many such techniques are described in detail in Maniatis, T., E. F. Fritsch and J. Sambrook (1982) *Molecular Cloning: a Laboratory Manual;* Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. DNA mutagenesis can be accomplished using the Stratagene (San Diego, CA) "QuickChange" kit according to the manufacturer's instructions.

Non-limiting examples of recombinant DNA techniques and genetic manipulation of microorganisms are described herein. In some embodiments, strains of engineered organisms described herein are mated to combine genetic backgrounds to further enhance carbon flux management through native and/or engineered pathways described herein, for the production of a desired target product (e.g., sebacic or dodecanedioic acid).

Example 1: Conversion of decane to sebacic acid in shakeflask fermentation

50mL of SP92 medium (6.7g/L yeast nitrogen base, 3.0 g/L yeast extract, 3.0 g/L $(NH_4)_2S0_4$, 1.0 g/L K_2HP0_4, 1.0 g/L KH_2P0_4, 75 g/L dextrose) was inoculated with a single colony of a completely beta-oxidation blocked strain of *Candida tropicalis* (ATCC20962) and the culture was grown overnight at 30°C, with shaking at about 300rpm. Cells were pelleted by centrifugation for 10 minutes at 4°C and 1,050 x *g* and the supernatant discarded. Cells were resuspended in 20mL TB- low nitrogen (low-N) media (1.7g/L yeast nitrogen base without ammonium sulfate, 3.0 g/L yeast extract, 1.0 g/L K_2HP0_4, 1.0 g/L KH_2P0_4) and transferred to a new sterile 250mL glass baffled flask and incubated at 30C, with shaking at about 250rpm, utilizing the following feeding schedule: dextrose fed to 0.1% at 0, 1, 2, 3, 4, and 5 hours, dextrose fed to 5% at 30 hours, decane fed to 0.7% at 0, 5, 30, and 48 hours. Samples were removed for gas chromatographic (GC) analysis at 0, 4, 30, and 72 hours. The GC profile showed that the culture accumulated the CIO dicarboxylic acid (sebacic acid) with very little accumulation of the CIO monocarboxylic acid (capric acid), as shown in FIG. 9. After 72 hours of incubation the concentration of sebacic acid was 0.94 g/L and the capric acid concentration was 0.01 g/L. There was no significant accumulation of any other monoacid or diacid.

Example 2: Conversion of capric acid to sebacic acid in shakeflask fermentation

5 mL of SP92-glycerol medium (6.7 g/L yeast nitrogen base, 3.0 g/L yeast extract, 3.0 g/L (NH4)2S04, 1.0 g/L K2HP04, 1.0 g/L KH2P04, 75 g/L glycerol) was inoculated with a single colony of Candida tropicalis (ATCC20962) and the starter culture was grown overnight at 30°C, with shaking at about 250rpm. Variations of SP92 media recipes are known, non-limiting examples of which include the addition of dextrose and/or glycerol, the like or combinations thereof. SP92 media, as referred to herein, can include dextrose and/or glycerol. The starter culture was then used to inoculate 25 mL cultures in the same medium to an initial OD_{600nm} of 0.4 and grown overnight at 30°C, with shaking at about 300 rpm. Cells were pelleted by centrifugation for 10 minutes at 4°C and 1,050 x g and the supernatant discarded. Cells were resuspended in 12.5 mL TB-lowN media + glycerol (1.7 g/L yeast nitrogen base without ammonium sulfate, 3.0 g/L yeast extract, 1.0 g/L K2HP04, 1.0 g/L KH2P04, 75 g/L glycerol) and transferred to a new sterile 250 mL glass baffled flask. Cultures were fed 0.05% or 0.1% capric acid and incubated at 30°C, with shaking at about 300 rpm. After 24 hours incubation cultures were fed glycerol to 75 g/L and incubation continued before sampling for GC at 48 hours. GC analysis showed that nearly all capric acid was converted to sebacic acid under both starting concentrations of capric acid, as shown in FIG. 10.

Example 3: Fermentation procedure for conversion of decane to sebacic acid

Filter sterilized modified SP92-glycerol fermentation medium (6.7 g/L yeast nitrogen base, 3.0 g/L yeast extract, 3.0 g/L (NH4)2S04, 1.0 g/L K2HP04, 1.0 g/L KH2P04, 20 g/L glycerol) is transferred to a sterile fermentation vessel. Growth of Candida tropicalis (ATCC20962) is innoculated to an initial OD_{600nm} of about 1.0 with a 5% inoculum and growth carried out under the following conditions: 30°C with shaking at about 1000 rpm, 1 volume per volume per minute aeration (wm), pH 5.8 and initial volume of 0.3 L. Growth proceeds for approximately 8 hours and the conversion phase is initiated by the addition of decane to 2 g/L. Continuous feeds for decane (1 g/L-h) and glucose (1.5 g/L-h) are initiated at the same time as the addition of the decane bolus. Fermentation conditions are maintained at 30°C, 1000 rpm, 1 wm, and pH 5.8 for 44 hours.

Samples were collected for GC analysis at 44 hours after initiating the conversion phase. The data, presented in FIG. 16, show that the decane was converted exclusively to the CIO

dicarboxylic acid, sebacic acid. Significant evaporative losses from the decane feed bottles prevented an accurate determination of product yield.

Example 4: Conversion of mixedfatty acidfeedstock to mixed diacidproducts containing sebacic acid in shakeflask fermentation

5 mL of SP92-glycerol medium (6.7 g/L yeast nitrogen base, 3.0 g/L yeast extract, 3.0 g/L (NH4)2S04, 1.0 g/L K2HP04, 1.0 g/L KH2P04, 75 g/L glycerol) is inoculated with a single colony of Candida tropicalis (ATCC20962) and grown as described in Example 2. 25 mL of the same media is inoculated using overnight cultures to an initial ODeoonm of 0.4 and grown overnight at 30°C, with shaking at about 300 rpm. Cells are pelleted by centrifugation for 10 minutes at 4°C and 1,050 x g and the supernatant discarded. Cells are resuspended in 12.5 mL TB-lowN media without carbon source (1.7 g/L yeast nitrogen base without ammonium sulfate, 3.0 g/L yeast extract, 1.0 g/L K2HP04, 1.0 g/L KH2P04) and transferred to a new sterile 250 mL glass baffled flask. Cultures are fed 0.05% capric acid, 0.05% methyl laurate, and 30 g/L glycerol and incubated at 30°C, 300 rpm. After 24 hours of incubation cultures are sampled for GC analysis.

The results, presented in FIG. 17, show that the C12 and CIO fatty acids were converted to dicarboxylic acids of the same chain length (e.g., C12 and CIO dicarboxylic acids), with no evidence of chain shortening of the diacids (e.g., no significant levels of monocarboxylic acids were detected).

Example 5: Conversion of long chainfatty acids to mixed diacids

SP92 fermentation medium was filter sterilized and transferred to a sterile fermentation vessel. Growth of a partially beta-oxidation blocked strain of Candida tropicalis (sAA106) was initiated with a 10%> inoculum (initial OD6oonm=3.0) and grown under the following conditions: of 30°C with shaking at about 1200 rpm, 1 wm, pH 6.1 and initial volume of 0.3 L. Growth continued until the glucose concentration dropped to less than 2 g/L at which time the conversion phase was initiated by increasing the pH to 8.0 by the addition of 6N KOH and by

the addition of methyl myristate to 30 g/L. Immediately following the methyl myristate bolus a continuous feed of glucose was initiated at a rate of 1.5 g/L-h. Fermentation conditions were maintained at 30°C, 1200 rpm, 1 wm, and pH 8.0 for 90 hours with boluses of 30 g/L methyl myristate at 24, 48, and 72 hours after initiation of conversion. Samples for GC were collected at 24, 48, 72, and 90 hours. The diacid profile graphically illustrated in FIG. 11 shows an accumulation of dicarboxylic acids ranging in chain-length from 6 to 14 carbons long, including sebacic acid. The methyl myristate substrate (methyl ester of myristic acid) is first converted to the C14 dicarboxylic acid via the □-oxidation pathway before being shortened by two carbon increments via the cyclic β-oxidation pathway. The glucose co-feed employed during the fermentation represses the β-oxidation pathway such that all chain-lengths of diacid accumulate. Manipulation of diacid chain-length distribution is being investigated by altering the glucose co-feed rate in the fermentation medium, thereby allowing growth under varying glucose concentrations.

Example 6: Fermentation procedure for conversion of mixed long-chainfatty acids to mixed diacids of shorter chain length

SP92 fermentation medium without glycerol was filter sterilized and transferred to a sterile fermentation vessel. Autoclaved virgin coconut oil was added to the vessel to a final concentration of 80 g/L. A partially beta-oxidation blocked Candida tropicalis strain (sAA496) was inoculated to an initial OD_{600nm} of 1.0 with a 5% inoculum and grown under the following conditions: 30°C with shaking at about 1200 rpm, 1 wm, initial pH 6.5 and initial volume of 1.0 L. The effect of pH on the distribution of fatty acid chain lengths was determined by manipulating the pH of the fermentation media. The pH of the fermentation was either 1) increased to pH 7.5 and controlled at that pH for the entire run, 2) allowed to drop naturally due to the growth of the culture before controlling at pH 6.0 for the rest of the run, or 3) allowed to drop naturally due to the growth of the culture before controlling at pH 4.5 for the rest of the run. Samples were collected for GC analysis after 140 hours of fermentation time. The product diacid composition was shown to shift to longer chain diacids with increasing pH, as shown in TABLE 6.

TABLE 6

Diacid composition (fraction of total diacids)			
C12 Diacid	Sebacic Acid	Suberic Acid	Adipic Acid

pH 4.5	0.00	0.00	0.68	0.32
pH 6.0	0.03	0.10	0.75	0.12
pH 7.5	0.16	0.17	0.62	0.05

Example 7: Conversion of capric acid to sebacic acid in shakeflask fermentations usingfully beta-oxidation blocked strains having additional genetic modifications in the omega oxidation pathway.

Various genetically modified strains of Candida tropicalis were inoculated into 5 mL of SP92 medium (6.7 g/L yeast nitrogen base, 3.0 g/L yeast extract, 3.0 g/L $(NH_4)_2S0_4$, 1.0 g/L K_2HP0_4, 1.0 g/L KH_2P0_4, 75 g/L glycerol). The strains included a completely beta-oxidation blocked strain of *Candida tropicalis* (sAA003), as well as derivatives of sAA003 with amplified components of the omega-oxidation pathway (e.g., various cytochrome P450s, cytochrome P450 reductase or combinations thereof) and the cultures grown overnight at 30°C, with shaking at about 250 rpm. These starter cultures were then used to inoculate 25 mL cultures in the same medium and grown overnight at 30°C, with shaking at about 250 rpm. Cells were pelleted by centrifugation for 10 minutes at 4°C and 1,050 x *g* and the supernatant discarded. Cells were resuspended in 12.5 mL TB-lowN media + glycerol (1.7 g/L yeast nitrogen base without ammonium sulfate, 3.0 g/L yeast extract, 1.0 g/L K_2HP0_4, 1.0 g/L KH_2P0_4, 75 g/L glycerol) and transferred to a new sterile 250 mL glass baffled flask. Cultures were fed 0.05% from a 5% capric acid solution in ethanol and incubated at 30°C, with shaking at about 300 rpm. After 24 hours incubation cultures were fed glycerol to 30 g/L and an additional bolus of 0.05%> capric acid. Incubation continued before sampling for GC at 24, 48, and 72 hours. The results are shown in FIG. 12. GC analysis showed that a greater proportion of capric acid was converted to sebacic acid when particular elements of the omega-oxidation pathway are amplified. The data are presented as % of theoretical maximum yield. Strains which include genetic modifications to CYPA18 and CYPA19 achieve approximately 80% of theoretical maximum yield in conversion of capric acid to sebacic acid. The strain designated +CPR+A18 has about 30 copies of CYPA18, whereas the strain designated +CPR+A19 has about 7 copies of CYPA19.

Example 8: Conversion of methyl-laurate to dodecanedioic acid in shakeflask fermentation

5 mL of SP92 glycerol medium (6.7 g/L yeast nitrogen base, 3.0 g/L yeast extract, 3.0 g/L (NH4)2S04, 1.0 g/L K2HP04, 1.0 g/L KH2P04, 75 g/L glycerol) was inoculated with a single colony of a completely beta-oxidation blocked strain of Candida tropicalis (ATCC20962), as well as, modified derivatives of this strain with amplified components of the omega-oxidation pathway, and the cultures grown overnight at 30°C, with shaking at about 250rpm. The starter cultures were then used to inoculate 25 mL cultures of the same medium and grown overnight at 30°C, with shaking at about 250 rpm. Cells were pelleted by centrifugation for 10 minutes at 4°C and 1,050 x g and the supernatant discarded. Cells were resuspended in 12.5 mL SP92 glycerol medium and transferred to a sterile 250 mL glass baffled flask. Cultures were fed 2% (v/v) methyl laurate and incubated at 30°C, with shaking at about 300 rpm. After 24 hours incubation, cultures were fed glycerol to 60 g/L and incubation continued before sampling for GC at 48 hours. GC analysis showed that amplification of certain components of the omega oxidation pathway allow for increased conversion to dodecanedioic acid (FIG. 13).

Example 9: Alteration ofAcyl CoA Oxidase Substrate Specificity

The substrate specificity of the peroxisomal acyl-CoA oxidase enzymes POX4 and POX5 have been shown to be involved in the control of the diacid product chain-length in fermentations of *Candida tropicalis* fed a mixed chain-length fatty acid feedstock. Reduction or elimination of POX4 activity, POX5 activity or POX4 activity and POX5 activity, effects the carbon chain-length distribution of dicarboxylic acids produced in Candida spp.. Acyl-CoA oxidase is the first enzyme in the cyclic beta-oxidation pathway that shortens a substrate by two carbons each cycle. Thus the acyl-CoA oxidase activity serves as the pathway entry point for substrates entering into the beta-oxidation pathway. Altering the substrate specificity an acyl-CoA oxidase activity such that it is not active on substrate carbon chains shorter than a desired carbon chain length (e.g., C8, CIO, C12, C14 and the like), can inhibit shortening of carbon chains below a chosen threshold, allowing accumulation of a desired target chain length and product (e.g., C12, dodecanedioic acid).

The native acyl-CoA oxidase isozymes in *Candida strain ATCC20336,* Pox4p and Pox5p have different substrate specificities. The Pox4p isozyme has a broad substrate specificity while the Pox5p isozyme has a narrow substrate specificity. In strains that are $Pox4^{-}$, $Pox5^{+}$ the chain length of the diacid product is determined by the substrate specificity of the Pox5p isozyme and the main product is adipic acid.

To maximize production of desired diacid products of longer chain lengths (e.g., C12) in fermentations, genetically modified organisms containing an acyl-CoA oxidase activity with a substrate chain-length specificity appropriate for the chain-length of the desired diacid product can be engineered, in some embodiments. The source of the acyl-CoA oxidase activity or the method of engineering the acyl-CoA oxidase activity may vary. Non-limiting examples of organisms which can be used to provide polynucleotide sequences suitable for use in engineering altered substrate specificity acyl-CoA oxidase activities include; plants (e.g., Arabidopsis, Cucurbita (e.g., pumpkin, squash), Oryza (e.g., rice)); animals (e.g., Bos (e.g., bovine), Cavia (e.g., guinea pig), Mus (e.g., mouse), Rattus (e.g., rat), Phascolarctos (e.g., Koala), primates (e.g., orangutans)); molds (e.g., Dictyostelium (e.g., slime molds)); insects (e.g., Drosophila); Yeast (e.g., Yarrowia lipolytica, Candida maltosa, Candida glabrata, Ashbya gossypii, Debaryomyces hansenii, Kluyveromyces lactis, Pichia pastoris, Saccharomyces cerevisiae); bacteria (e.g., Eschericia coli); cyanobacteria; nematodes (e.g., Caenorhabditis); and humans.

Acyl-CoA oxidase activities with different substrate chain-length specificities can be identified by:

1) Selecting acyl-CoA oxidase genes from heterologous organisms that contain different substrate chain-length specificities. The identified genes can be transferred into a *Candida* strain deleted for all acyl-CoA oxidase activity. The only acyl-CoA oxidase activity detectable in such a genetically modified organism may be that imparted by the heterologous gene.

2) Engineering an acyl-CoA oxidase gene library by domain swapping from multiple acyl-CoA oxidase genes to produce a library of non-native chimeric acyl-CoA oxidase genes. The library of chimeric genes can be transferred into a strain of *Candida* deleted for all acyl-CoA oxidase activity. The only detectable acyl-CoA oxidase activity may be that imparted by an engineered gene from the library of non-native chimeric acyl-CoA oxidase genes.

3) Engineering an acyl-CoA oxidase gene library by random mutagenesis. A naturally occurring or engineered acyl-CoA oxidase activity with a substrate chain-length

specificity close to that desired can be used as the basis for random mutagenesis, followed by screening and/or selection in an effort to generate and identify an altered activity with the desired substrate chain-length specificity. The library of genes can be transferred into a *Candida* strain deleted for all acyl-CoA oxidase activity. The only detectable acyl-CoA oxidase activity may be that imparted by the gene from the randomly mutagenized library.

4) Engineering an acyl-CoA oxidase gene by intelligent design and directed mutation using protein structural information to guide the position and identity of the amino acid(s) to be replaced. The engineered gene(s) can be transferred into a *Candida* strain deleted for all acyl-CoA oxidase activity. The only detectable acyl-CoA oxidase activity may be that imparted by the engineered gene(s).

A non-limiting example of a post-engineering method for selecting genes that impart the desired substrate chain-length specificity is provided herein. Selection is performed by growth on substrates of different chain lengths that are provided as the only carbon source. Growth of the cells on certain substrates but not others often reflects the substrate chain-length specificity of the acyl-CoA oxidase enzyme present in the strain. *Candida tropicalis* can utilize alkanes provided in the gas phase as its sole carbon source for growth. Alkanes of different chain lengths are provided by soaking a filter paper in the appropriate alkane, and inverting a solid growth media without a carbon source over the filter paper, with each specific carbon source (e.g., specific chain length alkane) provided in a different petri dish. Serially diluted Candida carrying the altered specificity acyl-CoA oxidase genes are spotted on the solid growth media as a growth selection for the chain-length specificity of the acyl-CoA oxidase enzyme in each strain. Shown in FIGS. 14 and 15 are a schematic representation of the selection process, which provides an alkane as a gas phase carbon source, as described herein. The solid growth media is an agar medium containing yeast nitrogen base without amino acids or any other carbon source. The plated cells are inverted over a lid containing a filter paper soaked with an alkane of appropriate chain length that evaporates and provides the carbon source through the gas phase, as shown in FIG. 14.

Candida strains containing altered acyl-CoA oxidase activities generated as described herein are selected and/or screened using the method described herein. Strains carrying different altered acyl-CoA oxidase activities (e.g., strain 1 (S1), strain 2 (S2), strain 3 (S3), strain 4 (S4)) are

grown overnight in a rich medium (e.g., YPD). Overnight cultures are centrifuged and washed to remove any traces of residual rich medium and serial dilutions of the cells are prepared in a phosphate buffered solution. The serial dilutions of each strain are spotted onto multiple YNB agar plates (growth medium having no amino acids or other carbon sources), the individual plates inverted over filter papers soaked in the appropriate chain length alkane, and the plate incubated at 30°C. The growth of the strains is dependent upon the chain-length specificity of the acyl-CoA oxidase. In order to utilize the particular alkane for growth the provided chain-length must be able to enter the beta-oxidation pathway. The shortest chain-length at which a certain strain is able to grow indicates the shortest chain-length of the acyl-CoA oxidase isozymes substrate specificity. An example is provided in FIG. 15. FIG. 15 illustrates that strain S4 can grow on decane, but is unable to grow on octane. Therefore the modified acyl-CoA oxidase activity of strain S4 has a substrate chain-length specificity that inhibits the utilization of 8 carbon molecules and the diacid product from fermentations with this strain typically result in an 8 carbon diacid. Acyl-CoA oxidase activities with any desired specificity can be selected and/or screened using the method described herein.

It will be understood that the example presented herein is a generalize method used to describe the selection/screening process. The feedstocks used for the selection and screening process are altered to suit the acyl-CoA oxidase activity being sought. For example, for acyl-CoA oxidases having specificity for longer chain substrates, feedstocks having longer carbon chain lengths could be substituted to allow selection and or screening for acyl-CoA oxidase activities with specificities for longer carbon chain lengths.

Example 10: Transformation of Candida spp. procedure

5mL YPD start cultures were inoculated with a single colony of Candida strain ATCC20336 and incubated overnight at 30°C, with shaking at about 200rpm. The following day, fresh 25mL YPD cultures, containing 0.05% Antifoam B, were inoculated to an initial OD_{600nm} of 0.4 and the culture incubated at 30°C, with shaking at about 200rpm until an OD_{600nm} of 1.0-2.0 was reached. Cells were pelleted by centrifugation at 1,000 x g, 4°C for 10 minutes. Cells were washed by resuspending in lOmL sterile water, pelleted, resuspended in ImL sterile water and transferred to a 1.5mL microcentrifuge tube. The cells were then washed in ImL sterile TE/LiOAC solution, pH 7.5, pelleted, resuspended in 0.25mL TE/LiOAC solution and incubated with shaking at 30°C for 30 minutes.

The cell solution was divided into 50uL aliquots in 1.5mL tubes to which was added 5-8ug of linearized DNA and 5uL of carrier DNA (boiled and cooled salmon sperm DNA, lOmg/mL). 300uL of sterile PEG solution (40% PEG 3500, IX TE, IX LiOAC) was added, mixed thoroughly and incubated at 30°C for 60 minutes with gentle mixing every 15 minutes. 40uL of DMSO was added, mixed thoroughly and the cell solution was incubated at 42°C for 15 minutes. Cells were then pelleted by centrifugation at 1,000 x *g* 30 seconds, resuspended in 500uL of YPD media and incubated at 30°C with shaking at about 200rpm for 2 hours. Cells were then pelleted by centrifugation and resuspended in lmL IX TE, cells were pelleted again, resuspended in 0.2mL IX TE and plated on selective media. Plates were incubated at 30°C for growth of transformants.

Example 11: Procedure for recycling of the URA3 marker

The URA3 gene was obtained from genomic DNA of Candida yeast culture ATCC20336. Candida strain ATCC20336 has a limited number of selectable marker, as compared to S. cerevisiae, therefore, the URA3 marker is "recycled" to allow multiple rounds of selection using URA3. To reutilize the URA3 marker for subsequent engineering of Candida spp., a single colony having the Ura$^+$ phenotype was inoculated into 3 mL YPD and grown overnight at 30°C with shaking. The overnight culture was then harvested by centrifugation and resuspended in 1 mL YNB+YE (6.7 g/L Yeast Nitrogen Broth, 3g/L Yeast Extract). The resuspended cells were then serially diluted in YNB+YE and 100 uL aliquots plated on YPD plates (incubation overnight at 30°C) to determine titer of the original suspension. Additionally, triplicate 100 uL aliquots of the undiluted suspension were plated on SC Dextrose (Bacto Agar 20g/L, Uracil 0.3 g/L, Dextrose 20 g/L, Yeast Nitrogen Broth 6.7 g/L, Amino Acid Dropout Mix 2.14 g/L) and 5-FOA.at 3 different concentrations (0.5, 0.75, 1 mg/mL).

Plates were incubated for at least 5 days at 30°C. Colonies arising on the SC Dextrose + 5-FOA plates were resuspended in 50 uL sterile, distilled water and 5 uL utilized to streak on to YPD and SC-URA (SC Dextrose medium without Uracil) plates. Colonies growing only on YPD and not on SC-URA plates were then inoculated into 3 mL YPD and grown overnight at 30°C with shaking. Overnight cultures were harvested by centrifugation and resuspended in 1.5 mL YNB (6.7 g/L Yeast Nitrogen Broth). The resuspended cells were serially diluted in YNB

and 100 uL aliquots plated on YPD plates and incubation overnight at 30°C to determine initial titer. 1 mL of each undiluted cell suspension also was plated on SC-URA and incubated for up to 7 days at 30°C. Colonies on the SC-URA plates are revertants and the isolate with the lowest reversion frequency (< 10^{-7}) was used for subsequent strain engineering.

Example 12: Cloning and Analysis of Candida Fatty alcohol oxidase (FAO) alleles

Isolation of fatty alcohol oxidase genes from Candida

Candida strain (ATCC20336) fatty alcohol oxidase genes were isolated by PCR amplification using primers generated to amplify the sequence region covering promoter, fatty alcohol oxidase gene (FAO) and terminator of the FAOl sequence (GenBank accession number of FAOl AY538780). The primers used to amplify the fatty alcohol oxidase nucleotide sequences from Candida strain ATCC20336 strain ATCC20336, are showing in TABLE 7 below.

TABLE 7

Oligonucleotides for cloning FAO alleles	
Oligo	Sequence
oAAOl 44	AACGACAAGATTAGATTGGTTGAGA
oAAOl 45	GTCGAGTTTGAAGTGTGTGTCTAAG
oAA02 68	AGATCTCATATGGCTCCATTTTTGCCCGACCAGGTCGACTACAAACACGTC
oAA02 69	ATCTGGATCCTCATTACTACAACTTGGCTTTGGTCTTCAAGGAGTCTGCCAAACCTAAC
oAA02 82	ACATCTGGATCCTCATTACTACAACTTGGCCTTGGTCT
oAA04 2 1	CACACAGCTCTTCTAGAATGGCTCCATTTTTGCCCGACCAGGTCGAC
oAA04	CACACAGCTCTTCCTTTCTACAACTTGGCTTTGGTCTTCAAGGAGTCTGC

Oligonucleotides for cloning FAO alleles	
22	
oAA0429	GTCTACTGATTCCCCTTTGTC
oAA0281	TTCTCGTTGTACCCGTCGCA

PCR reactions contained 25uL 2X master mix, 1.5 uL of oAA0144 and oAA0145 (lOuM), 3.0uL genomic DNA, and 19uL sterile H_20. Thermocycling parameters used were 98°C for 2 minutes, 35 cycles of 98°C 20 seconds, 52°C 20 seconds, 72°C 1 minute, followed by 72°C 5 minutes and a 4°C hold. PCR products of the correct size were gel purified, ligated into pCR-Blunt II-TOPO (Invitrogen) and transformed into competent TOP10 E. coli cells (Invitrogen). Clones containing PCR inserts were sequenced to confirm correct DNA sequence. Four FAO alleles were identified from sequence analysis and designated as FAO-13, FAO-17, FAO-18 and FAO-20. The sequence of the clone designated FAO-18 had a sequence that was substantially identical to the sequence of FAO 1 from GenBank. The resulting plasmids of the four alleles were designated pAA083, pAA084, pAA059 and pAA085, respectively. Sequence identity comparisons of FAO genes isolated as described herein are shown in the tables below.

DNA sequence identity							
	FAOl	FAO-18	FAO-17	FAO-13	FAO-20	FA02a	FA02b
FAOl	100	100	98	96	95	83	82
FAO-18		100	98	96	95	83	82
FAO-17			100	98	98	83	82
FAO-13				100	99	83	83
FAO-					100	83	83

20							
FA02a						100	96
FA02b							100

Protein sequence identity							
	FAO1	FAO-18	FAO-17	FAO-13	FAO-20	FAO2a	FAO2b
FAO1	100	100	99	98	98	81	80
FAO-18		100	99	98	98	81	80
FAO-17			100	99	99	82	81
FAO-13				100	99	82	81
FAO-20					100	82	81
FAO2a						100	97
FAO2b							100

Amino acid differences in FAO alleles										
	32	75	89	179	185	213	226	352	544	590
FAOl	E	M	G	L	Y	T	R	H	S	P
FAO-13	Q	T	A	L	Y	A	K	Q	A	A
FAO-20	Q	T	A	M	D	A	K	Q	A	A

Expression of FAQ alleles in E. coli

To determine the levels of FAO enzyme activity with respect to various carbon sources, the four isolated FAO alleles were further cloned and over-expressed in E. coli. The FAOs were amplified using the plasmids mentioned above as DNA template by PCR with primers OAA0268 and oAA0269 for FAO-13 and FAO-20 and oAA0268 and oAA0282 for FAO- 17 and FAO-18, using conditions as described herein. PCR products of the correct size were gel purified and ligated into pETl 1a vector between Ndel and BamHI sites and transformed into BL21 (DE3) E. coli cells. The colonies containing corresponding FAOs were confirmed by DNA sequencing. Unmodified pETl 1a vector also was transformed into BL21 (DE3) cells, as a control. The resulting strains and plasmids were designated sAA153 (pETl la), sAA154 (pAA079 containing FAO-13), sAA155 (pAA080 containing FAO-17), sAA156 (pAA081 containing FAO-18) and sAA157 (pAA082 containing FAO-20), respectively. The strains and plasmids were used for FAO over-expression in E. coli. One colony of each strain was transferred into 5 mL of LB medium containing 100 Dg/mL ampicillin and grown overnight at 37°C, 200 rpm. The overnight culture was used to inoculate a new culture to OD_{600nm} 0.2 in 25 ml LB containing 100 Dg/ml ampicillin. Cells were induced at OD_{600nm} 0.8 with 0.3 mM IPTG for 3 hours and harvested by centrifugation at 4°C 1,050xg for 10 minutes. The cell pellet was stored at -20 °C.

Expression of FAOs in a Candida strain

Two alleles, FAO-13 and FAO-20, were chosen for amplification in Candida based on their substrate specificity profile, as determined from enzyme assays of soluble cell extracts of E. coli with over expressed FAOs. DNA fragments containing FAO-13 and FAO-20 were amplified using plasmids pAA079 and pAA082 as DNA templates, respectively, by PCR with primers oAA0421 and oAA0422. PCR products of the correct sizes were gel purified and ligated into pCR-Blunt II-TOPO (Invitrogen), transformed into competent TOP 10 E. coli cells (Invitrogen) and clones containing FAO inserts were sequenced to confirm correct DNA sequence. Plasmids containing FAO-13 and FAO-20 were digested with Sapl and ligated into vector pAA105, which includes the Candida strain ATCC20336 PGK promoter and terminator. The resulting plasmids were confirmed by restriction digestion and DNA sequencing and designated as pAAl 15 (FAO-13) and pAAl 16 (FAO-20), respectively. Plasmids pAAl 15 and pAAl 16 were linearized with Spel, transformed into competent Candida spp. Ura^- strains

SAA002 (SU-2, ATCC20913) and sAA103. The integration of FAO-13 and FAO-20 was confirmed by colony PCR using primers oAA0429 and oAA028 1. The resulting strains were designated as sAA278 (pAAl 15 integrated in strain sAA002), sAA280 (pAAl 16 integrated in sAA002), sAA282 (pAAl 15 integrated in sAA103), and sAA284 (pAAl 16 integrated in sAA103), and were used for fatty alcohol oxidase over-expression in Candida spp..

One colony of each strain was inoculated into 5 ml YPD and grown overnight as described herein. The overnight culture was used to inoculate a new 25 mL YPD culture to about OD_{600nm} 0.5. FAO over-expression was regulated by the PGK promoter/terminator, induced with glucose in the medium and expressed constitutively. Strains sAA002 and sAA103 (e.g., untransformed starting strains) were included as negative controls for FAO over-expression. Cells were harvested at early log phase (OD_{600nm} = in the range of between about 3 to about 5) by centrifugation at 4°C for 10 minutes at 1,050 x *g*. Cell pellets were stored at -20 °C.

Cell extract preparation from E. coli

Cell pellets from 25 mL of FAO expressing E. coli cultures were resuspended in 10 mL phosphate-glycerol buffer containing 50 mM potassium phosphate buffer (pH7.6), 20% glycerol, 1 mM Phenylmethylsulfonyl fluoride (PMSF), 2uL Benzonase 25U/uL, 20uL lOmg/mL lysozyme. The cells were then lysed by incubation at room temperature for 50 minutes on a rotating shaker, and the cell suspension centrifuged for 30 minutes at 4°C using 15,000 x *g* for. The supernatant was aliquoted in 1.5 ml microcentrifuge tubes and stored at -20 °C for FAO enzyme activity assays.

Cell extract preparation from Candida

Frozen C. tropicalis cell pellets were resuspended in 1.2 ml of phosphate-glycerol buffer containing 50 mM potassium phosphate buffer (pH7.6), 20% glycerol, 1 mM Phenylmethylsulfonyl fluoride (PMSF). Resuspended cells were transferred to 1.5mL screw-cap tubes containing about 500uL of zirconia beads on ice. The cells were lysed with a Bead Beater (Biospec) using 2 minute pulses and 1 minute rest intervals on ice. The process was repeated 3 times. The whole cell extract was then transferred to a new 1.5 ml tube and centrifuged at 16,000 x *g* for 15 minutes at 4°C. The supernatant was transferred into a new tube and used for FAO enzyme activity assays.

Protein concentration determination

Protein concentration of the cell extracts was determined using the Bradford Reagent following manufacturers' recommendations (Cat# 23238, Thermo scientific).

FAQ enzyme activity assay

FAO enzyme activity assays were performed using a modification of Eirich et al., 2004). The assay utilizes a two-enzyme coupled reaction (e.g., FAO and horse radish peroxidase (HRP)) and can be monitored by spectrophotometry. 1-Dodecanol was used as a standard substrate for fatty alcohol oxidase enzymatic activity assays. FAO oxidizes the dodecanol to dodecanal while reducing molecular oxygen to hydrogen peroxide simultaneously. HRP reduces (2,2'-azino-bis 3-ethylbenzthiazoline-6-sulfonic acid; ABTS) in the two-enzyme coupled reaction, where the electron obtained from oxidizing hydrogen peroxide to ABTS, which can be measured by spectrometry at 405 nm. The assay was modified using aminotriazole (AT) to prevent the destruction of H_20_2 by endogenous catalase, thus eliminating the need for microsomal fractionation. The final reaction mixture (1.0 mL) for FAO enzyme assay consisted of 500 µL of 200 mM HEPES buffer, pH 7.6; 50 µL of a 10 mg/mL ABTS solution in deionized water; 10 µL of 5 mM solution of dodecanol in acetone; 40 µL of 1M AT and 5 µL of a 2 mg/mL horseradish peroxidase solution in 50 mM potassium phosphate buffer, pH 7.6. Reaction activity was measured by measuring light absorbance at 405 nm for 10 minutes at room temperature after adding the extract. The amount of extract added to the reaction mixture was varied so that the activity fell within the range of 0.2 to 1.0 $AA405_{nm}$/min. The actual amounts of extract used were about 1.69 U/mg for E. coli expressed FAO-13, 0.018U/mg for E. coli expressed FAO-17, 0.35U/mg for E. coli expressed FAO-18 (e.g., FAOl), 0.47U/mg E. coli expressed FAO-20, 0.036U/mg Candida (strain sAA278) expressed FAO-13, 0.016U/mg Candida (strain sAA282) expressed FAO-13, 0.032U/mg Candida (strain sAA280) expressed FAO-20 and 0.029U/mg Candida (strain sAA284) expressed FAO-20. FAO activity was reported as activity units/mg of total protein (1 unit = 1 Dmole substrate oxidized/min). An extinction coefficient at 405 nm of 18.4 was used for ABTS and was equivalent to 0.5 mM oxidized substrate. The results of the activity assays are shown in the tables below.

FAO activity (units/mg total protein) on primary alcohols

FAO activity (units/mg total protein) on primary alcohols								
	1-Butanol	1-Pentanol	1-Hexanol	1-Octanol	1-Decanol	1-Dodecanol	1-Tetradecanol	Hexadecanol
FAO-13	0.01	0.09	1.17	82.67	70.94	100	79.35	58.88
FAO-17	0.72	0.26	1.06	66.23	22.00	100	47.86	60.98
FAO-18	0.07	0.1 1	0.26	60.56	54.56	100	114.47	50.65
FAO-20	0.07	0.1 1	0.91	55.96	74.57	100	89.52	42.59

FAO activity (units/mg total protein) on omega hydroxy fatty acids					
	1-Dodecanol	6-OH-HA	10-OH-DA	12-OH-DDA	16-OH-HDA
FAO-13	100	4.18	4.14	6.87	8.57
FAO-17	100	1.18	0.00	0.59	0.94
FAO-18	100	0.00	0.00	4.87	2.94
FAO-20	100	0.03	0.04	2.25	7.46

Example 13: Construction of Candida shuttle vector pAA061

Vector pAA061 was constructed from a pUC19 backbone to harbor the selectable marker URA3 from Candida strain ATCC20336 as well as modifications to allow insertion of Candida promoters and terminators. A 1,507bp DNA fragment containing the promoter, ORF, and terminator of URA3 from Candida ATCC20336 was amplified using primers oAA0124 and oAA0125, shown in TABLE 8 below. The URA3 PCR product was digested with NdeI / MluI and ligated into the 2,505bp fragment of pUC19 digested with NdeI / BsmBI (an MluI compatible overhang was produced by BsmBI). In order to replace the lac promoter with a

short 21bp linker sequence, the resulting plasmid was digested with SphI / SapI and filled in with a linker produced by annealing oligos oAA0173 and oAA0174. The resulting plasmid was designated pAA06 1.

TABLE 8

Oligonucleotides for construction of pAA061		
Oligos	Sequence	PCR product (bp)
oAA012 4	cacacacatatgCGACGGGTACAACGAGAATT	1507
oAA012 5	cacacaacgcgtAGACGAAGCCGTTCTTCAAG	
oAA017 3	ATGATCTGCCATGCCGAACTC	21 (linker)
oAA017 4	AGCGAGTTCGGCATGGCAGATCATCATG	

Example 14: Cloning of Candida PGK promoter and terminator

Vector pAA105 was constructed from base vector pAA061 to include the phosphoglycerate kinase (PGK) promoter and terminator regions from Candida ATCC20336 with an intervening multiple cloning site (MCS) for insertion of open reading frames (ORF's). The PGK promoter region was amplified by PCR using primers oAA0347 and oAA0348, shown in TABLE 9. The 1,029bp DNA fragment containing the PGK promoter was digested with restriction enzymes PstI / XmaI. The PGK terminator region was amplified by PCR using primers oAA035 1 and oAA0352, also shown in TABLE 9. The 396bp DNA fragment containing the PGK terminator was digested with restriction enzymes XmaI / EcoRI. The 3,728bp PstI / EcoRI DNA fragment from pAA061 was used in a three piece ligation reaction with the PGK promoter and terminator regions to produce pAA105. The sequence between the PGK promoter and terminator contains restriction sites for incorporating ORF's to be controlled by the functionally linked constitutive PGK promoter.

TABLE 9

Oligonucleotides for cloning *Candida* PGK promoter and terminator		
Oligos	Sequence	PCR product (bp)
oAA0347	CACACACTGCAGTTGTCCAATGTAATAATTTT	1028
oAA0348	CACACATCTAGACCCGGGCTCTTCTTCTGAATAGGCAATTGATAAACTTACTTATC	
oAA0351	GAGCCCGGGTCTAGATGTGTGCTCTTCCAAAGTACGGTGTTGTTGACA	396
oAA0352	CACACACATATGAATTCTGTACTGGTAGAGCTAAATT	

Example 15: Cloning of the POX4 locus

Primers oAA0138 and oAA0141 (TABLE 10) were generated to amplify the entire sequence of NCBI accession number M12160 for the YSAPOX4 locus from genomic DNA prepared from Candida strain ATCC20336. The 2,845bp PCR product was cloned into the vector, pCR-Bluntll-TOPO (Invitrogen), sequenced and designated pAA052.

TABLE 10

Oligonucleotides for cloning of *POX4*		
Oligos	Sequence	PCR product (bp)
oAA0138	GAGCTCCAATTGTAATATTTCGGG	2845
oAA0141	GTCGACCTAAATTCGCAACTATCAA	

Example 16: Cloning of the POX5 locus

Primers oAA0179 and oAA0182 (TABLE 11) were generated to amplify the entire sequence of NCBI accession number M12161 for the YSAPOX5 locus from genomic DNA prepared from

Candida strain ATCC20336. The 2,624bp PCR product was cloned into the vector, pCR-Bluntll-TOPO (Invitrogen), sequenced and designated pAA049.

TABLE 11

Oligonucleotides for cloning of *POX5*		
Oligos	Sequence	PCR product (bp)
oAA0179	GAATTCACATGGCTAATTTGGCCTCGGTTCCACAACGCACTCAGCATTAAAAA	2624
oAA0182	GAGCTCCCCTGCAAACAGGGAAACACTTGTCATCTGATTT	

Example 17: Construction of strains with amplified CPR and CYP52 genes

Strains having an increased number of copies of cytochrome P450 reductase (CPR) and/or for cytochrome P450 monooxygenase (CYP52) genes were constructed to determine how over expression of CPR and CYP52 affected diacid production.

Cloning and integration of the CPR gene

A 3,019bp DNA fragment encoding the CPR promoter, ORF, and terminator from Candida ATCC750 was amplified by PCR using primers oAA0171 and oAA0172 (TABLE 12) incorporating unique SapI and SphI sites. The amplified DNA fragment was cut with the indicated restriction enzymes and ligated into plasmid pAA061, (described in Example 13) to produce plasmid pAA067. Plasmid pAA067 was linearized with ClaI and transformed into Candida Ura$^-$ strain sAA103 (ura3/ura3, pox4::ura3/pox4::ura3, pox5::ura3/pox5::ura3). Transformations were performed with plasmid pAA067 alone and in combination with plasmids harboring the CYP52A15 or CYP52A16 genes, described below.

Cloning and integration of CYP52A15 gene

A 2,842bp DNA fragment encoding the CYP52A15 promoter, ORF, and terminator from Candida ATCC20336 was amplified by PCR using primers oAA0175 and oAA0178 (see Table 12 below) and cloned into pCR-Bluntll-TOPO for DNA sequence verification. The cloned CYP52A15 DNA fragment was isolated by restriction digest with Xbal / BamHI (2,742bp) and ligated into plasmid pAA061, (described in Example 13), to produce plasmid pAA077. Plasmid pAA077 was linearized with Pmll and transformed into Candida Ura$^-$ strain sAA103 (ura3/ura3, pox4: :ura3/pox4: :ura3, pox5 ::ura3/pox5 ::ura3). pAA077 was cotransformed with plasmid pAA067 harboring the CPR gene.

Cloning and integration of CYP52A16 gene

A 2,728bp DNA fragment encoding the CYP52A16 promoter, ORF, and terminator from Candida ATCC20336 was amplified by PCR using primers oAA0177 and oAA0178 (TABLE 12) and cloned into pCR-Bluntll-TOPO for DNA sequence verification. The cloned CYP52A16 DNA fragment was amplified with primers oAA0260 and oAA026 1(TABLE 12) which incorporated unique Sacl / Xbal restriction sites. The amplified DNA fragment was digested with Sacl and Xbal restriction enzymes and ligated into plasmid pAA061 to produce plasmid pAA078. Plasmid pAA078 was linearized with Clal and transformed into Candida Ura$^-$ strain sAA103 (ura3/ura3, pox4::ura3/pox4::ura3, pox5::ura3/pox5::ura3). pAA078 was cotransformed with plasmid pAA067 harboring the CPR gene.

TABLE 12

Oligonucleotides for cloning of CPR, CYP52A15 and CYP52A16		
Oligos	Sequence	PCR product (bp)
oAA017 1	cacctcgctcttccAGCTGTCATGTCTATTCAATGCT TCGA	3019
oAA017 2	cacacagcatgcTAATGTTTATATCGTTGACGGTGA AA	
oAA017 5	cacaaagcggaagagcAAATTTTGTATTCTCAGTAG GATTTCATC	2842
oAA017	cacacagcatgCAAACTTAAGGGTGTTGTAGATAT	

8	CCC	
oAA0177	cacacacccgggATCGACAGTCGATTACGTAATCCATATTATTT	2772
oAA0178	cacacagcatgCAAACTTAAGGGTGTTGTAGATATCCC	
oAA0260	cacacagagctcACAGTCGATTACGTAATCCAT	2772
oAA0261	cacatctagaGCATGCAAACTTAAGGGTGTTGTA	

Preparation of genomic DNA

Genomic DNA was prepared from transformants for PCR verification and for Southern blot analysis. Isolated colonies were inoculated into 3 mL YPD and grown overnight at 30 °C with shaking. Cells were pelleted by centrifugation. To each pellet, 200 uL Breaking Buffer (2% Triton X-100, 1% SDS, 100 mM NaCl, 10 mM Tris pH 8 and, 1 mM EDTA) was added, and the pellet resuspended and transferred to a fresh tube containing 200 uL 0.5 mm Zirconia/Silica Beads. 200 uL Phenol:Chloroform:Isoamyl Alcohol (25:24:1) was added to each tube, followed by vortexing for 1 minute. Sterile distilled water was added (200 uL) to each tube and the tubes were centrifuged at 13000 rpm for 10 minutes. The aqueous layer was ethanol precipitated and washed with 70% ethanol. The pellet was resuspended in 100-200 □1 10 mM Tris, after drying. Genomic DNA preparation for southern blot analysis was performed using the same procedure on 25 mL cultures for each colony tested.

Characterization of strains with amplified CPR and CYP52 genes

Verification of integrated genes and/or sequences was performed by PCR using primers OAA0252 and oAA0256 (CPR), oAA0231 and oAA0281 (CYP52A15), and oAA242 and 0AAO257 (CYP52A16). The primers used for verification are shown in TABLE 13.

TABLE 13

Oligonucleotides for PCR verification of *CPR, CYP52A15 and CYP52A16*		
Oligos	Sequence	PCR product (bp)
0AAO252	TTAATGCCTTCTCAAGACAA	743
0AAO256	GGTTTTCCCAGTCACGACGT	
0AAO231	CCTTGCTAATTTTCTTCTGTATAGC	584
oAA0281	TTCTCGTTGTACCCGTCGCA	
0AAO242	CACACAACTTCAGAGTTGCC	974
0AAO257	TCGCCACCTCTGACTTGAGC	

Southern blot analysis was used to determine the copy number of the CPR, CYP52A15 and CYP52A16 genes. Biotinylated DNA probes were prepared with gene specific oligonucleotides using the NEBlot Phototope Kit from New England BioLabs (Catalog #N7550S) on PCR products generated from each gene target as specified in TABLE 14 below. Southern Hybridizations were performed using standard methods (e.g., Sambrook, J. and Russell, D. W. (2001) Molecular Cloning: A Laboratory Manual, (3rd ed.), pp. 6.33-6.64. Cold Spring Harbor Laboratory Press). Detection of hybridized probe was performed using the Phototope-Star Detection Kit from New England BioLabs (Catalog #N7020S). Copy number was determined by densitometry of the resulting bands.

TABLE 14

Oligonucleotides for Probe Template PCR of *CPR, CYP52A15 and CYP52A16*				
Oligos	Sequence	Gene	Template	PCR product (bp)
OAA0250	AATTGAACATCAGAAGAGGA	CPR	pAA067	1313
0AAO254	CCTGAAATTTCCAAATGGTGT CTAA			
0AAO227	TTTTTTGTGCGCAAGTACAC	CYP52A15	pAA077	905
0AAO235	CAACTTGACGTGAGAAACCT			
0AAO239	AGATGCTCGTTTTACACCCT	CYP52A16	pAA078	672
0AAO247	ACACAGCTTTGATGTTCTCT			

Example 18: Addition and/or Amplification of Monooxygenase and Monooxygenase reductase activities.

Cytochrome P450's often catalyze a monooxygenase reaction, e.g., insertion of one atom of oxygen into an organic substrate (RH) while the other oxygen atom is reduced to water:

RH + 0 2 + 2H+ + 2e- → ROH + H20

The substrates sometimes are of a homogeneous carbon chain length. Enzymes with monooxygenase activity sometimes recognize substrates of specific carbon chain lengths, or a subgroup of carbon chain lengths with respect to organic substrates of homogenous carbon chain length. Addition of novel cytochrome activities (e.g., B. megaterium BM3) and/or amplification of certain or all endogenous or heterologous monooxygenase activities (e.g., CYP52A12 polynucleotide, CYP52A13 polynucleotide, CYP52A14 polynucleotide, CYP52A15 polynucleotide, CYP52A16 polynucleotide, CYP52A17 polynucleotide, CYP52A18 polynucleotide, CYP52A19 polynucleotide, CYP52A20 polynucleotide, CYP52D2 polynucleotide, BM3 polynucleotide) can contribute to an overall increase in carbon flux through native and/or engineered metabolic pathways, in some embodiments. In certain embodiments, adding a novel monooxygenase or increasing certain or all endogenous or heterologous monooxygenase activities can increase the flux of substrates of specific carbon chain length or subgroups of substrates with mixtures of specific carbon chain lengths. In some

embodiments, the selection of a monooxygenase activity for amplification in an engineered strain is related to the feedstock utilized for growth of the engineered strain, pathways for metabolism of the chosen feedstock and the desire end product (e.g., dodecanedioic acid). Strains engineered to utilize plant-based oils for conversion to dodecanedioic acid can benefit by having one or more monooxygenase activities with substrate specificity that matches the fatty acid chain-length distribution of the oil. For example, the most prevalent fatty acid in coconut oil is lauric acid (12 carbons long), therefore, the monooxygenase activity chosen for a coconut oil-utilizing strain can have a substrate preference for C12 fatty acids. For strains engineered to utilize other plant based oils with different fatty acid chain-length distributions it may be desirable to amplify a monooxygenase activity that has a matching substrate preference. In some embodiments, a genetic modification that alters monooxygenase activity increases the activity of one or more monooxygenase activities with a substrate preference for feedstocks having carbon chain lengths of between about 12 and about 24 carbons (e.g., mixed chain length alkanes, mixed chain length fatty acids, soapstocks, the like and combinations thereof). In certain embodiments, the genetic modification increases the activity of a monooxygenase activity with a preference for fatty acids having a carbon chain-length distribution of between about 10 carbons and about 16 carbons.

As mentioned previously, the enzymes that carry out the monooxygenase activity are reduced by the activity of monooxygenase reductase, thereby regenerating the enzyme. Selection of a CPR for amplification in an engineered strain depends upon which P450 is amplified, in some embodiments. A particular CPR may interact preferentially with one or more monooxygenase activities, in some embodiments, but not well with other monooxygenases. A monooxygenase reductase from Candida strain ATCC750, two monooxygenase reductase activities from Candida strain ATCC20336 and a monooxygenase reductase activity from Bacillus megaterium are being evaluated for activity with the added and/or amplified monooxygenases described herein. Provided in the tables below are nucleotide sequences used to add or amplify monooxygenase and monooxygenase reductase activities.

Example 19: Amplification of Selected Beta Oxidation Activities

Described herein are methods of amplifying a POX5 beta oxidation activity. Substantially similar methods can be utilized to amplify different beta oxidation activities including non-POX (e.g., acyl-CoA oxidase) activities and/or acyl-CoA oxidase activities with altered substrate specificities, as described herein.

Construction of POX5 amplified strains

Plasmid pAA166 (P$_{POX4}$POX 5T$_{POX4}$)

A PCR product containing the nucleotide sequence of POX5 was amplified from Candida 20336 genomic DNA using primers oAA540 and oAA541 . The PCR product was gel purified and ligated into pCR-Blunt II-TOPO (Invitrogen), transformed into competent TOP 10 E. coli cells (Invitrogen) and clones containing PCR inserts were sequenced to confirm correct DNA sequence. One such plasmid was designated, pAA165. Plasmid pAA165 was digested with BspQI and a 2-kb fragment was isolated. Plasmid pAA073 which contained a POX4 promoter and POX4 terminator was also digested with BspQI and gel purified. The isolated fragments were ligated together to generate plasmid pAA166. Plasmid pAA166 contains a Ppox4POX5Tp$_{0x4}$ fragment.

Plasmid pAA204 (thiolase deletion construct)

A PCR product containing the nucleotide sequence of a short-chain thiolase (e.g., acetyl-CoA acetyltransferase) was amplified from Candida 20336 genomic DNA using primers oAA640 and oAA641 . The PCR product was gel purified and ligated into pCR-Blunt II-TOPO (Invitrogen), transformed into competent TOP 10 E. coli cells (Invitrogen) and clones containing PCR inserts were sequenced to confirm correct DNA sequence. One such plasmid was designated, pAA184. A URA3 PCR product was amplified from pAA061 using primers 0AA66O and 0AA66 I . The PCR product was gel purified and ligated into pCR-Blunt II-TOPO (Invitrogen), transformed as described and clones containing PCR inserts were sequenced to confirm the correct DNA sequence. One such plasmid was designated pAA192. Plasmid pAA184 was digested with Bglll/Sall and gel purified. Plasmid pAA192 was digested with Bglll/Sall and a 1.5 kb fragment was gel purified. The isolate fragments were ligated together to generate pAA199. An alternative PU$_R$A$_3$ PCR product was amplified from plasmid pAA061 using primers oAA684 and oAA685. The PCR product was gel purified and ligated into pCR-

Blunt II-TOPO (Invitrogen), transformed as described and clones containing PCR inserts were sequenced. One such plasmid was designated, pAA201 . Plasmid pAA199 was digested with Sail and gel purified. Plasmid pAA201 was digested with Sail and a 0.43 kb PU_RA_3 was gel purified. The isolated fragments were ligated to generate plasmid pAA204 that contains a direct repeat of PURA3.

Plasmid pAA221 ($P_Pox_4POX5T_Pox_4$ in thiolase deletion construct)

A PCR product containing the nucleotide sequence of $P\rho o\chi_4POX\ 5T_P o\chi 4$ was amplified from plasmid pAA166 DNA using primers oAA728 and oAA729. The PCR product was gel purified and ligated into pCR-Blunt II-TOPO, transformed as described and clones containing PCR inserts were sequenced to confirm the sequence of the insert. One such plasmid was designated, pAA220. Plasmid pAA204 was digested with Bglll, treated with shrimp alkaline phosphatase (SAP), and a 6.5 kb fragment was gel purified. Plasmid pAA220 was digested with Bglll and a 2.7 kb fragment containing $P\rho o\chi_4POX\ 5T_P o\chi_4$ was gel purified. The isolated fragments were ligated to generate plasmid pAA221.

Strain sAA617 ($P_Pox4POX5T_Pox4$ in sAA451)

Strain sAA45 1 is a ura-, partially β-oxidation blocked strain (ura3/ura3 pox4a::ura3/pox4b::ura3 POX5/POX5). Plasmid pAA221 was digested with EcoRI to release a DNA fragment containing $P\rho o\chi_4POX\ 5T\rho o\chi_4$ in a thiolase deletion construct. The DNA was column purified and transformed to strain sAA45 1 to plate on SCD-ura plate. After two days, colonies were streaked out on YPD plates, single colonies selected and again streaked out on YPD plates. Single colonies were selected from the second YPD plates and characterized by colony PCR. The insertion of $P\rho o\chi_4POX\ 5T\rho o\chi_4$ in strain sAA451, disrupting the short-chain thiolase gene, was confirmed by PCR and one such strain was designated sAA617.

Strain sAA620

Strain sAA617 was grown overnight on YPD medium and plated on SCD+URA+5-FOA, to select for loop-out of URA3. Colonies were streaked out onto YPD plates twice as described for strain sAA617, and single colonies characterized by colony PCR. The loop-out of URA3 by

direct repeats of PURA3 was confirmed by PCR. One such strain was designated sAA620. Strain sAA620 has one additional copy of POX5 under control of the POX4 promoter.

Plasmid pAA156

A PCR product containing the nucleotide sequence of CYP52A19 was amplified from Candida strain 20336 genomic DNA, using primers oAA525 and oAA526. The PCR product was gel purified and ligated into pCR-Blunt II-TOPO, transformed as described, and clones containing PCR inserts were sequenced to confirm correct DNA sequence. One such plasmid was designated, pAA144. Plasmid pAA144 was digested with BspQI and a 1.7-kb fragment was isolated. Plasmid pAA073, which includes a POX4 promoter and POX4 terminator, also was digested with BspQI and gel purified. The isolated fragments were ligated together to generate plasmid, pAA156. Plasmid pAA156 included P_POX4CYP 52A19$T_{POX}4$ fragment and URA3.

Strain sAA496

Plasmid pAA156 was digested with ClaI and column purified. Strain sAA45 1 was transformed with this linearized DNA and plated on SCD-ura plate. Colonies were checked for CYP52A19 integration. Colonies positive for plasmid integration were further analyzed by qPCR to determine the number of copies of CYP52A19 integrated. One such strain, designated sAA496 contained about 13 copies of the monooxygenase activity encoded by CYP52A19.

Strains sAA632 and sAA635

Strain sAA620 was transformed with linearized pAA156 DNA and plated on SCD-ura plates. Several colonies were checked for CYP52A19 integration. Colonies positive for plasmid integration were further analyzed by qPCR to determine the number of copies of CYP52A19 integrated. One such strain, designated sAA632 contained about 27 copies of the monooxygenase activity encoded by CYP52A19. Another strain, designated sAA635, contained about 12 copies of the monooxygenase activity encoded by CYP52A19.

Example 20: Cloning of Candida ACH genes

ACH PCR product was amplified from *Candida strain ATCC20336* genomic DNA using primers oAA1095 and oAA1096, shown in TABLE 15. The PCR product was gel purified and ligated into pCR-Blunt II-TOPO (Invitrogen), transformed into competent TOP10 *E. coli* cells (Invitrogen) and clones containing PCR inserts were sequenced to confirm correct DNA sequence.

Sequence analysis of multiple transformants revealed the presence of allelic sequences for the ACH gene, which were designated ACHA and ACHB. A vector containing the DNA sequence for the ACHA allele was generated and designated pAA310. A vector containing the DNA sequence for the ACHB allele was generated and designated pAA311.

TABLE 15

Primer	sequence
oAA1095	CACACACCCGGGATGATCAGAACCGTCCGTTATCAAT
oAA1096	CACACATCTAGACTCTCTTCTATTCTTAATTGCCGCTTCCACTAAA CGGCAAAGTCTCCACG

Example 21: Cloning of Candida FAT1 gene

FAT1 PCR product was amplified from *Candida* 20336 genomic DNA using primers oAA1023 and oAA1024, shown in TABLE 16 below. The PCR product was gel purified and ligated into pCR-Blunt II-TOPO (Invitrogen), transformed into competent TOP10 *E. coli* cells (Invitrogen) and clones containing PCR inserts were sequenced to confirm correct DNA sequence. A vector containing the DNA sequence for the FAT1 gene was designated pAA296.

TABLE 16

Primer	sequence
oAA1023	GATATTATTCCACCTTCCCTTCATT
oAA1024	CCGTTAAACAAAAATCAGTCTGTAAA

Example 22: Cloning of Candida ARE1 and ARE2 genes

AREl and *AREl* PCR products were amplified from *Candida* 20336 genomic DNA using primers oAA2006/oAA2007 and oAA1012/oAA1018, respectively, shown in TABLE 17 below. The PCR products were gel purified and ligated into pCR-Blunt II-TOPO (Invitrogen), transformed into competent TOP10 *E. coli* cells (Invitrogen) and clones containing PCR inserts were sequenced to confirm correct DNA sequence. A vector containing the DNA sequence for the AREl gene was designated pAA318. A vector containing the DNA sequence for the ARE2 gene was designated pAA301.

TABLE 17

Primer	sequence
oAA1012	ATGTCCGACGACGAGATAGCAGGAATAGTCAT
oAA1018	TCAGAAGAGTAAATACAACGCACTAACCAAGCT
oAA2006	ATGCTGAAGAGAAAGAGACAACTCGACAAG
oAA2007	GTGGTTATCGGACTCTACATAATGTCAACG

Example 23: Construction of an optimized TESA gene for expression in Candida

The gene sequence for the *E. coli TESA* gene was optimized for expression in *Candida* by codon replacement. A new *TESA* gene sequence was constructed using codons from *Candida* with similar usage frequency for each of the codons in the native *E. coli TESA* gene (avoiding the use of the CTG codon due to the alternative yeast nuclear genetic code utilized by *Candida).* The optimized *TESA* gene was synthesized with flanking BspQI restriction sites and provided in vector pIDTSMART-Kan (Integrated DNA Technologies). The vector was designated as pAA287. Plasmid pAA287 was cut with BspQI and the 555bp DNA fragment was gel purified. Plasmid pAA073 also was cut with BspQI and the linear DNA fragment was gel purified. The two DNA fragments were ligated together to place the optimized *TESA* gene under the control of the *Candida POX4* promoter. The resulting plasmid was designated pAA294.

Example 24: Cloning of Candida DGA1 gene

DGA1 PCR product was amplified from *Candida* 20336 genomic DNA using primers oAA996 and oAA997, shown in TABLE 18 below. The PCR product was gel purified and ligated into pCR-Blunt II-TOPO (Invitrogen), transformed into competent TOP10 *E. coli* cells (Invitrogen) and clones containing PCR inserts were sequenced to confirm correct DNA sequence. A vector containing the DNA sequence of the DGA1 gene was designated pAA299.

TABLE 18

Primer	Sequence
oAA996	ATGACTCAGGACTATAAAGACGATAGTCCTACGTCCACTGAGTTG
oAA997	CTATTCTACAATGTTTAATTCAACATCACCGTAGCCAAACCT

Example 25: Cloning of Candida LROl gene

LROl PCR product was amplified from *Candida* 20336 genomic DNA using primers oAA998 and oAA999, shown in TABLE 19 below. The PCR product was gel purified and ligated into pCR-Blunt II-TOPO (Invitrogen), transformed into competent TOP10 *E. coli* cells (Invitrogen) and clones containing PCR inserts were sequenced to confirm correct DNA sequence. A vector containing the DNA sequence of the LROl gene was designated pAA300.

TABLE 19

Primer	sequence
oAA998	ATGTCGTCTTTAAAGAACAGAAAATC
oAA999	TTATAAATTTATGGCCTCTACTATTTCT

Example 26: Cloning of Candida ACS1 gene and construction of deletion cassette

ACS1 PCR product was amplified from *Candida* 20336 genomic DNA using primers oAA951 and oAA952, shown in TABLE 20 below. The PCR product was gel purified and ligated into pCR-Blunt II-TOPO (Invitrogen), transformed into competent TOP10 *E. coli* cells (Invitrogen) and clones containing PCR inserts were sequenced to confirm the DNA sequence. One such plasmid was designated pAA275. Plasmid pAA280 was digested with BamHI to release a 2.0

kb P_{URA3}*URA3*$T_{URA3}P_{URA3}$ cassette. Plasmid pAA275 was digested with BglII and gel purified. The two pieces were ligated together to generate plasmid pAA276 and pAA282. Plasmid pAA276 and pAA282 have the PU_RA_3*URA3*$TU_{RA3}PU_RA_3$ cassette inserted into the ACS gene in opposite orientations.

TABLE 20

Primer	sequence
oAA951	CCTACTTCCACAGCTTTAATCTACTATCAT
oAA952	TTTAAGAAAACAACTAAGAGAAGCCAC

Example 27: Construction of Strain sAA722 (pox4a::ura3/pox4b::ura3 POX5/POX5 ACSl/acsl:: $PURA3URA3T_{URA3}PURA3)$

Plasmid pAA276 was digested with BamHI/XhoI and column purified. Strain sAA329 *(ura3/ura3 pox4a::ura3/pox4b::ura3 POX5/POX5)* was transformed with the linearized DNA and plated on SCD-ura plate. Several colonies were checked *for ACS1* disruption. One such strain was designated sAA722.

Example 28: Construction of Strain sAA 741 (pox4a::ura3/pox4b::ura3 POX5/POX5 ACSl/acs 1::$PU_RA3)$

Strain sAA722 was grown in YPD media overnight and plated on 5-FOA plate. Colonies that grew in the presence of 5-FOA were PCR screened for the looping out of the *URA3* gene leaving behind only the *URA3* promoter (PU_RA_3) in the *ACS1* site. Out of 30 colonies analyzed, only one strain showed the correct genetic modification. The strain was designated sAA741 .

Example 29: Construction of Strain sAA 776 (pox4a::ura3/pox4b::ura3 POX5/POX5 acsl::$P_{URA3}URA3T_{URA3}P_{URA3}$*/acs 1::*$P_{URA3})$

Plasmid pAA282 was digested with BamHI/XhoI and column purified. Strain sAA741 (see Example 28) was transformed with the linearized DNA and plated on SCD-ura plate. Several colonies were checked for double *ACS1* knockout by insertional inactivation. One such strain was designated sAA776.

Example 30: Construction of Strain sAA 779 (pox4a::ura3/pox4b::ura3 POX5/POX5 acsl:: P_{URA3}/acsl:: PJJ_RA_3)

Strain sAA776 (see Example 29) was grown in YPD media overnight and plated on 5-FOA plates. Colonies that grew in the presence of 5-FOA were PCR screened for the looping out of the *URA3* gene leaving behind only the *URA3* promoter (PU_RA_3) in both *ACS1* copies. One such strain was designated sAA779.

Example 31: Construction of Strain sAA811 (pox4a::ura3/pox4b::ura3 POX5/POX5 acsl:: P_{JRA3}/acsl:: $P\,U_RA_3$ ura3::3xPpox$_4$P450A19)

Plasmid pAA156 containing a *P450A19* integration cassette was digested with ClaI and column purified. Strain sAA779 (see Example 30) was transformed with the linearized DNA and plated on SCD-ura plate. Several colonies were checked for *P450A19* integration. From those colonies, qPCR was performed to check the copy number of *P450A19* integration. One strain, designated sAA81 1, contained 3 copies of *P450A19*.

Example 32: Construction of Strain sAA810 (pox4a::ura3/pox4b::ura3 POX5/POX5 acsl:: P_{JRA3}/acsl:: $P\,U_RA_3$ ura3::5xPpox$_4$P450A19 ura3::8xPpox$_4$TESA)

Plasmid pAA156 containing a P450-A19 integration cassette was digested with ClaI and column purified. Plasmid pAA294 containing a *TESA* integration cassette also was digested with ClaI and column purified. Strain sAA779 was cotransformed with both linearized DNAs and plated on SCD-ura plate. Several colonies were checked for both *P450A19* integration and *TESA* integration. Colonies that were positive for both TESA and P450A19 were further analyzed by qPCR. qPCR was performed to check the copy number of the *P450A19* and *TESA* integration events. One strain, designated sAA810, contained 5 copies of *P450A19* and 8 copies of *TESA*.

Example 33. General Techniques & Methods.

Growth Media, Reagents and Conditions

YPD, ScD-ura media and plates, and 5-FOA containing plates were made as described in Methods in Yeast Genetics: a Cold Spring Harbor Laboratory Manual /David C. Amberg, Daniel J. Burke, Jeffrey Strathern, - 2005 ed.).

SP92 + glycerol was made by adding 6.7 g of Bacto yeast nitrogen base without amino acids (BD, Franklin Lakes, NJ, USA), 3.0 g of Bacto yeast extract (BD, Franklin Lakes, NJ, USA), 3.0 g of ammonium sulfate, 1.0 g of potassium phosphate monobasic, 1.0 g of potassium phosphate dibasic, and 75 g of glycerol to water to a final volume of one liter. The media was then filtered sterilized.

TB-low N Media was made by adding 1.7 g Bacto yeast nitrogen base without ammonium sulfate, 3 g of Bacto yeast extract, 1 g of potassium phosphate monobasic and 1g potassium phosphate dibasic per liter of water. The media was filtered sterilized.

Overnight cultures were typically grown in 2 to 5 ml of either ScD-ura media or YPD media in standard culture tubes at 30 C on a shaker at about 250 rpm.

Molecular Methods

Gel purifications of DNA fragments were done as recommended by the manufacturer using either the GeneJET Gel Extraction Kit (Fermentas Inc., Glen Burnie, MD USA) or the Zymoclean Gel DNA Recovery Kit (ZymoResearch, Irvine, CA).

PCR was performed using either PFU Ultra II DNA Polymerase (Agilent Technologies, Santa Clara, CA), Taq DNA polymerase (New England Biolabs, Ipswich, MA, USA), DreamTaq PCR Master Mix (Fermentas Inc, Glen Burnie, MD USA) or Quick Load Midas Mix (Monserate, San Diego, CA USA). Each enzyme was used according to the manufacturer's instructions.

Restriction enzyme digestions were conducted as recommended by each manufacturer (New England Biolabs, Ipswich, MA, USA or Fermentas Inc, Glen Burnie, Maryland USA). DNA ligations were conducted using either the Rapid Ligation Kit (Fermentas Inc, Glen Burnie, Maryland USA) or using T4 DNA Ligase (New England Biolabs, Ipswich, MA, USA) according to the manufacturer's instructions.

Yeast transformations were performed as described in Example 10.

Genomic DNA Preparation

The URA3 gene was obtained from genomic DNA of Candida yeast culture ATCC20336.

Genomic DNA from Candida strain ATCC20336 was prepared as follows: About 1.5 ml of an overnight culture of cells was and the pellet was resuspended in about 200 µï of a solution containing 2% Triton X-100, 1% SDS, 100 mM NaCl, 10 MM Tris pH 8.0, and 1 mM EDTA. About 200 µï of acid washed glass beads were added with about 200 µï of phenol:chloroform:isoamyl alcohol (25:24:1) at a pH of about 8.0. The sample was vortexed for about 2 minutes after which about 200 µï of water was added. The sample was then centrifuged at 13000 rpm for about 10 minutes. The aqueous layer was transferred to a new microcentrifuge tube and an equal volume of chloroform:isoamyl alcohol (24:1) solution was added. This sample was vortexed for 10 seconds and then centrifuged at 13000 rpm for about 2 minutes. The aqueous layer was transferred to a new microfuge tube and 1 ml of ethanol was added. The tube was then placed at -80 °C for about 15 minutes and then spun at 13000 rpm for 15 minutes to pellet the DNA The DNA was washed with 70% ethanol and air-dried. The DNA was then resuspended in about 500 µï of water.

Genomic DNA for *Klyveromyces lactis* (ATCC8585) was purchased from the American Type Culture Collection (Manassas, VA, USA).

To calculate gene copy number, a qPCR method was used as described by Jin et al (Appl. Environ. Microbiol. January 2003 vol. 69, no. 1, 495-503). qPCR was performed according to the manufacturer's instructions using either the Brilliant III Ultra-Fast SYBR® Green QPCR Master Mix (Agilent Technologies, Englewood, CO USA) or the QuantiTect Multiplex PCR NoROX Kit (Qiagen). Genomic DNA from Candida strain ATCC20336 or plasmid DNA containing the actin gene from ATCC20336 and the gene of interest were used as standards.

Primers and probes used throughout these Examples were made via standard DNA synthesis techniques by Integrated DNA Technologies (Coralville, IA, USA).

Example 34: Construction of Cloning Plasmid AA073

The plasmid pAA073 was designed to contain the POX4 promoter and terminator from *Candida* strain ATCC20336 (this strain is also referred to herein as strain sAAOOl). This plasmid was derived from the publicly available plasmid pUC19 which contains an ampicillin resistance marker. pAA073 was designed to have two SapI restriction enzyme sites located between the POX4 promoter and terminator which allows unidirectional cloning of any gene of interest in tandem with the *POX4* promoter. The *Candida* strain ATCC20336 *URA3* gene including the open reading frame and the endogenous regulatory regions was also placed into pAA073 as a selection marker for transformants. Plasmid pAA073 allows the direct integration of multiple copies of any gene of interest by digesting the plasmid with a unique restriction enzyme such as SpeI, ClaI or BstZ171. These multiple cloning sites for are contained in the URA3 auxotrophic marker region and can be selectively be used to avoid cutting the gene of interest (i.e., the DNA sequence for the gene of interest can be searched for particular restriction enzyme cut sites and those enzymes can be avoided). In addition, this plasmid can serve as a template to generate an antibiotic free-DNA cassette containing the gene of interest and the POX 4 regulatory regions inserted between the 3' and 5' regions of the URA3 gene; this cassette can be PCR amplified using the plasmid as a template, and the isolated PCR product can be inserted into any microorganism strain.
A diagram of pAA073 is set forth in **FIG. 18** and the sequence of pAA073 is set forth as SEQ ID NO. 3704.

Example 35. *Cloning enoyl-CoA isomerase (ECI) genes from ATCC 20336*

The amino acid sequence for EciI (i.e. Eci) from S. cerevisiae S288c (SEQ ID NO. 3705) was used to identify homologs from Candida species ATCC MYA-3404 and ATCC20336. The BLAST search revealed two Ecilp homologs in each strain of Candida, which have been named Ecilp and Eci2p (TABLE 21). The percent amino acid identities for the homologs are shown below:

TABLE 21

Amino acid percent identity

	Eci2p_MYA-3404 SEQ ID NO. 3707	**Eci1p_MYA-3404** SEQ ID NO. 3706	**Eci1p_S.c.** SEQ ID NO. 3705	**Eci2p_20336** SEQ ID NO. 3709	**Eci1p_20336** SEQ ID NO. 3708
Eci2p_MYA-3404 SEQ ID NO. 3707		58	36	84	57
Eci1p_MYA-3404 SEQ ID NO. 3706			39	57	92
Eci1p_S.c. SEQ ID NO. 3705				37	40
Eci2p_20336 SEQ ID NO. 3709					57
Eci1p_20336 SEQ ID NO. 3708					

The ECU gene encoding the N-terminal 241 residues of SEQ ID NO. 3708 was amplified from genomic DNA (ATCC 20336) using oligonucleotides oAA2835 (SEQ ID NO. 3712) and oAA2836 (SEQ ID NO. 3713) that also incorporated unique SapI restriction sites. The 770 bp PCR product was gel purified and ligated into pCR-Blunt II-TOPO (Life Technologies), transformed into competent TOP10 E. coli cells (Life Technologies) and clones containing PCR inserts were sequenced to confirm the correct DNA sequence. One such plasmid was named pAA574 (SEQ ID NO. 3710).

The full length ECI2 gene encoding Eci2p (SEQ ID NO. 3709) was amplified from genomic DNA (ATCC 20336) using oligonucleotides oAA2837 (SEQ ID NO. 3714) and oAA2838 (SEQ ID NO. 3715) that also incorporated unique SapI restriction sites. The 851 bp PCR product was gel purified and ligated into pCR-Blunt II-TOPO (Life Technologies), transformed into competent TOP 10 E. coli cells (Life Technologies) and clones containing PCR inserts were sequenced to confirm the correct DNA sequence. One such plasmid was named pAA575 (SEQ ID NO. 371 1).

Example 36 - Generation of Strain sAA1764 (ura3/ura3 pox4a::ura3/pox4b::ura3 POX5/POX5 acsl:: PURA3/acsl:: PURA3 fatl-Al ::PURA3/fatl-A2::PURA3 ecil-A1::URA3/ECI1)

Deletion of the first allele of ECU was achieved by transforming cells (strain sAA886 (pox4a::ura3/pox4b::ura3 POX5/POX5 acsl:: PURA3/acsl:: PURA3 fatl-Al::PURA3/fatl-A2::PURA3 ura3/ura3)) with linear DNA cassettes constructed by overlap extension PCR (OE-PCR). A deletion cassette for the first ECU allele in strain sAA886 was generated from three DNA fragments. A first DNA fragment (ECU 5' homology) was amplified from ATCC20336 gDNA using primers oAA3085 (SEQ ID NO. 3716) and oAA3086 (SEQ ID NO. 3717). A second DNA fragment (PURA3URA3TURA3PURA3) was amplified from plasmid pAA298 (FIG. 29, and SEQ ID NO: 3784) using primers oAA3087 (SEQ ID NO. 3718) and oAA3088 (SEQ ID NO. 3719). The third DNA fragment (ECU 3' homology) was amplified from ATCC20336 gDNA using primers oAA3089 (SEQ ID NO. 3720) and oAA3090 (SEQ ID NO. 3721). All three DNA fragments were combined in the same reaction to generate the full-length deletion cassette (FIG. 19) by OE-PCR using primers oAA3085 (SEQ ID NO. 3716) and OAA3090 (SEQ ID NO. 3721).

Strain sAA886 was transformed with the full-length deletion cassette and plated on SCD-Ura plate. Several colonies were screened by PCR for integration of the deletion cassette at the first ECU allele. One such strain was named sAA1764.

Example 37. *Generation of Strain sAA1860 (ura3/ura3 pox4a::ura3/pox4b::ura3 POX5/POX5 acsl:: PURA3/acsl:: PURA3 fatl-Al ::PURA3/fatl-A2::PURA3 ecil-A1::PURA3/ECI1)*

Strain sAA1764 was grown in YPD media overnight and plated on 5-FOA plate. Colonies that grew in the presence of 5-FOA were PCR screened for the looping out of the URA3 gene leaving behind only the URA3 promoter (PURA3) in the first ECU allele. One such strain was named sAA1860.

Example 38. *Construction of a double ECU knockout strain (ura3/ura3 pox4a::ura3/pox4b::ura3 POX5/POX5 acsl:: PURA3/acsl:: PURA3 fatl-Al::PURA3/fatl-A2::PURA3 ecil-Al ::PURA3/ecil-A2:: URA3)*

Deletion of the second allele of ECU is achieved by transforming cells with linear DNA cassettes constructed by overlap extension PCR (OE-PCR). A deletion cassette for the second ECU allele in sAA1860 (ura3/ura3 pox4a::ura3/pox4b::ura3 POX5/POX5 acsl:: PURA3/acsl:: PURA3 fatl-Al::PURA3/fatl-A2::PURA3 ecil-Al::PURA3/ECIl) was generated from three DNA fragments. A first DNA fragment (ECU 5' homology) was amplified from ATCC20336 gDNA using primers oAA3212 (SEQ ID NO. 3722) and oAA3213 (SEQ ID NO. 3723). A second DNA fragment (PURA3URA3TURA3PURA3) was amplified from plasmid pAA298 (FIG. 29, and SEQ ID NO: 3784) using primers oAA3214 (SEQ ID NO. 3724) and oAA3215 (SEQ ID NO. 3725). A third DNA fragment (ECU 3' homology) was amplified from ATCC20336 gDNA using primers oAA3216 (SEQ ID NO. 3726) and oAA3217 (SEQ ID NO. 3727). All three DNA fragments were combined in the same reaction to generate the full-length deletion cassette (FIG. 20) by OE-PCR using primers oAA3212 (SEQ ID NO. 3722) and oAA3217 (SEQ ID NO. 3727).

To generate a double ECU knockout strain, sAA1860 is transformed with the full-length deletion cassette and plated on SCD-Ura plate. Several colonies are screened by PCR for integration of the deletion cassette at the second ECU allele.

EXAMPLE 39 - *Cloning of acyl CoA oxidase proteins*

Acyl-CoA oxidases from a range of organisms were cloned into the *E. coli* expression vector pET26b (EMD4Biosciences, Darmstadt, Germany), which contains a kanamycin resistance cassette. The source of the acyl-CoA oxidase, the name of the gene, the primers and restriction enzymes used to clone the acyl CoA oxidase coding sequence into pET26b and the coding

sequence are described herein. The acyl CoA oxidase coding sequences were amplified by PCR using the appropriate primers designed from cDNA libraries, published cDNA or genomic DNA sequences corresponding to the organism. In the event that a template source was not available, the coding sequences were synthesized as gBlocks (IDT) and stitched together by standard overlap extension PCR. The PCR products were then cloned into pCRII-Blunt TOPO vector (Life Sciences) and the products were sequenced to verify that they did not contain undesired mutations. The coding sequences were released from the TOPO vector using the appropriate restriction enzymes and ligated into pET26b that had been digested with the same restriction enzymes. The resulting expression plasmids were then transformed into Rosetta II BL21 cells (Novagen).

Example 40 - Expression of acyl-CoA oxidases in E. coli

To express an enzyme, a colony from each transformation of Rosetta cells was used to start a 5 ml overnight culture of LB containing the antibiotics kanamycin (to select for pET26b) and chloramphenicol (to select for a second plasmid found in Rosetta II cells that mediates improved translation of eukaryotic proteins expressed in *E. coli)* grown at 37°C. The next morning, the overnight culture was used to seed 30 ml of LB containing kanamycin and chloramphenicol to an OD_{600} reading of 0.1. The 30 ml cultures were grown at 37°C for 2 hours and then placed on ice for 10 minutes. To induce expression, isopropyl-beta-D-thiogalactopyranoside (IPTG) was added to the culture to a final concentration of 0.1 mM. In some cases induction was performed using Novagen Overnight Express Autoinduction System 1 (CAT# 71300-3). The cells were then shaken at 15°C overnight to express the acyl CoA oxidase.

Example 41 - Acyl-CoA Oxidase Activity Assay

To test the activity of the acyl CoA oxidase, cells from the overnight induction were pelleted at 1046 x g at 4°C in 50 ml conical tubes. The cell pellets were resuspended in 1 ml of 50mM KPO_4, pH 7.6, 50 µM FAD buffer and then transferred to a 2 ml centrifuge tube. A Misonix Sonicator 3000 (QSonica, Newtown, CT) was used to lyse the cells, which were sonicated at a power setting of 2 for 2 pulses of 20 seconds each. The lysates were placed on ice for 30 seconds in between each pulse. To obtain a supernatant, cell debris was pelleted at 16,100 x g for 10 mins in a 4°C microcentrifuge. The supernatant was transferred to a 1.5 ml centrifuge

tube and Bradford assays (Thermoscientific) were performed on cell lysates according to manufacturer's specifications to determine protein concentration in preparation for the acyl-CoA oxidase assays. A Beckman Coulter DTX-800 Multimode Detector spectrophotometer was used for the assays. The spectrophotometer was set to read for 5 minutes at 500 nm, 30°C. Each reaction was 200 µï in volume and contained 10 µg of cell lysate in 50mM $KP0_4$, pH 7.6, 200 µg/ml BSA, 0.05% Triton X-100, 250µM fatty acyl-CoA substrate, 50µM FAD, 10U horseradish peroxidase, 25mM p-hydroxybenzoic acid and ImM 4-aminoantipyrine. Fatty acyl-CoA substrates covered a range from hexanoyl CoA (six carbon chain length) to oleoyl CoA (eighteen carbon chain length).

In TABLE 22 unshaded blocks indicated the sample was not tested. Dark shading indicates that no activity was detected. Light shading indicates that minimal activity (i.e. poor activity) was detected at less than or equal to 0.1 umol/min/ug (umol substrate/minute/ug total protein). Medium shading indicates that good activity was detected at > 0.1 umol/min/ug.

The results in TABLE 22 indicated that several enzymes are not functional when expressed in E.coli. Furthermore, the remaining enzymes that are functional when expressed in E.coli showed broad substrate specificity or were similar in their substrate specificity to Pox5 from Candida strain ATCC20336 (i.e. not very active on a C6 substrate, show peak activity on a C12 substrate and are active from C8 all the way to C18:l).

TABLE 22 - Activity of Heterologous Acyl CoA Oxidases

E. coli Strain number	**Description**	**C6**	**C8**	**CI 0**	**CI 2**	**CI 4**	**CI 6**	**C18: 1**
1570	*Aspergillus nidulans* (An)AOXa in pET26b							
1557	AnAOXa (N-term. GST fusion) in pET26b							
1573	*Arabidopsis thaliana* (At)ACX2L (N-term. GST fusion) in pET26b							
1514	AtACX2L in pGEX							
1503	AtACX2L in pET26b							
1508	AtACX2L (N-term. GST fusion) in							

E. coli Strain number	Description	C6	C8	C10	C12	C14	C16	C18:1
	pET26b							
1510	AtACX1 in pET26b							
1511	AtACX3 in pET26b							
1366	AtACX2S in pET26b							
1507	AtACX2S (N-term. GST fusion) in pET26b							
1525	AtACX2S (N-term. GST fusion) in pET26b							
1663	AtACX2 N terminus-Pox5 C-terminus fusion in pET26b							
1369	AtACX5 in pET26b							
1362	Candida strain ATCC20336 Pox5 in pET26b							
1888	Candida strain ATCC20336 Pox4 in pET26b							
1365	*Debaryomyces hansenii* (Dh)7248p in pET26b							
1572	*Glycine max* (Gm)Aco2 pET26b							
1558	GmAco2 (N-term. GST fusion) in pET26b							
1515	GmAco2 in pGEX							
1367	*Pichia stipitis* (Ps)Acox2 in pET26b							
1504	*Rattus norvegicus* (Rn)Aco1 in pET26b							
1505	RnAco1b in pET26b							
1574	RnAco1 (N-term. GST fusion) in pET26b							
1575	RnAco1b (N-term. GST fusion) in pET26b							

E. coli Strain number	Description	C6	C8	C10	C12	C14	C16	C18:1
1501	*Yarrowia lipolytica* (Yl)Aco1 in pET26b							
1361	YlAco2 in pET26b							
1502	YlAco3 in pET26b							
1499	YlAco4 in pET26b							
1509	YlAco4 (N-term. GST fusion) in pET26b							
1500	YlAco5 in pET26b							
1368	YlAco6 in pET26b							
1363	*Zea mays* (Zm)Aco1 in pET26b							
1552	ZmAco2 in pET26b							
1555	ZmAco2 (N-term. GST fusion) in pET26b							
1364	ZmAco4 pET26b							
1506	ZmAco4 (N-term. GST fusion) in pET26b							
1795	*Cucumis sativus* Aco in pET26b							
1775	*Arthrobacter ureafaciens* Aco in pET26b							
1776	*Salinobacter ruber* Aco in pET26b							
1777	*Thermobifidafusca* Aco in pET26b							

Example 42 - Genetic Modification of Candida Pox4, Pox5 and R. norvegicus VLCAD

The objective was to design mutations in 1) the Pox4 and Pox5 acyl CoA oxidases of Candida strain ATCC20336 (Pox4 and Pox5, respectively) to alter their respective substrate specificities and 2) the *R. norvegicus* very long chain acyl-CoA dehydrogenase (VLCAD) to convert it into

an acyl CoA oxidase. When introduced into Candida, these mutant enzymes may mediate selective conversion of fatty acid substrates to sebacic, dodecanedioic acid or longer chain diacids by beta oxidation.

Site-directed mutagenesis of Pox4 and Pox5- Methodology

Several approaches were used to identify regions and/or residues of Pox4 and Pox5 of Candida strain ATCC20336 that determine the substrate specificities of these enzymes. In rat liver, a single gene with an alternatively spliced third exon produces two splice forms, Acol (acyl CoA oxidase-I, *R. norvegicus, RnAcol)* and AcoII (acyl CoA oxidase-II, *R. norvegicus, RnAcoII),* which are identical in amino acid length and differ in amino acid sequence only at the region encoded by the differentially spliced exon (Miyazawa, S., Hayashi, H., Hijikata, M., Ishii, N., Furuta, S., Kagamiyama, H., Osumi, T., Hashimoto, T. (1987) *Complete nucleotide sequence of cDNA andpredicted amino acid sequence of rat acyl-CoA oxidase.* J. Biol. Chem. 262(17):8138-43.; Osumi, T., Ishii, N., Miyazawa, S., Hashimoto, T. (1987) *Isolation and structural characterization of the rat acyl-CoA oxidase gene.* J. Biol. Chem. 262 (17):8138-43; Setoyama, C , Tamaoki, H., Nishina, Y., Shiga, K., Miura, R. (1995) *Functional expression of twoforms of rat acyl-CoA oxidase and their substrate specificities.* Biochem. Biophys. Res. Commun. 217(2):482-7). A comparison of the primary amino acid sequences of Acol and AcoII revealed differences in residues 90 to 133 as a result of the alternatively spliced exon (underlined residues, FIG. 23). The splicing event resulted in two enzymes, Acol and AcoII, that display different substrate activity profiles. RnAcol prefers substrates with few carbons (e.g., fatty acids with 8 or 10 carbons). RnAcoII prefers substrates with longer carbon chains (e.g., 14 carbons). The crystal structure of RnAcoII has been solved (PDB: 1IS2 (without substrate); PDB: 2DDH (with dodecanoate substrate)) and the region encoded by the alternatively spliced exon ends at the boundary between the N-terminal alpha helical domain and the subsequent beta sheet domain, both of which are characteristic structural features of acyl CoA oxidases (Acyl-CoA dehydrogenase (ACAD) superfamily, NCBI Conserved Domains Accession cl0993) and have been identified as a region that may determine substrate specificity.

To verify that this region of an acyl CoA oxidase plays a role in determining substrate specificity, the HotSpot Wizard algorithm was utilized (Pavelka, A., Chovancova, E., Damborsky, J. HotSpot Wizard: a Web Server for Identification of Hot Spots in Protein

Engineering, Nucleic Acids Research 37: W376-W383, 2009. http://loschmidt.chemi.muni.cz/hotspotwizard/). HotSpot Wizard is a program that identifies regions of a protein for engineering of substrate specificity or activity. The program utilizes structural, functional and sequence homology data from numerous databases, such as PDB, UniProt and NCBI, to identify regions and/or residues that are "hot spots" for mutagenesis. The search relies on a PDB file corresponding to a crystal structure of the enzyme of interest. In the case of Pox4 or Pox5, no such structure was available. Therefore, the structures of both proteins were determined by modeling with the crystal structure of *R. norvegicus* AcoII as the template (PDB: 1IS2). The SWISS-MODEL program was used to generate these models (Arnold K., Bordoli L., Kopp J., and Schwede T. (2006). *The SWISS-MODEL Workspace: A web-based environmentfor protein structure homology modeling.* Bioinformatics, 22,195-201; Kiefer F, Arnold K, Kiinzli M, Bordoli L, Schwede T (2009) The SWISS-MODEL Repository and associated resources. Nucleic Acids Research. 37, D387-D392. Peitsch, M. C. (1995) Protein modeling by E-mail Bio/Technology 13: 658-660; http://swissmodel.expasy.org/). The resulting models, which were PDB files, were entered in HotSpot Wizard as the "Query structure". The results of a HotSpot Wizard analysis are summarized in FIG 24A, 24B and 25). Residues highlighted in grey are proposed mutagenic "hot spots". Dark grey shading indicates residues with greater variability than those with light grey shading. Residues shown in bold are found within or close to the substrate binding pocket (discussed below).

All three enzymes were aligned to show areas of homology (FIG. 26). In FIG. 26 light grey shading indicates identity between all three enzymes. Darker shades of grey indicate partial identity or homology between the three enzymes (e.g., a dark grey shading may indicate identity for two of the three proteins in the alignment). In some cases, dark grey shading indicates sequence similarity (i.e., the residues are similar because they are acidic, basic, polar or non-polar). The underlined region (FIG. 23 and FIG. 26) indicates the alternatively spliced exon of AcoII.

Molecular modeling alignments were used to identify residues in Pox4 and Pox5 that are found within or close to the substrate binding pocket (residues shown in bold, FIG. 24A and 24B). The molecular structure of AcoII complexed with its substrate dodecanoate (PDB:2DDH) as determined from its crystal structure, was aligned with the predicted molecular models of Pox4 and Pox5. Residues located in the N-terminal loop and first part of alpha helix D (TABLE 23) appear at the surface and lining of the substrate entry/exit channel. In Pox4 these residues

correspond to the sequence IDTFNK (a.a. 95-100 of Pox4 from Candida strain ATCC20336) and PDQQAQ (a.a. 80-85 of Pox5 from Candida strain ATCC20336). According to HotSpot Wizard this entire sequence is category 9, which means highly variable.

TABLE 23 - *N-terminal loop andfirst part of alpha helix D*

Protein	**Sequence**	**Residue(s)**
1IS2 (ACOII)	ISDPEE	79-84
ACOI	ISDPEE	79-84
Pox4	IDTFNK	95-100
Pox5	PDQQAQ	80-85

Residues located in the loop between alpha helices D and E' form part of the substrate binding pocket. None of these residues were identified as contact residues for the 12 carbon substrate but may be contact residues for longer substrates. This stretch of four amino acids is located within the divergent exon splice site of ACO-I and ACO-II. In Pox4 and Pox5, three of these amino acids are highly conserved (TABLE 24). The fourth amino acid is different (113G in Pox4 & F98 in Pox5). Of the four amino acids in this region, the divergent residue is closest to the substrate and the amino acid character of this residue is drastically different between Pox4 and Pox5. For this reason, this residue is of particular interest. Additionally, according to HotSpot this residue is highly variable.

TABLE 24 - *Loop between alpha helices D and E'*

Protein	**Sequence**	**Residue(s)**
1IS2 (ACOII)	RGHP	94-97
ACOI	ANFV	94-97

Pox4	PQVG	110-1 13
Pox5	PQVF	95-98

The residue DlOl is a contact residue for the substrate carbons 6 through 9 in the 2DDH crystal structure of RnACOII. This residue is located at the beginning of alpha helix E' which is part of the substrate binding pocket. Since this is a contact residue and is located in the small region of sequence that differs between ACO-I and ACO-II, the corresponding amino acid in Pox4 and Pox5 (TABLE 25) is of interest. This residue is of interest since it contacts the substrate at carbons 6-9 and ACO-II has lower activity on substrates of chain-length 6-12 compared to ACO-I. If either Pox4 or Pox5 is modified at this position to aspartate, it is expected that there would be a decrease in activity on adipic acid and lead to increases in yield of larger diacids. This residue has a score of 6 from HotSpot.

TABLE 25 - *Residue making contact with substrate carbons 6-9*

Protein	**Sequence**	**Residue(s)**
1IS2 (ACOII)	D	101
ACOI	G	101
Pox4	G	117
Pox5	G	102

Residue F284 is a contact residue for the substrate carbons 10 through 12 in the 2DDH crystal structure of RnACOII. This residue is conserved between RnAcol and RnAcoII. The corresponding amino acid in Pox4 and Pox5 (TABLE 26) is one of the very few substrate contact residues that differ between Pox4 and Pox5. The ACOI, ACOII, and Pox4 enzymes all have a large hydrophobic residue at this location whereas the Pox5 enzyme has a small polar residue. The HotSpot score for this residue is 9.

TABLE 26 - *Residue making contact with substrate carbons 10-12*

Protein	**Sequence**	**Residue(s)**

1IS2 (ACOII)	F	284
ACOI	F	284
Pox4	M	302
Pox5	T	287

The loop C-terminal to alpha helix L is much smaller in Pox5 than it is in Pox4 or ACO-I/ACO-II (TABLE 27). This loop appears to display structural flexibility and may have implications for the structure of the substrate pocket and how much the substrate-binding pocket "breathes". The residues in this region vary in HotSpot analysis, however the residues just downstream of this region in Pox5 are all highly variable (shown next).

TABLE 27 - *Alpha helix L and loop C-terminal to alpha helix L*

Protein	**Sequence**	**Residue(s)**
1IS2 (ACOII)	IYDQVRSGKLVGGMVSYLNDLPSQRIQPQQVA	438-469
ACOI	IYDQVRSGKLVGGMVSYLNDLPSQRIQPQQVA	438-469
Pox4	QVISIEDAGKTVRGSTAFLNQLKDYTGSNSSKV	473-505
Pox5	DLLKEPEQKGL	453-463

The loop between alpha helix L and M does not appear to be as variable between Pox4 and Pox5 (TABLE 28), although HotSpot analysis assigns this stretch of residues with scores of 9 with high variability. It is expected that this loop, including the previous section mentioned above, is a target for mutagenesis.

TABLE 28 - *Loop between a-helix L and M*

Protein	**Sequence**	**Residue(s)**
1IS2 (ACOII)	VWPTMV	470-475

ACOI	VWPTMV	470-475
Pox4	VLNTVA	506-51 1
Pox5	VLSSVA	464-469

For both Pox4 and Pox5, the HotSpot Wizard analyses, combined with molecular modeling alignments, determined that residues within the same approximate regions are good targets for mutagenesis. The multiple sequence alignment shows that the alternatively spliced exon of AcoII overlaps with hot spot residues in all three acyl CoA oxidases (FIG. 26).

Some additional regions in RnACoII, indicated by their secondary structure and position of amino acid residues, are shown in TABLE 29 and corresponding regions in POX5, POX4 and Aco2, Aco3 and Aco5 of Yarrowia lipolytica are also shown.

TABLE 29

RnAcoII (1IS2 or 2DDH)		Corresponding residues in aox enzymes				
Secondary Structure	aa residues	POX5	YlAco2	POX4	YlAco3	YlAco5
Helices 1 - 6	1-76	1-77	1-97	1-92	1-98	1-96
Helices 7 - 10	82-129	83-133	103-152	98-148	104-153	102-151
β-sheet domain	134-268	138-271	157-291	153-286	158-292	156-290
Helical domain 3	278-460	281-462	301-492	296-500	302-493	300-494
Helical domain 4	475-655	469-637	502-683	511-684	502-682	503-682

Site-directed mutagenesis of Candida Pox4 and Pox5 to alter substrate specificity - Method

Using the HotSpot Wizard and molecular modeling results as a guide, specific amino acids in Pox4 and Pox5 were mutated (i.e. added, deleted or substituted) by converting primarily polar or charged residues in the hot spot regions to alanine. TABLES 30A and 30B

below show a summary of Candida strain ATCC20336 Pox5 and Pox4 mutations that were made and tested. The summary of the acyl CoA activity profile associated with some of the mutants in TABLES 30A and 30B are shown in FIG. 27 (Pox5) and FIG. 28 (Pox4). The number of carbons in each substrate tested is shown below each bar in FIG. 27 and FIG. 28. Pox5 Mutant I (grey highlight in TABLE 30A) results from "ACAD-based mutagenesis" (see discussion below).

TABLE 30A

Pox5

MUTANT	POSITION	AMINO ACID(S)	MUTATION
A	81, 82	DQ	AA
B	86, 88	RLS	ALA
C	93, 94	FD	AA
D	291, 292	DS	AA
E	95, 96	PQ	AA
F	294, 295	RM	AA
G	287	T	A
H	290, 291	MD	AA
I	284/436	GE	EG
J	291	D	G
K	292	S	A
L	93	F	A
M	94	D	G
CT1	102	G	D
N	86	R	A
O	88	S	A
P	98	F	G
Q	83, 85	QAQ	AAA
CT2	453-463	DLLKEPEQKGL	QVISIEDAGKTVRGSTAFLNQLKDYTGSNSSKV

TABLE 30B

Pox4

MUTANT	POSITION	AMINO ACID(S)	MUTATION
A	98, 99, 100	FNK	AAA
B	102, 103	LS	AA
C	96	D	A

D	90	R	A
E	88	R	A
F	302	M	A
G	309, 310	RM	A
H	98	F	A
I	99	N	A
J	100	K	A
K	102	L	A
L	103	S	A
CT3	473-505	QVISIEDAGKTVRGSTAFLNQLKDYTGSNSSKV	DLLKEPEQKGL

Pox4 and Pox5 from Candida strain ATCC20336 were cloned into pET26b for expression in E. coli and assayed for acyl CoA oxidase activity in vitro. The activity profiles of the genetically modified Pox4 and Pox5 were compared to the activity profile of the wild type enzymes. To alter their substrate activity profile, site directed mutagenesis was performed on several locations in Pox4 and Pox5. Complementary primers encoding the point mutation(s) were used to amplify the coding sequences of Pox4 or Pox5 generating two to four PCR products that were then "stitched" together to regenerate the entire coding region using overlap extension PCR (FIG. 22). As shown in FIG. 22, overlap extension PCR was performed using primers A, B, C and D. Primers B and C are complementary and contain the introduced genetic modifications (e.g. mutations). PCR was performed using oligonucleotides A and B to produce a product with overlap to a PCR product generated using oligonucleotides C and D. The A-B product was used as a primer for the C-D product, and vice versa, for overlap extension. Several mutagenic primer pairs, for example, like the B-C primer pair, were used to produce mutations at different locations that were "stitched" together, i.e. A-B, C-D, E-F, etc. to generate an intact, full length coding region. To produce more of the final product containing the mutation(s), a PCR using the A primer and the most 3' reverse primer was performed. Primers A and D were used to amplify the entire coding sequence of Pox4 and Pox5 and to incorporate the restriction enzyme sites (REl and RE2) for cloning into an E. coli expression vector. The primers used for the site-directed mutagenesis for Pox5 (Candida strain ATCC20336) are listed in TABLE 31 to TABLE 40. The primers used for the site-directed mutagenesis for Pox4 (Candida strain ATCC20336) are listed in TABLE 41 to TABLE 47.

TABLE 31 - Pox5 (Candida strain ATCC20336)

Mutant Name	Position	Native Amino Acid(s)	Introduced Mutation	Primer	Primer Sequence (5'-3')	Restriction Sites
A				PRIMER A	GTTCACTGCCATATGCCTACCGAACTTCAAAAAGAAAGAGAACTC	NdeI
A	81, 82	DQ	AA	PRIMER B	GATCGACAATCTCTGGGCCTGAGCAGCTGGGTACTCGTGCTCAAAG	
A	81, 82	DQ	AA	PRIMER C	CTTTGAGCACGAGTACCCAGCTGCTCAGGCCCAGAGATTGTCGATC	
A				PRIMER D	CTTCGAGATGCGGCCGCTTAACTGGACAAGATTTCAGCAGCTTCTTCG	NotI
B				PRIMER A	GTTCACTGCCATATGCCTACCGAACTTCAAAAAGAAAGAGAACTC	NdeI
B	86, 88	RLS	ALA	PRIMER B	GTGGGTCAAAGACACCGAGGATAGCCAAAGCCTGGGCCTGTTGGTCTGGGTAC	
B	86, 88	RLS	ALA	PRIMER C	GTACCCAGACCAACAGGCCCAGGCTTTGGCTATCCTCGGTGTCTTTGACCCAC	
B				PRIMER D	CTTCGAGATGCGGCCGCTTAACTGGACAAGATTTCAGCAGCTTCTTCG	NotI

TABLE 32 - Pox5 (Candida strain ATCC20336)

Mutant Name	Position	Native Amino Acid(s)	Introduced Mutation	Primer	Primer Sequence (5'-3')	Restriction Sites
C				PRIMER A	GTTCACTGCCATATGCCTACCGAACTTCAAAAAGAAAGAGAACTC	NdeI
C	93, 94	FD	AA	PRIMER B	GATTCTGGTGAAGACTTGTGGAGCAGCGACACCGAGGATCGACAATC	
C	93, 94	FD	AA	PRIMER C	GATTGTCGATCCTCGGTGTCGCTGCTCCACAAGTCTTCACCAGAATC	
C				PRIMER D	CTTCGAGATGCGGCCGCTTAACTGGACAAGATTTCAGCAGCTTCTTCG	NotI
D				PRIMER A	GTTCACTGCCATATGCCTACCGAACTTCAAAAAGAAAGAGAACTC	NdeI
D	291, 292	DS	AA	PRIMER B	GAATCTACTGGTCATTCTGTAAGCAGCCATCATCATGGTGACTCTACC	
D	291, 292	DS	AA	PRIMER C	GGTAGAGTCACCATGATGATGGCTGCTTACAGAATGACCAGTAGATTC	
D				PRIMER D	CTTCGAGATGCGGCCGCTTAACTGGACAAGATTTCAGCAGCTTCTTCG	NotI

TABLE 33 - Pox5 (Candida strain ATCC20336)

Mutant Name	Position	Native Amino Acid(s)	Introduced Mutation	Primer	Primer Sequence (5'-3')	Restriction Sites
E				PRIMER A	GTTCACTGCCATATGCCTACCGAACTTCAAAAAGAAAGAGAACTC	NdeI
E	95, 96	PQ	AA	PRIMER B	CACCGATTCTGGTGAAGACAGCAGCGTCAAAGACACCGAGGATCG	
E	95, 96	PQ	AA	PRIMER C	CGATCCTCGGTGTCTTTGACGCTGCTGTCTTCACCAGAATCGGTG	
E				PRIMER D	CTTCGAGATGCGGCCGCTTAACTGGACAAGATTTCAGCAGCTTCTTCG	NotI
F				PRIMER A	GTTCACTGCCATATGCCTACCGAACTTCAAAAAGAAAGAGAACTC	NdeI
F	294, 295	RM	AA	PRIMER B	GGTGATGAATCTACTGGTCGCGGCGTAGGAGTCCATCATCATG	
F	294, 295	RM	AA	PRIMER C	CATGATGATGGACTCCTACGCCGCGACCAGTAGATTCATCACC	
F				PRIMER D	CTTCGAGATGCGGCCGCTTAACTGGACAAGATTTCAGCAGCTTCTTCG	NotI

TABLE 34 - Pox5 (Candida strain ATCC20336)

Mutant Name	Position	Native Amino Acid(8)	Introduced Mutation	Primer	Primer Sequence (5'-3')	Restriction Sites
G				PRIMER A	GTTCACTGCCATATGCCTACCGAACTTCAAAAAGAAAGAGAACTC	NdeI
G	287	T	A	PRIMER B	GAGTCCATCATCATGGCGACTCTACCACCAATC	
G	287	T	A	PRIMER C	GATTGGTGGTAGAGTCGCCATGATGATGGACTC	
G				PRIMER D	CTTCGAGATGCGGCCGCTTAACTGGACAAGATTTCAGCAGCTTCTTCG	NotI
H				PRIMER A	GTTCACTGCCATATGCCTACCGAACTTCAAAAAGAAAGAGAACTC	NdeI
H	290, 291	MD	AA	PRIMER B	CTGGTCATTCTGTAGGAGGCTGCCATCATGGTGACTCTACC	
H	290, 291	MD	AA	PRIMER C	GGTAGAGTCACCATGATGGCAGCCTCCTACAGAATGACCAG	
H				PRIMER D	CTTCGAGATGCGGCCGCTTAACTGGACAAGATTTCAGCAGCTTCTTCG	NotI

TABLE 35 - Pox5 (Candida strain ATCC20336)

Mutant Name	Position	Native Amino Acid(s)	Introduced Mutation	Primer	Primer Sequence (5'-3')	Restriction Sites
I				PRIMER A	GTTCACTGCCATATGCCTACCGAACTTCAAAAAGAAAGAGAACTC	NdeI
I	284/436	GE	EG	PRIMER B	CATCATCATGGTGACTCTTTCACCAATCAAAGCCGAG	
I	284/436	GE	EG	PRIMER C	CTCGGCTTTGATTGGTGAAAGAGTCACCATGATGATG	
I	284/436	GE	EG	PRIMER D	GTTGTTGTCACCTCCCCAGGTACATTGG	
I	284/436	GE	EG	PRIMER E	CCAATGTACCTGGGGAGGTGACAACAAC	
I				PRIMER F	CTTCGAGATGCGGCCGCTTAACTGGACAAGATTTCAGCAGCTTCTTCG	NotI
J				PRIMER A	GTTCACTGCCATATGCCTACCGAACTTCAAAAAGAAAGAGAACTC	NdeI
J	291	D	G	PRIMER B	GGTCATTCTGTAGGAGCCCATCATCATGGTGAC	
J	291	D	G	PRIMER C	GTCACCATGATGATGGGCTCCTACAGAATGACC	
J				PRIMER D	CTTCGAGATGCGGCCGCTTAACTGGACAAGATTTCAGCAGCTTCTTCG	NotI

TABLE 36 - Pox5 (Candida strain ATCC20336)

Mutant Name	Position	Native Amino Acid(s)	Introduced Mutation	Primer	Primer Sequence (5'-3')	Restriction Sites
K				PRIMER A	GTTCACTGCCATATGCCTACCGAACTTCAAAAAGAAAGAGAACTC	NdeI
K	292	S	A	PRIMER B	CTGGTCATTCTGTAGGCGTCCATCATCATGGTG	
K	292	S	A	PRIMER C	CACCATGATGATGGACGCCTACAGAATGACCAG	
K				PRIMER D	CTTCGAGATGCGGCCGCTTAACTGGACAAGATTTCAGCAGCTTCTTCG	NotI
L				PRIMER A	GTTCACTGCCATATGCCTACCGAACTTCAAAAAGAAAGAGAACTC	NdeI
L	93	F	A	PRIMER B	GTGAAGACTTGTGGGTCTGCGACACCGAGGATCGAC	
L	93	F	A	PRIMER C	GTCGATCCTCGGTGTCGCAGACCCACAAGTCTTCAC	
L				PRIMER D	CTTCGAGATGCGGCCGCTTAACTGGACAAGATTTCAGCAGCTTCTTCG	NotI

Mutant Name	Position	Native Amino Acid(s)	Introduced Mutation	Primer	Primer Sequence (5'-3')	Restriction Sites
M				PRIMER A	GTTCACTGCCATATGCCTACCGAACTTCAAAAAGAAAGAGAACTC	Nde1
M	94	D	G	PRIMER B	GGTGAAGACTTGTGGGCCAAAGACACCGAGGATC	
M	94	D	G	PRIMER C	GATCCTCGGTGTCTTTGGCCCACAAGTCTTCACC	
M				PRIMER D	CTTCGAGATGCGGCCGCTTAACTGGACAAGATTTCAGCAGCTTCTTCG	Not1
CT1*				PRIMER A	CACACAAGGGGAATTGTGAGCGGATAAC	XbaI
CT1*	102	G	D	PRIMER B	AACCCAAGTTGACGTCGATTCTGG	
CT1*	102	G	D	PRIMER C	CCAGAATCGACGTCAACTTGGGTT	
CT1*				PRIMER D	CACACAAACTGGATCCAACCGTTATCG	BamHI

TABLE 37 - Pox5 (Candida strain ATCC20336)

* Mutant CT1 produces a smaller fragment that is used to replace the sequence in between the Xbal and BamHI sites of wild type Pox5 cloned into pET26b.

TABLE 38 - Pox5 (Candida strain ATCC20336)

Mutant Name	Position	Native Amino Acid(s)	Introduced Mutation	Primer	Primer Sequence (5'-3')	Restriction Sites
N				PRIMER A	GTTCACTGCCATATGCCTACCGAACTTCAAAAAGAAAGAGAACTC	NdeI
N	86	R	A	PRIMER B	CAAAGACACCGAGGATCGACAAAGCCTGGGCCTGTTGGTCTGGGTAC	
N	86	R	A	PRIMER C	GTACCCAGACCAACAGGCCCAGGCTTTGTCGATCCTCGGTGTCTTTG	
N				PRIMER D	CTTCGAGATGCGGCCGCTTAACTGGACAAGATTTCAGCAGCTTCTTCG	NotI
0				PRIMER A	GTTCACTGCCATATGCCTACCGAACTTCAAAAAGAAAGAGAACTC	NdeI
0	88	S	A	PRIMER B	CAAAGACACCGAGGATCGCCAATCTCTGGGCCTGTTG	
0	88	S	A	PRIMER C	CAACAGGCCCAGAGATTGGCGATCCTCGGTGTCTTTG	
0				PRIMER D	CTTCGAGATGCGGCCGCTTAACTGGACAAGATTTCAGCAGCTTCTTCG	NotI

TABLE 39 - Pox5 (Candida strain ATCC20336)

Mutant Name	Position	Native Amino Acid(s)	Introduced Mutation	Primer	Primer Sequence (5'-3')	Restriction Sites
P				PRIMER A	GTTCACTGCCATATGCCTACCGAACTTCAAAAAGAAAGAGAACTC	NdeI
P	98	F	G	PRIMER B	GTTGACACCGATTCTGGTTCCGACTTGTGGGTCAAAGAC	
P	98	F	G	PRIMER C	GTCTTTGACCCACAAGTCGGAACCAGAATCGGTGTCAAC	
P				PRIMER D	CTTCGAGATGCGGCCGCTTAACTGGACAAGATTTCAGCAGCTTCTTCG	NotI
Q				PRIMER A	GTTCACTGCCATATGCCTACCGAACTTCAAAAAGAAAGAGAACTC	NdeI
Q	83, 85	QAQ	AAA	PRIMER B	CAAAGACACCGAGGATCGACAATCTAGCCGCAGCTTGGTCTGGGTACTCGTGCTCAAAG	
Q	83, 85	QAQ	AAA	PRIMER C	CTTTGAGCACGAGTACCCAGACCAAGCTGCCGCTAGATTGTCGATCCTCGGTGTCTTTG	
Q				PRIMER D	CTTCGAGATGCGGCCGCTTAACTGGACAAGATTTCAGCAGCTTCTTCG	NotI

TABLE 40 - Pox5 (Candida strain ATCC20336)

Mutant Name	Position	Native Amino Acid(8)	Introduced Mutation	Primer	Primer Sequence (5'-3')	Restriction Sites
CT2				PRIMER A	GTTCACTGCCATATGCCTACCGAACTTCAAAAAGAAAGAGAACTC	NdeI
CT2	453-463	DLLKEPEQKGL	QVISIEDAGKTVRGSTAFLNQLKDYTGSNSSKV	PRIMER B	CGGCATCTTCAATGCTGATAACTTGCTCTAACCATTGGCTTGGCA	
CT2	453-463	DLLKEPEQKGL	QVISIEDAGKTVRGSTAFLNQLKDYTGSNSSKV	PRIMER C	TGCCAAGCCAATGGTTAGAGCAAGTTATCAGCATTGAAGATGCC	
CT2	453-463	DLLKEPEQKGL	QVISIEDAGKTVRGSTAFLNQLKDYTGSNSSKV	PRIMER D	TCGGCAACGCTGGAGAGAACAACCTTGGAGCTGTTGGAACCAGTGT	
CT2	453-463	DLLKEPEQKGL	QVISIEDAGKTVRGSTAFLNQ	PRIMER E	ACACTGGTTCCAACAGCTCCAAGGTTGTTCTCTCCAGCGTTGCCGA	

			LKDYTGSNSSKV			
CT2				PRIMER F	CTTCGAGATGCGGCCGCTTAACTGGACAAGATTTCAGCAGCTTCTTCG	NotI

TABLE 41 - Pox4 (Candida strain ATCC20336)

Mutant Name	Position	Native Amino Acid(s)	Introduced Mutation	Primer	Primer Sequence (5'-3')	Restriction Sites
A				PRIMER A	GTTCACTGCCATATGACTTTTACAAAGAAAAACGTT AGTGTA	Nde1
A	98, 99, 100	FNK	AAA	PRIMER B	CAAAGATACCAATCAAGGACAATCTAGCAGCAGCA GTGTCGATGGATTCTTGTTCTCTG	
A	98, 99, 100	FNK	AAA	PRIMER C	CAGAGAACAAGAATCCATCGACACTGCTGCTGCTA GATTGTCCTTGATTGGTATCTTTG	
A				PRIMER D	CTTCGAGATGCGGCCGCTTATTACTTGGACAAGAT AGCAGCGGTTTCATCAGA	Not1
B				PRIMER A	GTTCACTGCCATATGACTTTTACAAAGAAAAACGTT AGTGTA	Nde1
B	102, 103	LS	AA	PRIMER B	GTGGGTCAAAGATACCAATCAAAGCAGCTCTCTTG TTGAAAGTGTCGATG	
B	102, 103	LS	AA	PRIMER C	CATCGACACTTTCAACAAGAGAGCTGCTTTGATTG GTATCTTTGACCCAC	
B				PRIMER D	CTTCGAGATGCGGCCGCTTATTACTTGGACAAGAT AGCAGCGGTTTCATCAGA	Not1

TABLE 42 - Pox4 (Candida strain ATCC20336)

Mutant Name	Position	Native Amino Acid(s)	Introduced Mutation	Primer	Primer Sequence (5'-3')	Restriction Sites
C				PRIMER A	GTTCACTGCCATATGACTTTTACAAAGAAAAACGTT AGTGTA	NdeI
C	96	D	A	PRIMER B	GGACAATCTCTTGTTGAAAGTAGCGATGGATTCTT GTTCTCTG	
C	96	D	A	PRIMER C	CAGAGAACAAGAATCCATCGCTACTTTCAACAAGA GATTGTCC	
C				PRIMER D	CTTCGAGATGCGGCCGCTTATTACTTGGACAAGAT AGCAGCGGTTTCATCAGA	NotI
D				PRIMER A	GTTCACTGCCATATGACTTTTACAAAGAAAAACGTT AGTGTA	NdeI
D	90	R	A	PRIMER B	GTGTCGATGGATTCTTGTTCAGCGTATCTGGCGAT TCTGTTG	
D	90	R	A	PRIMER C	CAACAGAATCGCCAGATACGCTGAACAAGAATCCA TCGACAC	
D				PRIMER D	CTTCGAGATGCGGCCGCTTATTACTTGGACAAGAT AGCAGCGGTTTCATCAGA	NotI

TABLE 43 - Pox4 (Candida strain ATCC20336)

Mutant Name	Position	Native Amino Acid(8)	Introduced Mutation	Primer	Primer Sequence (5'-3')	Restriction Sites
E				PRIMER A	GTTCACTGCCATATGACTTTTACAAAGAAAAACGTTAGTGTA	NdeI
E	88	R	A	PRIMER B	GATGGATTCTTGTTCTCTGTAAGCGGCGATTCTGTTGATCTTGAC	
E	88	R	A	PRIMER C	GTCAAGATCAACAGAATCGCCGCTTACAGAGAACAAGAATCCATC	
E				PRIMER D	CTTCGAGATGCGGCCGCTTATTACTTGGACAAGATAGCAGCGGTTTCATCAGA	NotI
F				PRIMER A	GTTCACTGCCATATGACTTTTACAAAGAAAAACGTTAGTGTA	NdeI
F	302	M	A	PRIMER B	GAGTCCAAAACCATCGCGACTCTACCACCCAAC	
F	302	M	A	PRIMER C	GTTGGGTGGTAGAGTCGCGATGGTTTTGGACTC	
F				PRIMER D	CTTCGAGATGCGGCCGCTTATTACTTGGACAAGATAGCAGCGGTTTCATCAGA	NotI

TABLE 44 - Pox4 (Candida strain ATCC20336)

Mutant Name	Position	Native Amino Acid(8)	Introduced Mutation	Primer	Primer Sequence (5'-3')	Restriction Sites
G				PRIMER A	GTTCACTGCCATATGACTTTTACAAAGAAAAACGTTAGTGTA	Ndel
G	309, 310	RM	A	PRIMER B	GTGGACATTCTAGCCAACGCGGCGTAGGAGTCCAAAACCATC	
G	309, 310	RM	A	PRIMER C	GATGGTTTTGGACTCCTACGCCGCGTTGGCTAGAATGTCCAC	
G				PRIMER D	CTTCGAGATGCGGCCGCTTATTACTTGGACAAGATAGCAGCGGTTTCATCAGA	NotI
H				PRIMER A	GTTCACTGCCATATGACTTTTACAAAGAAAAACGTTAGTGTA	Ndel
H	98	F	A	PRIMER B	CAAGGACAATCTCTTGTTGGCAGTGTCGATGGATTCTTG	
H	98	F	A	PRIMER C	CAAGAATCCATCGACACTGCCAACAAGAGATTGTCCTTG	
H				PRIMER D	CTTCGAGATGCGGCCGCTTATTACTTGGACAAGATAGCAGCGGTTTCATCAGA	NotI

TABLE 45 - Pox4 (Candida strain ATCC20336)

Mutant Name	Position	Native Amino Acid(8)	Introduced Mutation	Primer	Primer Sequence (5'-3')	Restriction Sites
I				PRIMER A	GTTCACTGCCATATGACTTTTACAAAGAAAAACGTTAGTGTA	NdeI
I	99	N	A	PRIMER B	CAATCAAGGACAATCTCTTCGCGAAAGTGTCGATGGATTC	
I	99	N	A	PRIMER C	GAATCCATCGACACTTTCGCGAAGAGATTGTCCTTGATTG	
I				PRIMER D	CTTCGAGATGCGGCCGCTTATTACTTGGACAAGATAGCAGCGGTTTCATCAGA	NotI
J				PRIMER A	GTTCACTGCCATATGACTTTTACAAAGAAAAACGTTAGTGTA	NdeI
J	100	K	A	PRIMER B	CAATCAAGGACAATCTCGCGTTGAAAGTGTCGATG	
J	100	K	A	PRIMER C	CATCGACACTTTCAACGCGAGATTGTCCTTGATTG	
J				PRIMER D	CTTCGAGATGCGGCCGCTTATTACTTGGACAAGATAGCAGCGGTTTCATCAGA	NotI

TABLE 46 - Pox4 (Candida strain ATCC20336)

Mutant Name	Position	Native Amino Acid(s)	Introduced Mutation	Primer	Primer Sequence (5'-3')	Restriction Sites
K				PRIMER A	GTTCACTGCCATATGACTTTTACAAAGAAAAACGTTAGTGTA	NdeI
K	102	L	A	PRIMER B	CAAAGATACCAATCAAGGAGGCTCTCTTGTTGAAAGTGTCG	
K	102	L	A	PRIMER C	CGACACTTTCAACAAGAGAGCCTCCTTGATTGGTATCTTTG	
K				PRIMER D	CTTCGAGATGCGGCCGCTTATTACTTGGACAAGATAGCAGCGGTTTCATCAGA	NotI
L				PRIMER A	GTTCACTGCCATATGACTTTTACAAAGAAAAACGTTAGTGTA	NdeI
L	103	S	A	PRIMER B	GTCAAAGATACCAATCAAGGCCAATCTCTTGTTGAAAGTG	
L	103	S	A	PRIMER C	CACTTTCAACAAGAGATTGGCCTTGATTGGTATCTTTGAC	
L				PRIMER D	CTTCGAGATGCGGCCGCTTATTACTTGGACAAGATAGCAGCGGTTTCATCAGA	NotI

TABLE 47 - Pox4 (Candida strain ATCC20336)

Mutant Name	Position	Native Amino Acid(s)	Introduced Mutation	Primer	Primer Sequence (5'-3')	Restriction Sites
CT3				PRIMER A	GTTCACTGCCATATGACTTTTACAAAGAAAAACGTTAGTGTA	NdeI
CT3	473-505	QVISIEDAGKTVRGSTAFLNQLKDYTGSNSSKV	DLLKEPEQKGL	PRIMER B	CAATCCCTTTTGTTCTGGCTCCTTCAACAAGTCCTTGACAATTGGCTTACCAA	
CT3	473-505	QVISIEDAGKTVRGSTAFLNQLKDYTGSNSSKV	DLLKEPEQKGL	PRIMER C	GACTTGTTGAAGGAGCCAGAACAAAAGGGATTGGTTTTGAACACTGTTGCTGA	
CT3				PRIMER D	CTTCGAGATGCGGCCGCTTATTACTTGGACAAGATAGCAGCGGTTTCATCAGA	NotI

In vitro acyl CoA oxidase Assay

E.coli lysates were tested for acyl CoA oxidase activity as described in Example 41.

In vitro Activity Assayfor Pox4 Mutants

TABLE 48 shows the acyl CoA oxidase activity profile associated with Pox4 mutants and TABLE 49 shows the acyl CoA oxidase activity profile associated with Pox5 mutants. The carbon length of the substrates tested is indicated above the data as C6 (6 carbons), C8 (8 carbons), CIO (10 carbons), C12 (twelve carbons), C14 (fourteen carbons), C16 (sixteen carbons) and CI8.1 (eighteen carbons). In TABLE 48 and 49, unshaded blocks indicated the sample was not tested. Dark shading indicates that no activity was detected. Light shading indicates that minimal activity (i.e. poor activity) was detected at less than or equal to 0.1 umol/min/ug (umol substrate/minute/ug total protein). Medium shading indicates that good activity was detected at > 0.1 umol/min/ug.

The Pox4 Mutant C, although displaying good activity across all substrates tested, demonstrated reduced overall activity for all substrates (TABLE 48). Pox4 Mutant D showed a similar result. Activity on C12 and CI8:1 substrates was abolished in Pox4 Mutants B, A, E and G (TABLE 48) and CT3 (not shown).

TABLE 48 - *In vitro Activity Assayfor Pox4 Mutants*

Mutant Name	**SDM** Candida strain ATCC20336 **Pox4 Mutants (all in pET26b)**	**C6**	**C8**	**C10**	**C12**	**C14**	**C16**	**C18:1**
WT	Pox4 WT							
B	Pox4 with L102A, S103A							
A	Pox4 with F98A, N99A, K100A							
C	Pox4 with D96A							
C*	Pox4 with D96A, L692S							
D	Pox4 with R90A							

E	Pox4 with R88A							
F	Pox4 with M302A							
G	Pox4 with R309A, M310A							

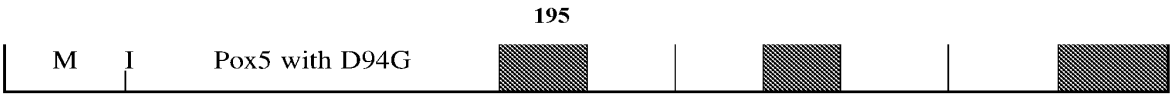
* Indicates a secondary mutation (1692S) that occurred during PCR for cloning Candida strain ATCC20336 POX5

In vitro Activity Assayfor Pox5 Mutants - Results

Acyl CoA oxidase activity was abolished in Pox5 Mutants B, C, F and M at least on substrates C6, C12 and C18:1 (TABLE 49). Mutants CT1 and CT2 were also inactive (not shown). Mutants A, E and I showed no change when compared to the activity of the wild type protein. However, Pox5 Mutants D, H, G, and J displayed altered substrate specificity when compared to wild type Pox5. Pox5 mutants D, H and J demonstrated reduced activity on C6 and/or C8 substrates. Pox5 mutant G displayed increased activity on C18:1 substrates.

TABLE 49 - *In vitro Activity Assayfor Pox5 Mutants*

Mutant Name	**SDM** Candida strain ATCC20336 **Pox5 Mutants (all in pET26b)**	**C6**	**C8**	**C10**	**C12**	**C14**	**C16**	**C18:1**
WT	Pox5 WT							
A	Pox5 with D81A, Q82A							
B	Pox5 with R86A, S88A							
C	Pox5 with F93A, D94A							
D	Pox5 with D291A, S292A							
G	Pox5 with T287A							
H	Pox5 with M290A, D291A							
E	Pox5 with PQ95AA							
F	Pox5 with R294A, M295A							
J	Pox5 with D291G							
K	Pox5 with S292A							
L	Pox5 with F93A							



Acyl-CoA dehydrogenase-based mutagenesis

Acyl-CoA oxidases and acyl-CoA dehydrogenases (ACAD) both utilize similar but distinct mechanisms to catalyze dehydrogenation of an acyl-CoA substrate to produce a *2-trans-enoyl*-CoA, the first step in β-oxidation (Arent, S., Pye, V.E., Henriksen, A. (2008). Structure and function of plant acyl CoA oxidases. Plant Phys. Biochem. 46:292-301). There are acyl-CoA dehydrogenases of different classes and they are grouped according to their substrate specificities: very long, long, medium and short chain (VLCAD, LCAD, MCAD, SCAD, respectively) (Kim, J.J., Miura, R. (2004). Acyl-CoA dehydrogenases and acyl CoA oxidases. Structural basis for mechanistic similarities and differences. Eur. J. Biochem., 271(3):483-93.). The crystal structures of several of these enzymes have been solved and these data show structural differences that very likely contribute to their respective differences in substrate specificity. The crystal structure of VLCAD (PDB: 3B96) has revealed regions and amino acid residues of the protein that make it structurally, and more than likely functionally, different from MCAD (PDB: 3MDE) (McAndrew, R.P., Wang, Y., Mohsen, A.W., He, M., Vockley, J., Kim, J.J. (2008). Structural basis for substrate fatty acyl chain specificity: crystal structure of human very-long-chain acyl-CoA dehydrogenase. *J. Biol. Chem.* 283(14):9435-43). In some cases, a more significant difference is the location of the catalytic residue. In MCAD, the catalytic glutamate is located at position 376 on the loop connecting helix J and K while in LCAD, the catalytic glutamate is at position 255 on the adjacent helix G (Nandy, A., Kieweg, V., Krautle, F.G., Vock, P., Kuchler, B., Brass, P., Kim, J.J., Rasched, I., Ghisla, S. (1996). *Medium-long-chain chimeric human Acyl-CoA dehydrogenase: medium-chain enzyme with the active center base arrangement of long-chain Acyl-CoA dehydrogenase.* Biochemistry, 35(38):12402-1 1; Lee, H.J., Wang, M., Paschke, R., Nandy, A., Ghisla, S., Kim, J.J. (1996). *Crystal structures of the wild type and the Glu376Gly/Thr255Glu mutant of human medium-chain acyl-CoA dehydrogenase: influence of the location of the catalytic base on substrate specificity.* Biochemistry, 35(38): 12412-20).

The crystal structure of VLCAD (PDB: 3B96) has also revealed regions and amino acid residues of the protein that make it structurally, and more than likely functionally, different from MCAD (PDB: 3MDE) (McAndrew et al, 2008). VLCAD is larger than other acyl-CoA dehydrogenase proteins and forms a dimer, like a typical acyl CoA oxidase. Its substrate

binding cavity is larger compared to other acyl-CoA dehydrogenase proteins and resembles an acyl CoA oxidase substrate binding pocket. The larger and more spacious pocket is necessary for accommodating the longer fatty acyl-CoA substrates that it acts upon. However, the crystal structures of rat AcoII and Arabidopsis thaliana ACX1 (PDB ID: 1W07) also reveal large substrate binding pockets and this feature does not necessarily explain the substrate specificities of each enzyme (Arent et al., 2008). Structural differences between MCAD and VLCAD offer some insight. At the base of the MCAD substrate binding pocket, there are two polar/charged residues (Q95 and E99) that are different from the analogous residues in VLCAD (G175 and G178). The increased hydrophobicity of the base of the VLCAD substrate binding pocket may be a factor, in addition to pocket size and depth, which contributes to substrate specificity. The corresponding residues in Candida strain ATCC20336 Pox5 are F98 and G102. Mutant P (F98G) is the mutation that should more closely reproduce the base of the VLCAD substrate binding pocket.

A double mutation in MCAD (e.g., E376G, T255E) can change its substrate specificity profile. This double mutation produced somewhat of a chimeric enzyme (MLCAD)(Nandy, et.al.1996). MCAD has a broad substrate profile (C4 - C18) with peak activity at C6 and C8. LCAD has a similarly broad profile with peak activity at C10 and C12. MLCAD has a more defined substrate profile (C10 - C18) compared to MCAD or LCAD with peak activity at C12. However, the overall enzymatic activity of the MLCAD was also reduced (V_{max} of MLCAD for C12 substrate is approximately 25% of V_{max} of LCAD for C12 substrate).

Based in part on the results of the above studies, Pox5 was mutated as described in TABLE 30 (Mutant I) to shift its substrate profile to preferentially act on longer chain substrates.

VLCAD mutagenesis

VLCAD has a substrate profile that is appropriate for production of longer chain diacids, such as sebacic or dodecanedioic acid. Activity of the enzyme ranges from acyl substrates that are 10 carbons to 22 carbons long and peak activity is on a C16 substrate. However, the enzymatic mechanism of VLCAD differs from that of a typical acyl CoA oxidase with respect to the final electron acceptor; in VLCAD, the enzyme is reoxidized by electron transfer ferroprotein (ETF) and AOXs are reoxidized by oxygen to produce hydrogen peroxide (Arent et al, 2008; Kim and Miura, 2004). To accommodate the difference in mechanisms, the substrate binding pocket of

an acyl CoA oxidase, such as A. thaliana ACX1, is more spacious than that of VLCAD to allow oxygen into the pocket to act as the final electron acceptor and reoxidize the flavine adenine dinucleotide, or FAD, cofactor required for dehydrogenation of the acyl-CoA substrate. ETF performs this function in a typical acad and reoxidation of FAD by oxygen is inhibited while substrate is bound (Kumar, N.R., Srivastava, D.K. (1995). Facile and restricted pathways for the dissociation of octenoyl-CoA from the medium-chain fatty acyl-CoA dehydrogenase (MCAD)-FADH2-octenoyl-CoA charge-transfer complex: energetics and mechanism of suppression of the enzyme's oxidase activity. Biochemistry, 34(29): 9434-43). This is reflected in the shape of the substrate binding pocket with respect to FAD. In an acyl CoA oxidase, FAD is more solvent exposed, but in MCAD, the entire flavin ring is embedded in the protein and is only accessible to solvent when substrate is not present (Kim and Miura, 2004). In order for an acad to have oxidase activity, the substrate binding pocket must become more solvent accessible and permit oxidation of the reduced FAD cofactor by oxygen. Mutagenesis studies of MCAD have identified a residue that can achieve this result. Tyrosine 375 in MCAD, when changed to a lysine, confers significantly increased (~200-fold increase relative to wild type MCAD) acyl CoA oxidase activity (Zeng, J., Liu, Y., Wu, L., Li, D. (2007). Mutation of Tyr375 to Lys375 allows medium-chain acyl-CoA dehydrogenase to acquire acyl CoA oxidase activity. Biochim. Biophys. Acta, 1774(12): 1628-34).

Molecular modeling suggests that the mutation increases solvent accessibility near the FAD moiety in the active site. In order for VLCAD to function as an acyl CoA oxidase with the appropriate substrate specificity profile, an analogous mutation in VLCAD is made. Tyrosine 375 in MCAD corresponds to phenylalanine 461 in human and rat VLCAD. A F461K mutation in VLCAD is tested to see if it will now have acyl CoA oxidase activity.

Example 43: Generation of Strain sAA2220

A DNA fragment encoding the *ECU* second allele was amplified from strain sAA1764 genomic DNA using primers oAA3091 and oAA3092. The PCR product was gel purified and ligated into pCR-Blunt II-TOPO (Invitrogen), transformed into competent TOP10 *E. coli* cells (Invitrogen) and clones containing PCR inserts were sequenced to confirm correct DNA sequence. One such plasmid was named pAA756.

The deletion of the second *ECU* allele in sAA1860 was created by a PCR based nested strategy. The second *ECU* deletion cassette was PCR amplified from three pieces using primer oAA3212 and oAA3217. The first piece (ECIl_N_Nested2) was amplified from pAA756 using primer oAA3212 and oAA3213. The second piece (URA3 cassette with direct repeat) was amplified from plasmid pAA298 using primer oAA3214 and oAA3215. The third piece (ECIl_C_Nested 2) was amplified from pAA756 using primer oAA3216 and oAA3217.

Strain sAA1860 was transformed with the second *ECU* deletion cassette described above and plated on a SCD-ura plate. Several colonies were checked for double *ECU* deletion. One such positive strain was named sAA2220

TABLE 50 - Primer List

Primer	sequence
oAA3085	ATCGTTACCACCATCCCTACAAT
oAA3086	CCGAAACAACCGTAGATACCTTTAAGCTACAACACTATACACGATAATTCCC
oAA3087	GGGAATTATCGTGTATAGTGTTGTAGCTTAAAGGTATCTACGGTTGTTTCGG
oAA3088	CTTGGACATTTCGACCTTGGCGGTACCGAGCTCTGCGAATT
oAA3089	AATTCGCAGAGCTCGGTACCGCCAAGGTCGAAATGTCCAAG
oAA3090	GCTTGTTCTGCAAAATGGAGTCA
oAA3091	AGGAAAGACGACCATCTTGTACAA
oAA3092	TGGTCTCTGGACAACTTCAACAAT
oAA3212	GGGGGAGATCGTTACCACCA
oAA3213	AATTCGCAGAGCTCGGTACCGCTGCTGCTGCTGCTGTTTT
oAA3214	AAAACAGCAGCAGCAGCAGCGGTACCGAGCTCTGCGAATT
oAA3215	TTCGTTGTTGGCTCTCTCCATTAAAGGTATCTACGGTTGTTTCGG
oAA3216	CCGAAACAACCGTAGATACCTTTAATGGAGAGAGCCAACAACGAA
oAA3217	CAAAGGCATCGGTCAACTCC

Example 44: Shake flask characterization of sAA875 and sAA2220

A 250 mL glass, non-baffled, flask containing 25 mL of SP92 media was inoculated with 5 mL of an overnight YPD culture (initial OD = 0.4) and incubated for 24h at 30°C, with shaking at 300 rpm (2" throw incubator). The cells were centrifuged and the cell pellet

resuspended in 12.5 mL of TB-lowN media (yeast nitrogen base without amino acids and without ammonium sulfate, 1.7 g/L; yeast extract, 3.0 g/L; potassium phosphate monobasic, 1.0 g/L; potassium phosphate dibasic, 1.0 g/L) and was transferred to 250 mL glass bottom-baffled flasks. Either 0.281 mL of oleic acid or linoleic acid was added to start adipic acid production and the cultures were incubated at 30°C, with shaking at 300 rpm. Samples were taken after 48 hours for gas chromatographic (GC) analysis. Shown below are the diacid profiles produced from oleic acid and linoleic acid with strains sAA875 (ECI1^{+} ECI2^{+}) and sAA2220 (ecil^{-} ECI2^{+}). In the case of linoleic acid, all unsaturated diacid products, whether having one or two double bonds, are combined and listed as the monounsaturated product. The results are shown in TABLE 51 AND TABLE 52 below; "g/L" refers to grams of the indicated di-acid per liter of fermentation broth.

TABLE 51

	Diacid Profile, Linoleic acid Bioconversion (g/L)					
	C12:1 DA	C10:1 DA	C8:1 DA	C6:1 DA	C8:0 DA	C6:0 DA
ECI1+ ECI2+	0.00	0.78	2.26	0.00	2.88	5.82
eci1- ECI2+	6.17	2.18	0.47	0.00	0.38	0.09

TABLE 52

	Diacid Profile, Oleic acid Bioconversion (g/L)					
	C12:1 DA	C10:1 DA	C8:1 DA	C6:1 DA	C8:0 DA	C6:0 DA
ECI1+ ECI2+	0.00	0.11	0.46	0.00	4.88	7.77
eci1- ECI2+	1.23	2.44	3.61	0.00	0.66	0.16

Example 45: Construction of Strains sAA2428 and sAA1269,

Construction of sAA1269

Candida strain sAA988 was grown overnight in YPD and plated on 5-FOA plate. Colonies that grew in the presence of 5-FOA were PCR screened for the looping out of the URA3 gene leaving behind only the URA3 terminator. One such strain was named sAA1269.

Construction of Strain sAA2428

Candida strain SAA2428 was constructed by knocking out both alleles of POX5 from the genome of the starting strain sAA886. Both POX5 alleles were deleted by transforming the strain with the knock out plasmid construct pAA9 18.

	1st copy KO		after FOA		2nd copy KO		after FOA
sAA886 →	sAA2291	-→	sAA2310	-→	sAA2399	-→	sAA2428

Knock out plasmid construct pAA918: Approximately 600 bp of the 5' flanking region of the Candida POX5 gene was amplified using oligos oAA2656/oAA2657 and genomic DNA from ATCC20336. This fragment was gel purified and cloned into a PCR Blunt II TOPO vector to create plasmid pAA494. Approximately 500 bp of the 3' flanking region of the POX5 gene was amplified with primers oAA2658/oAA2659 using genomic DNA of ATCC20336 and this fragment was cloned into a PCR Blunt II TOPO vector to create plasmid pAA495. *EcoRIIBamHI* digested fragment of pAA494, *HindlllIBamHI* fragment of pAA495 and *HindlllIEcoRI* fragment of pUC19 were ligated together to construct plasmid pAA496. Subsequently, a DNA fragment containing the *URA3* terminator -*URA3* promoter - *URA3* coding sequence - *URA3* terminator flanked by *NotI* restriction sites was cloned into the *NotI* site of pAA496 to create the POX5 knockout construct plasmid pAA918. (FIG. 32).

TABLE 53 - Primer List

Primer	Sequence
oAA2656	AAGCTTTTAATTAAAGATAATCACAGGGGTAGAGACCTTG
oAA2657	GGATCCGCATGCGGCCGGCCGATAGCGTGGTATGAATGAATAAGTGTG
oAA2658	GGATCCGAGCTCGCGGCCGCGAGCACTAGGTTTTGATAATTTGGTTCTTAC
oAA2659	GAATTCTTAATTAACGGCGAAGAACATAGTGTGATG

Example 46: Generation of Strain sAA2291

Candida strain sAA886 was transformed with i¾c/-digested fragment of pAA918 and plated on SCD-Ura plate. Several colonies were screened by PCR for integration of the deletion cassette in the first *POX5* allele. A positive colony was named sAA2291 .

Example 47: Generation of Strain sAA2310

Strain sAA2291 was grown in YPD media overnight and plated on a 5-FOA plate. Colonies that grew in the presence of 5-FOA were PCR screened for the looping out of the *URA3* gene to leave behind only the *URA3* terminator (T_{URA3}) in the first *POX5* allele. This strain was named SAA2310.

Example 48: Generation of Strain sAA2399

The deletion of the second *POX5* allele was performed by transformation of i¾c/-digested fragment of pAA918 into sAA2310 and plating on a SCD-Ura plate. Colonies that grew on the SCD-Ura plate were PCR screened for integration of the knockout cassette in the second *POX5* allele. A positive colony was named sAA2399.

Example 49: Generation of Strain sAA2428

Strain sAA2399 was grown in YPD media overnight and plated on a 5-FOA plate. Colonies that grew in the presence of 5-FOA were PCR screened for the looping out of the *URA3* gene to leave behind only the *URA3* terminator (T_{URA3}) in the second *POX5* allele. This strain was named sAA2428.

Example 50: Construction of ATCC20336 POX4 and POX5 mutants in Strain sAA1269 or sAA2428

Mutants of ATCC20336 POX4 and POX5 had previously been constructed for expression in *E. coli (as described above).* To test these mutants for activity when expressed in strains derived from ATCC20336, they were cloned by PCR and inserted into either the vector pAA073 or pAA335.

The pAA335 vector was generated as follows. The promoter and terminator of PEX1 1 were PCR amplified from genomic DNA obtained from Candida strain ATCC20336 using oligos oAA2164 and oAA2138 or oAA2135 and oAA2136, respectively. The two pieces were digested with either PstI and SmaI or NdeI and SmaI and ligated into the PstI and NdeI sites of pAA61 to form pAA335.

If a gene is cloned into pAA073 or pAA335, it is under the regulation of the POX4 or PEX1 1 promoter, respectively. The oligonucleotides oAA540 and oAA541 were used to amplify POX5 sequences for seamless cloning into pAA073. The oligonucleotides oAA2984 and oAA2985 were used to amplify POX5 sequences for seamless cloning into pAA335. The oligonucleotides oAA3452 and oAA3453 were used to amplify POX4 sequences for seamless cloning into pAA335. The coding sequence for wildtype POX4 was amplified from genomic DNA of ATCC20336. All POX5 mutants except for T287A were transformed into sAA2428. The construct for the POX5 T287A mutant was transformed into sAA1269 to produce the strain sAA2058. The wildtype POX4 construct was transformed into sAA2428 to produce the strains SAA2778-2780 while the mutant POX4 D96A construct was transformed into sAA1269 to produce the strain sAA2109. All PCR amplifications were performed using Pfu Ultra II DNA Polymerase (Agilent Technologies, Santa Clara, CA) according to the manufacturer's instructions. Clones were selected by growth on SCD-ura plates and subsequently tested for growth on minimal media containing 0.1% Tween 80 with 0.1% oleic acid. Clones that grew well on media containing Tween 80 and oleic acid were tested in shake flask fermentations. In the case of the T287A mutant of POX5 and the D96A mutant of POX4, clones to be tested were identified by qPCR to determine how many copies of the mutant gene had been integrated. qPCR was performed using Brilliant III Ultra-Fast SYBR Green qPCR Master Mix (Agilent Technologies, Santa Clara, CA) according to the manufacturer's instructions. The oligonucleotides oAA2718 and oAA2719 were used to amplify a portion of the actin gene while the oligonucleotides oAA3515 and oAA3516 were used to amplify a portion of the PEX1 1 promoter as the means of quantitating gene copy number. The tables below shows the sequences of oligonucleotides used for the above cloning procedures and qPCRs and the resulting strains from integration of the constructs by transformation.

TABLE 54 - Cloning Primers List

Primer	Primer Sequence (5'-3')	Restriction Site(s)
oAA540	CACACAGCTCTTCCATAATGCCTACCGAACTTCAAAAAGAAAG	BspQl
oAA541	CACACAGCTCTTCCCTCTCTTCTATTCTTAACTGGACAAGATTTCAGCAGC	BspQl
oAA2984	CACACAGCTCTTCAGCCATGCCTACCGAACTTCAAAAAGAAAGAGAACTC	BspQl
oAA2985	CACACAGCTCTTCGAGCTTACAATTTAGAACTGGACAAGATTTCAGCAGCTTCTTCG	BspQl
oAA3452	CACACAGCTCTTCAGCCATGACTTTTACAAAGAAAAACGTTAGTGTATCACAAGG	BspQl
oAA3453	CACACAGCTCTTCGAGCTTACTTGGACAAGATAGCAG	BspQl
oAA2135	GAGCCCGGGTCTAGATGTGTGCTCTTCCGCTCCAGGCTTGTTATGACTCTAGAGAGAAGTGTG	SmaI
oAA2136	CACACACATATGAATTCGGTCGGGTTTTGACCTTGGATATGAAACTCAAAAATCATCAAATT	NdeI
oAA2138	CACACATCTAGACCCGGGCTCTTCTGGCTGCGTTGTGTATGGGTT	SmaI
oAA2164	CACACACTGCAGGAGGATGAAGAAGACGAAGA	PstI

TABLE 55 - qPCR Primer List

Primer	Primer Sequence (5'-3')	Restriction Site(s)
oAA2718	AGAAGCTTTGTTCAGACCAGCCGA	n/a
oAA2719	ACCACCGGACATGACAATGTTACC	n/a
oAA3515	GCAGAGTTAAGCCCGAGAAAGCAA	n/a

oAA3516	TTGCCTTCCTCTATTCGGCTACCA	n/a

TABLE 56 - POX5 Mutants

Mutant	Position(s)	Native Amino Acid(s)	Introduced Mutation(s)	Primer Combination	Resulting Strain
O	88	S	A	oAA540, 541	sAA2645
I	284, 436	G, E	E, G	oAA540, 541	sAA2646
K	292	S	A	oAA540, 541	sAA2648
E	95, 96	P, Q	A, A	oAA540, 541	sAA2651
P	98	F	G	oAA540, 541	sAA2570
G	287	T	A	oAA2984, 2985	sAA2058

TABLE 57 – POX4 Mutants

Mutant	Position(s)	Native Amino Acid(s)	Introduced Mutation(s)	Primer Combination	Resulting Strain
C	96	D	A	oAA3452, 3453	sAA2109

Example 51: Mutagenic target sites for *Y. lipolytica* Pox2, Pox3 and Pox5

Structural models of *Y. lipolytica* Pox2, Pox3 and Pox5, based on the crystal structure of *R. norvegicus* Acoll (PDB ID: 1IS2A), were obtained from MODBASE (ModBase: Database of Comparative Protein Structure Models - http://modbase.compbio.ucsf.edu/modbase-cgi/index.cgi) and used for HotSpot Wizard analyses of these enzymes (Hotspot Wizard - loschmidt.chemi.muni.cz/hotspotwizard/). The MODBASE identification numbers for the structural models of *Y. lipolytica* Pox2, Pox3 and Pox5 (in addition to their UniProtKB identification numbers) are shown in TABLE 58 below:

TABLE 58

Y. lipolytica acyl-CoA oxidase	MODBASE Database ID	UniProt Knowledgebase ID
Pox2	TR O74935	O74935
Pox3	TR O74936	O74936
Pox5	TR O74938	O74938

All three of the *Y. lipolytica* enzymes contain similar "hotspots" for mutagenesis as those found in both POX4 and POX5 of ATCC20336. TABLE 59 below shows the hotspots for POX4 and POX5 of ATCC20336 and the corresponding residues in Pox2, Pox3 and Pox5 of *Y. lipolytica* based on the multiple sequence alignment shown above. An amino acid at one or more positions identified in TABLE 59 can be substituted with a non-native amino acid for the purpose of modifying substrate specificity.

TABLE 59

ATCC20336 POX5		YIAco2		YIAco3		YIAco5	
Position	Residue	Position	Residue	Position	Residue	Position	Residue
81	D	101	D	102	K	100	K
82	Q	102	D	103	E	101	E
83	Q	103	I	104	A	102	A
84	A	104	K	105	L	103	I
85	Q	105	K	106	L	104	L
86	R	106	R	107	R	105	R
88	S	108	A	109	Q	107	Q
93	F	113	A	114	V	112	V
94	D	114	D	115	D	113	D
95	P	115	M	116	M	114	M
96	Q	116	G	117	G	115	G
98	F	118	Y	119	R	117	R
102	G	122	G	123	G	121	A
284	G	304	G	305	G	303	G
287	T	307	S	308	T	306	T
290	M	310	S	311	A	309	A
291	D	311	D	312	D	310	E
292	S	312	S	313	S	311	S
294	R	314	Q	315	F	313	L
295	M	315	V	316	V	314	L
428	D	446	D	447	D	446	D
429	W	447	W	448	W	447	W
436	E	454	E	455	E	454	E
453	D	471	S	472	S	471	S
454	L	472	A	473	C	472	C
455	L	473	V	474	L	473	I
456	K	474	A	475	G	474	A
457	E	475	L	476	H	475	M
458	P	476	R	477	R	476	R
459	E	477	K	478	K	477	K
460	Q	478	G	479	G	478	K
461	K	494	E	494	K	495	K
462	G	495	G	495	G	496	G
463	L	496	N	496	S	497	K

ATCC20336 POX4		YIAco2		YIAco3		YIAco5	
Position	Residue	Position	Residue	Position	Residue	Position	Residue
88	R	93	L	94	G	92	R
90	R	95	V	96	W	94	L
96	D	101	D	102	K	100	K
98	F	103	I	104	A	102	A
99	N	104	K	105	L	103	I
100	K	105	K	106	L	104	L
102	L	107	F	108	S	106	G
103	S	108	A	109	Q	107	Q
105	I	110	I	111	H	109	H
106	G	111	G	112	G	110	G
107	I	112	I	113	I	111	I
302	M	307	S	308	T	306	T
309	R	314	Q	315	F	313	L
310	M	315	V	316	V	314	L
374	D	373	A	374	I	372	Q
377	H	376	S	377	D	375	N
474	V	472	A	473	C	472	C
475	I	473	V	474	L	473	I
476	S	474	A	475	G	474	A
477	I	475	L	476	H	475	M
478	E	476	R	477	R	476	R
479	D	477	K	478	K	477	K
480	A	478	G	479	G	478	K
481	G	479	E	480	K	479	K
482	K					480	G
483	T	480	P	481	P	481	H
484	V	481	V	482	L	482	V
485	R	482	G	483	G	483	G
486	G	483	N	484	S	484	K
487	S	484	A	485	S	485	S
488	T	485	V	486	V	486	V
489	A	486	S	487	G	487	E
490	F	487	Y	488	Y	488	Y
491	L	488	L	489	L	489	L
492	N	489	K	490	A	490	Q
493	Q	490	R	491	N	491	R
494	L	491	Y			492	R
495	K	492	K	492	K	493	D
496	D	493	D	493	G	494	E
497	Y	494	L	494	L	495	L
498	T	495	A	495	E	496	Q
499	G	496	N	496	Q	497	N
500	S	497	A	497	A	498	A
501	N	498	K	498	T	499	R
502	S	499	L	499	L	500	V
503	S	500	N	500	S	501	D
504	K	501	G	501	G	502	N
505	V						

Example 52: Functional Analyses of ATCC20336 POX4 and POX5 mutants

Functional analyses of "hotspot" residues in POX5 and one residue in POX4 were performed by primarily substituting alanine or glycine at those positions to test if they changed the activity or substrate specificity of the enzyme. The (G284E, E436G) mutant is based on the ACAD mutagenesis studies previously described. The resulting mutants were cloned into either pAA073 or pAA335 and integrated into sAA2428. These clones were tested in shake flask fermentations of oleic acid as follows over a 48- or 72-hour period. Starter cultures (5mL) of the strains to be tested were grown in SP92 glycerol media (6.7g/L Difco yeast nitrogen base, 3.0g/L Difco yeast extract, 3.0g/L ammonium sulfate, l.Og/L potassium phosphate monobasic, l.Og/L potassium phosphate dibasic, 75g/L glycerol) were incubated overnight at 30°C, 250 rpm and used to inoculate 25mL fresh SP92 glycerol media to an initial OD600nm of 0.4 and incubated for approximately 18 hours at 30°C, 300 rpm. Cells were then pelleted by centrifugation for 10 minutes at 4,000xg, 4°C then resuspended in 12.5mL of TB-lowN media (1.7g/L Difco yeast nitrogen base without amino acids and ammonium sulfate, 3.0g/L Difco yeast extract, l.Og/L potassium phosphate monobasic, l.Og/L potassium phosphate dibasic). Oleic acid (562 µï) was added to start diacid production. TABLE 060 below shows the resulting diacids produced from these fermentations:

TABLE 60

		g/L Product (saturated + unsaturated diacid)				
Strain	POX5 mutation	C14 diacid	C12 diacid	C10 diacid	C8 diacid	C6 diacid
sAA2645	S88A	0.05	0.12	2.71	9.36	1.64
sAA2646	G284E, E436G	0.00	0.06	0.70	8.54	5.45
sAA2648	S292A	0.00	0.08	0.88	7.58	3.04
sAA2651	P95A, Q96A	0.00	0.05	0.95	7.72	3.35
sAA2570	F98G	0.00	0.00	0.00	0.16	12.02
sAA875	wt	0.00	0.04	0.14	4.24	7.48
sAA2058	T287A*		0.04	0.48	10.35	4.1
sAA875	wt*		0.02	0.17	4.84	6.69
Strain	POX4 mutation	C14 diacid	C12 diacid	C10 diacid	C8 diacid	C6 diacid
sAA2780	wildtype*	0.02	0.02	0.04	0.19	10.13
sAA2109	D96A*	0	0.01	0.01	0.12	9.01

^Results from a separate 48 hour fermentation

TABLE 61 - Description of some strains referenced herein.

Strain	**Genetic Modifications**
SAA496	poχ4Δ, CPR750, P450A19
SAA617	poχ4Δ, CPR750, acoataA/ACOATB, POX5
SAA620	poχ4Δ, CPR750, acoataA/ACOATB, POX5, ura3
SAA632	pox4A, CPR750, acoataA/ACOATB, P450A19
sAA635	pox4A, CPR750, acoataA/ACOATB, P450A19
sAA722	pox4A, acsl A/ACS1
sAA741	pox4A, acsl A/ACS1, ura3
sAA776	pox4A, acsl A
sAA779	pox4A, acsl A, ura3
sAA81 1	pox4A, acslA, P450A19
SAA810	pox4A, acslA, P450A19, EcTESA
sAA865	pox4A, acsl A, fat1A/FAT1
sAA869	pox4A, acsl A, fat1A/FAT1, ura3
sAA875	pox4A, acsl A, fatlA
SAA886	pox4A, acsl A, fatlA, ura3
SAA1764	pox4A, acsl A, fatlA, ecilA/ECIl

SAA1860	poχ4Δ, acslA, fatlA, ecilA/ECIl, ura3
SAA2058	poχ4Δ, poχ5Δ, acslA, fatlA, POX5(T287A)
SAA2109	poχ4Δ, poχ5Δ, acslA, fatlA, POX4(D96A)
SAA2220	poχ4Δ, acslA, fatlA, ecilA
SAA2291	poχ4Δ, poχ5Δ/POX 5, acslA, fatlA
SAA2310	poχ4Δ, poχ5Δ/POX 5, acslA, fatlA, ura3
SAA2399	poχ4Δ, poχ5Δ, acslA, fatlA
SAA2428	poχ4Δ, poχ5Δ, acslA, fatlA, ura3
SAA2570	poχ4Δ, poχ5Δ, acslA, fatlA, POX5(F98G)
SAA2645	poχ4Δ, poχ5Δ, acslA, fatlA, POX5(S88A)
SAA2646	poχ4Δ, poχ5Δ, acslA, fatlA, POX5(G284E)
SAA2648	poχ4Δ, poχ5Δ, acslA, fatlA, POX5(S292A)
SAA2651	poχ4Δ, poχ5Δ, acslA, fatlA, POX5(P95A,Q96A)
**sAA2780	poχ4Δ, poχ5Δ, acslA, fatlA, P0X4

Note: Genes in lower case and/or with a Δ symbol indicate a deleted gene. Strains sAA617, sAA620 and sAA632 comprise a deletion of one allele of "acoat" and are heterozygous for the acoat gene knock out (e.g., acetoacetyl CoA thiolase $^{-/+}$).

**Strain AA2780 comprises a deletion of the endogenous POX4 gene (i.e. poχ4Δ) and re-introduction of the wild type POX4 gene (i.e., POX4) under the control of a PEX1 1 promoter.

Example 53: Nucleotide and Amino Acid Sequences Usedfor Manipulations described herein

SEQ ID NO:	**Description**	**Sequence**
SEQ ID NO:l	Thioesterase activity Cuphea lanceolata Amino acid (A.A. Seq)	MVAAAATSAFFPVPAPGTSPKPGKSGNWPSSLSPTFKPKSIPN AGFQVKANASAHPKANGSAVNLKSGSLNTQEDTSSSPPPRAF LNQLPDWSMLLTAITTVFVAAEKQWTMLDRKSKRPDMLVDS VGLKSIVRDGLVSRQSFLIRSYEIGADRTASIETLMNHLQETSI NHCKSLGLLNDGFGRTPGMCKNDLIWVLTKMQIMVNRYPT WGDTVEINTWFSQSGKIGMASDWLISDCNTGEILIRATSVWA MMNQKTRRFSRLPYEVRQELTPHFVDSPHVIEDNDQKLHKFD VKTGDSIRKGLTPRWNDLDVNQHVSNVKYIGWILESMPIEVL

SEQ ID NO:	Description	Sequence
		ETQELCSLTVEYRRECGMDSVLESVTAVDPSENGGRSQYKHL LRLEDGTDIVKSRTEWRPKNAGTNGAISTSTAKTSNGNSAS
SEQ ID NO:2	FAO-13 (fatty alcohol oxidase activity) C. Tropicalis Nucleotide (Nuc. Seq)	atggctccatttttgcccgaccaggtcgactacaaacacgtcgacacccttatgttattatgtgacgg gatcatccacgaaaccaccgtcgaccaaatcaaagacgttattgctcctgacttccctgctgacaa gtacgaagagtacgtcaggacattcaccaaaccctccgaaaccccagggttcagggaaaccgtc tacaacacagtcaacgcaaacaccacggacgcaatccaccagttcattatcttgaccaatgttttgg catccagggtcttggctccagctttgaccaactcgttgacgcctatcaaggacatgagcttggaag accgtgaaaaattgttggcctcgtggcgcgactccccaatcgctgccaaaaggaagttgttcaggt tggtttctacgcttaccttggtcacgttcacgagattggccaatgagttgcatttgaaagccattcatta tccaggaagagaagaccgtgaaaaggcttatgaaacccaggagattgacccttttaagtaccagtt tttggaaaaaccgaagttttacggcgctgagttgtacttgccagatattgatgtgatcattattggatct ggtgccggtgctggtgttgtggcccacactttggccaacgatggcttcaagagtttggttttggaaa agggcaaatactttagcaactccgagttgaactttgatgacaaggacggcgttcaagaattatacca aagtggaggtactttgactacagtcaaccaacagttgtttgttcttgctggttccacttttggtggcggt accactgtcaattggtcagcctgtcttaagacgccattcaaggtgcgtaaggaatggtatgatgagt ttggtgttgactttgctgctgatgaagcatacgataaagcgcaggattatgtttggcagcaaatggg agcttctaccgaaggcatcacccactctttggctaacgagattattattgaaggtggtaagaaattag gttacaaggccaaggtattagaccaaaacagcggtggtcatcctcagcacagatgcggtttctgtt atttgggctgtaagcacggtatcaagcagggttctgttaataactggtttagagacgcagctgccca cggttcccagttcatgcaacaggttagagttttgcaaatacttaacaagaaggggatcgcttacggt atcttgtgtgaggatgttgtaaccggcgccaagttcaccattactggccccaaaaagtttgttgttgct gccggtgctttgaacactccatctgtgttggtcaactccggcttcaagaacaagaacatcggtaaga acttaactttgcacccagtttctgtcgtgtttggtgattttggcaaagacgttcaagcagaccacttcc acaactccatcatgactgccctttgttcagaagccgctgatttagacggcaagggccatggatgca gaattgaaaccatcttgaacgctccattcatccaggcttcattcttaccatggagaggtagtaacgag gctagacgagacttgttgcgttacaacaacatggtggcgatgttgctccttagtcgtgacaccacca gtggttccgtttctgctcatccaaccaaacctgaagctttggttgtcgagtacgacgtgaacaagttt gacagaaactcgatcttgcaggcattgttggtcactgctgacttgttgtatatccaaggtgccaaga gaatccttagtccacaggcatgggtgccaatttttgaatccgacaagccaaaggataagagatcaa tcaaggacgaggactatgtcgaatggagagccaaggttgccaagattcctttcgacacctacggct

SEQ ID NO:	Description	Sequence
		caccttatggttcggcacatcaaatgtcttcttgccgtatgtcaggtaagggtcctaaatacggtgct gttgacaccgatggtagattgtttgaatgttcgaatgtttatgttgccgatgcaagtcttttgccaactg caagcggtgccaaccctatggtcaccaccatgactcttgccagacatgttgcgttaggtttggcag actccttgaagaccaaagccaagttgtag
SEQ ID NO:3	FAO-13 (fatty alcohol oxidase activity) C. Tropicalis A.A. Seq	MAPFLPDQVDYKHVDTLMLLCDGIIHETTVDQIKDVIAPDFPA DKYEEYVRTFTKPSETPGFRETVYNTVNANTTDAIHQFIILTN VLASRVLAPALTNSLTPIKDMSLEDREKLLASWRDSPIAAKRK LFRLVSTLTLVTFTRLANELHLKAIHYPGREDREKAYETQEID PFKYQFLEKPKFYGAELYLPDIDVIIIGSGAGAGVVAHTLAND GFKSLVLEKGKYFSNSELNFDDKDGVQELYQSGGTLTTVNQQ LFVLAGSTFGGGTTVNWSACLKTPFKVRKEWYDEFGVDFAA DEAYDKAQDYVWQQMGASTEGITHSLANEIIIEGGKKLGYKA KVLDQNSGGHPQHRCGFCYLGCKHGIKQGSVNNWFRDAAA HGSQFMQQVRVLQILNKKGIAYGILCEDVVTGAKFTITGPKKF VVAAGALNTPSVLVNSGFKNKNIGKNLTLHPVSVVFGDFGKD VQADHFHNSIMTALCSEAADLDGKGHGCRIETILNAPFIQASF LPWRGSNEARRDLLRYNNMVAMLLLSRDTTSGSVSAHPTKP EALVVEYDVNKFDRNSILQALLVTADLLYIQGAKRILSPQAW VPIFESDKPKDKRSIKDEDYVEWRAKVAKIPFDTYGSPYGSAH QMSSCRMSGKGPKYGAVDTDGRLFECSNVYVADASLLPTAS GANPMVTTMTLARHVALGLADSLKTKAKL
SEQ ID NO:4	FAO-17(fatty alcohol oxidase activity) C. Tropicalis Nuc. Seq	atggctccatttttgcccgaccaggtcgactacaaacacgtcgacacccttatgttattatgtgacgg gatcatccacgaaaccaccgtggacgaaatcaaagacgtcattgcccctgacttccccgccgaca aatacgaggagtacgtcaggacattcaccaaaccctccgaaaccccagggttcagggaaaccgt ctacaacaccgtcaacgcaaacaccatggatgcaatccaccagttcattatcttgaccaatgttttgg gatcaagggtcttggcaccagctttgaccaactcgttgactcctatcaaggacatgagcttggaag accgtgaaaagttgttagcctcgtggcgtgactcccctattgctgctaaaaggaagttgttcaggttg gtttctacgcttaccttggtcacgttcacgagattggccaatgagttgcatttgaaagccattcattatc caggaagagaagaccgtgaaaaggcttatgaaacccaggagattgacccttttaagtaccagttttt ggaaaaaccgaagttttacggcgctgagttgtacttgccagatattgatgtgatcattattggatctg

SEQ ID NO:	Description	Sequence
		gtgccggtgctggtgttgtggcccacactttggccaacgatggcttcaagagtttggttttggaaaa gggcaaatactttagcaactccgagttgaactttgatgacaaggacggcgttcaagaattataccaa agtggaggtactttgactacagtcaaccaacagttgtttgttcttgctggttccacttttggtggcggta ccactgtcaattggtcagcctgtcttaagacgccattcaaggtgcgtaaggaatggtatgatgagttt ggtgttgactttgctgctgatgaagcatacgataaagcgcaggattatgtttggcagcaaatgggag cttctaccgaaggcatcacccactctttggctaacgagattattattgaaggtggtaagaaattaggtt acaaggccaaggtattagaccaaaacagcggtggtcatcctcagcacagatgcggtttctgttattt gggttgtaagcacggtatcaagcagggctctgttaataactggtttagagacgcagctgcccacgg ttctcagttcatgcaacaggttagagttttgcaaatccttaacaagaagggcatcgcttatggtatctt gtgtgaggatgttgtaaccggtgccaagttcaccattactggccccaaaaagtttgttgttgccgcc ggcgccttaaacactccatctgtgttggtcaactccggattcaagaacaagaacatcggtaagaac ttaactttgcatccagtttctgtcgtgtttggtgattttggcaaagacgttcaagcagaccacttccaca actccatcatgactgccctttgttcagaagccgctgatttagacggcaagggccatggatgcagaa ttgaaaccatcttgaacgctccattcatccaggcttcattcttaccatggagaggtagtaacgaggct agacgagacttgttgcgttacaacaacatggtggcgatgttgctccttagtcgtgacaccaccagtg gttccgtttctgctcatccaaccaaacctgaagctttggttgtcgagtacgacgtgaacaagtttgac agaaactcgatcttgcaggcattgttggtcactgctgacttgttgtatatccaaggtgccaagagaat ccttagtccacaggcatgggtgccaatttttgaatccgacaagccaaaggataagagatcaatcaa ggacgaggactatgtcgaatggagagccaaggttgccaagattcctttcgacacctacggctcac cttatggttcggcacatcaaatgtcttcttgccgtatgtcaggtaagggtcctaaatacggtgctgttg acaccgatggtagattgtttgaatgttcgaatgtttatgttgccgatgcaagtcttttgccaactgcaa gcggtgccaaccctatggtcaccaccatgactcttgcaagacatgttgcgttaggtttggcagactc cttgaagaccaaggccaagttgtag
SEQ ID NO:5	FAO-17(fatty alcohol oxidase activity) C. Tropicalis A.A. Seq	MAPFLPDQVDYKHVDTLMLLCDGIIHETTVDEIKDVIAPDFPA DKYEEYVRTFTKPSETPGFRETVYNTVNANTMDAIHQFIILTN VLGSRVLAPALTNSLTPIKDMSLEDREKLLASWRDSPIAAKPvK LFRLVSTLTLVTFTRLANELHLKAIHYPGREDREKAYETQEID PFKYQFLEKPKFYGAELYLPDIDVIIIGSGAGAGVVAHTLAND GFKSLVLEKGKYFSNSELNFDDKDGVQELYQSGGTLTTVNQQ LFVLAGSTFGGGTTVNWSACLKTPFKVRKEWYDEFGVDFAA

SEQ ID NO:	Description	Sequence
		DEAYDKAQDYVWQQMGASTEGITHSLANEIIIEGGKKLGYKA KVLDQNSGGHPQHRCGFCYLGCKHGIKQGSVNNWFRDAAA HGSQFMQQVRVLQILNKKGIAYGILCEDVVTGAKFTITGPKKF VVAAGALNTPSVLVNSGFKNKNIGKNLTLHPVSVVFGDFGKD VQADHFHNSIMTALCSEAADLDGKGHGCRIETILNAPFIQASF LPWRGSNEARRDLLRYNNMVAMLLLSRDTTSGSVSAHPTKP EALVVEYDVNKFDRNSILQALLVTADLLYIQGAKRILSPQAW VPIFESDKPKDKRSIKDEDYVEWRAKVAKIPFDTYGSPYGSAH QMSSCRMSGKGPKYGAVDTDGRLFECSNVYVADASLLPTAS GANPMVTTMTLARHVALGLADSLKTKAK1
SEQ ID NO:6	FAO-20(fatty alcohol oxidase activity) C. Tropicalis Nuc. Seq	atggctccatttttgcccgaccaggtcgactacaaacacgtcgacacccttatgttattatgtgacgg gatcatccacgaaaccaccgtcgaccaaatcaaagacgttattgctcctgacttccctgctgacaa gtacgaagagtacgtcaggacattcaccaaaccctccgaaaccccagggttcagggaaaccgtc tacaacacagtcaacgcaaacaccacggacgcaatccaccagttcattatcttgaccaatgttttgg catccagggtcttggctccagctttgaccaactcgttgacgcctatcaaggacatgagcttggaag accgtgaaaaattgttggcctcgtggcgcgactccccaatcgctgccaaaaggaaattgttcaggt tggtttccacgcttaccttggttactttcacgagattggccaatgagttgcatttgaaagccattcacta tccaggaagagaagaccgtgaaaaggcttatgaaacccaggagattgacccttttcaagtaccagt ttatggaaaagccaaagtttgacggcgctgagttgtacttgccagatattgatgttatcattattggat ctggtgccggtgctggtgttgtggcccacactttggccaacgatggcttcaagagtttggttttggaa aagggcaaatactttagcaactccgagttgaactttgatgacaaggacggcgttcaagaattatacc aaagtggaggtactttgactacagtcaaccaacagttgtttgttcttgctggttccacttttggtggcg gtaccactgtcaattggtcagcctgtcttaagacgccattcaaggtgcgtaaggaatggtatgatga gtttggtgttgactttgctgctgatgaagcatacgataaagcgcaggattatgtttggcagcaaatgg gagcttctaccgaaggcatcacccactctttggctaacgagattattattgaaggtggtaagaaatta ggttacaaggccaaggtattagaccaaaacagcggtggtcatcctcagcacagatgcggtttctgt tatttgggctgtaagcacggtatcaagcagggttctgttaataactggtttagagacgcagctgccc acggttcccagttcatgcaacaggttagagttttgcaaatacttaacaagaaggggatcgcttacgg tatcttgtgtgaggatgttgtaaccggcgccaagttcaccattactggccccaaaaagtttgttgttgc tgccggtgctttgaacactccatctgtgttggtcaactccggcttcaagaacaagaacatcggtaag

SEQ ID NO:	Description	Sequence
		aacttaactttgcacccagtttctgtcgtgtttggtgattttggcaaagacgttcaagcagaccacttc cacaactccatcatgactgccctttgttcagaagccgctgatttagacggcaagggccatggatgc agaattgaaaccatcttgaacgctccattcatccaggcttcattcttaccatggagaggtagtaacga ggctagacgagacttgttgcgttacaacaacatggtggcgatgttgctccttagtcgtgacaccacc agtggttccgtttctgctcatccaaccaaacctgaagctttggttgtcgagtacgacgtgaacaagtt tgacagaaactcgatcttgcaggcattgttggtcactgctgacttgttgtatatccaaggtgccaaga gaatccttagtccacaggcatgggtgccaatttttgaatccgacaagccaaaggataagagatcaa tcaaggacgaggactatgtcgaatggagagccaaggttgccaagattcctttcgacacctacggct caccttatggttcggcacatcaaatgtcttcttgccgtatgtcaggtaagggtcctaaatacggtgct gttgacaccgatggtagattgtttgaatgttcgaatgtttatgttgccgatgcaagtcttttgccaactg caagcggtgccaaccctatggtcaccaccatgactcttgccagacatgttgcgttaggtttggcag actccttgaagaccaaagccaagttgtag
SEQ ID NO:7	FAO-20(fatty alcohol oxidase activity) C. Tropicalis A.A. Seq	MAPFLPDQVDYKHVDTLMLLCDGIIHETTVDQIKDVIAPDFPA DKYEEYVRTFTKPSETPGFRETVYNTVNANTTDAIHQFIILTN VLASRVLAPALTNSLTPIKDMSLEDREKLLASWRDSPIAAKRK LFRLVSTLTLVTFTRLANELHLKAIHYPGREDREKAYETQEID PFKYQFMEKPKFDGAELYLPDIDVIIIGSGAGAGVVAHTLAND GFKSLVLEKGKYFSNSELNFDDKDGVQELYQSGGTLTTVNQQ LFVLAGSTFGGGTTVNWSACLKTPFKVRKEWYDEFGVDFAA DEAYDKAQDYVWQQMGASTEGITHSLANEIIIEGGKKLGYKA KVLDQNSGGHPQHRCGFCYLGCKHGIKQGSVNNWFRDAAA HGSQFMQQVRVLQILNKKGIAYGILCEDVVTGAKFTITGPKKF VVAAGALNTPSVLVNSGFKNKNIGKNLTLHPVSVVFGDFGKD VQADHFHNSIMTALCSEAADLDGKGHGCRIETILNAPFIQASF LPWRGSNEARRDLLRYNNMVAMLLLSRDTTSGSVSAHPTKP EALVVEYDVNKFDRNSILQALLVTADLLYIQGAKRILSPQAW VPIFESDKPKDKRSIKDEDYVEWRAKVAKIPFDTYGSPYGSAH QMSSCRMSGKGPKYGAVDTDGRLFECSNVYVADASLLPTAS GANPMVTTMTLARHVALGLADSLKTKAKL
SEQ	FAO-2a(fatty	atgaataccttcttgccagacgtgctcgaatacaaacacgtcgacacccttttgttattgtgtgacgg

SEQ ID NO:	Description	Sequence
ID NO:8	alcohol oxidase activity) C. Tropicalis Nuc. Seq	gatcatccacgaaaccacagtcgatcagatcaaggacgccattgctcccgacttccctgaggacc agtacgaggagtatctcaagaccttcaccaagccatctgagacccctgggttcagagaagccgtc tacgacacgatcaacgccaccccaaccgatgccgtgcacatgtgtattgtcttgaccaccgcattg gactccagaatcttggcccccacgttgaccaactcgttgacgcctatcaaggatatgaccttgaag gagcgtgaacaattgttggcctcttggcgtgattccccgattgcggcaaagagaagattgttcagat tgatttcctcgcttaccttgacgacgtttacgagattggccagcgaattgcacttgaaagccatccac taccctggcagagacttgcgtgaaaaggcgtatgaaacccaggtggttgacccttcaggtacctg tttatggagaaaccaaagtttgacggcgccgaattgtacttgccagatatcgacgtcatcatcattgg atcaggcgccggtgctggtgtcatggcccacactctcgccaacgacgggttcaagaccttggtttt ggaaaagggaaagtatttcagcaactccgagttgaactttaatgacgctgatggcgtgaaagagtt gtaccaaggtaaaggtgctttggccaccaccaatcagcagatgtttattcttgccggttccactttgg gcggtggtaccactgtcaactggtctgcttgccttaaaacaccatttaaagtgcgtaaggagtggta cgacgagtttggtcttgaatttgctgccgatgaagcctacgacaaagcgcaggattatgtttggaaa caaatgggtgcttcaacagatggaatcactcactccttggccaacgaagttgtggttgaaggaggt aagaagttgggctacaagagcaaggaaattgagcagaacaacggtggccaccctgaccaccca tgtggtttctgttacttgggctgtaagtacggtattaaacagggttctgtgaataactggtttagagac gcagctgcccacgggtccaagttcatgcaacaagtcagagttgtgcaaatcctcaacaagaatgg cgtcgcttatggtatcttgtgtgaggatgtcgaaaccggagtcaggttcactattagtggccccaaa aagtttgttgtttctgctggttctttgaacacgccaactgtgttgaccaactccggattcaagaacaag cacattggtaagaacttgacgttgcacccagtttccaccgtgtttggtgactttggcagagacgtgc aagccgaccatttccacaaatctattatgacttcgctttgttacgaggttgctgacttggacggcaag ggccacggatgcagaatcgaaaccatcttgaacgctccattcatccaagcttctttgttgccatgga gaggaagtgacgaggtcagaagagacttgttgcgttacaacaacatggtggccatgttgcttatca cgcgtgataccaccagtggttcagtttctgctgacccaaagaagcccgacgctttgattgtcgacta tgagattaacaagtttgacaagaatgccatcttgcaagctttcttgatcacttccgacatgttgtacatt gaaggtgccaagagaatcctcagtccacagccatgggtgccaatctttgagtcgaacaagccaaa ggagcaaagaacgatcaaggacaaggactatgttgagtggagagccaaggctgctaagatacct ttcgacacctacggttctgcatatgggtccgcacatcaaatgtccacctgtcgtatgtccggaaagg gtcctaaatacggtgctgttgatactgatggtagattgtttgaatgttcgaatgtctatgttgctgatgct agtgttttgcctactgccagcggtgccaacccaatgatatccaccatgacctttgctagacagattgc

SEQ ID NO:	Description	Sequence
		gttaggtttggctgactccttgaagaccaaacccaagttgtag
SEQ ID NO:9	FAO-2a(fatty alcohol oxidase activity) C. Tropicalis A.A. Seq	MNTFLPDVLEYKHVDTLLLLCDGIIHETTVDQIKDAIAPDFPE DQYEEYLKTFTKPSETPGFREAVYDTINATPTDAVHMCIVLTT ALDSRILAPTLTNSLTPIKDMTLKEREQLLASWRDSPIAAKRR LFRLISSLTLTTFTRLASELHLKAIHYPGRDLREKAYETQVVDP FRYSFMEKPKFDGAELYLPDIDVIIIGSGAGAGVMAHTLANDG FKTLVLEKGKYFSNSELNFNDADGVKELYQGKGALATTNQQ MFILAGSTLGGGTTVNWSACLKTPFKVRKEWYDEFGLEFAA DEAYDKAQDYVWKQMGASTDGITHSLANEVVVEGGKKLGY KSKEIEQNNGGHPDHPCGFCYLGCKYGIKQGSVNNWFRDAA AHGSKFMQQVRVVQILNKNGVAYGILCEDVETGVRFTISGPK KFVVSAGSLNTPTVLTNSGFKNKHIGKNLTLHPVSTVFGDFGR DVQADHFHKSIMTSLCYEVADLDGKGHGCRIETILNAPFIQAS LLPWRGSDEVRRDLLRYNNMVAMLLITRDTTSGSVSADPKKP DALIVDYEINKFDKNAILQAFLITSDMLYIEGAKRILSPQPWVP IFESNKPKEQRTIKDKDYVEWRAKAAKIPFDTYGSAYGSAHQ MSTCRMSGKGPKYGAVDTDGRLFECSNVYVADASVLPTASG ANPMISTMTFARQIALGLADSLKTKPKL
SEQ ID NO:l 0	FAO-2b(fatty alcohol oxidase activity) C. Tropicalis Nuc. Seq	atgaataccttcttgccagacgtgctcgaatacaaacacgtcgatacccttttgttattatgtgacggg atcatccacgaaaccacagtcgaccagatcagggacgccattgctcccgacttccctgaagacca gtacgaggagtatctcaagaccttcaccaagccatctgagacccctgggttcagagaagccgtct acgacacgatcaacagcaccccaaccgaggctgtgcacatgtgtattgtattgaccaccgcattgg actcgagaatcttggcccccacgttgaccaactcgttgacgcctatcaaggatatgaccttgaaag agcgtgaacaattgttggctgcctggcgtgattccccgatcgcggccaagagaagattgttcagat tgatttcctcacttaccttgacgacctttacgagattggccagcgacttgcacttgagagccatccac taccctggcagagacttgcgtgaaaaggcatatgaaacccaggtggttgacccttcaggtacctg tttatggaaaaaccaaagtttgacggcaccgagttgtacttgccagatatcgacgtcatcatcattgg atccggtgccggtgctggtgtcatggcccacactttagccaacgacgggtacaagaccttggtttt ggaaaagggaaagtatttcagcaactccgagttgaactttaatgatgccgatggtatgaaagagttg taccaaggtaaatgtgcgttgaccaccacgaaccagcagatgtttattcttgccggttccactttggg

SEQ ID NO:	Description	Sequence
		cggtggtaccactgttaactggtctgcttgtcttaaaacaccatttaaagtgcgtaaggagtggtacg acgagtttggtcttgaatttgctgccgacgaagcctacgacaaagcacaagactatgtttggaaaca aatgggcgcttctaccgaaggaatcactcactctttggcgaacgcggttgtggttgaaggaggtaa gaagttgggttacaagagcaaggaaatcgagcagaacaatggtggccatcctgaccacccctgt ggtttctgttacttgggctgtaagtacggtattaagcagggttctgtgaataactggtttagagacgca gctgcccacgggtccaagttcatgcaacaagtcagagttgtgcaaatcctccacaataaaggcgtc gcttatggcatcttgtgtgaggatgtcgagaccggagtcaaattcactatcagtggccccaaaaagt ttgttgtttctgcaggttctttgaacacgccaacggtgttgaccaactccggattcaagaacaaacac atcggtaagaacttgacgttgcacccagtttcgaccgtgtttggtgactttggcagagacgtgcaag ccgaccatttccacaaatctattatgacttcgctctgttacgaagtcgctgacttggacggcaaggg ccacggatgcagaatcgagaccatcttgaacgctccattcatccaagcttctttgttgccatggaga ggaagcgacgaggtcagaagagacttgttgcgttacaacaacatggtggccatgttgcttatcacc cgtgacaccaccagtggttcagtttctgctgacccaaagaagcccgacgctttgattgtcgactatg acatcaacaagtttgacaagaatgccatcttgcaagctttcttgatcacctccgacatgttgtacatcg aaggtgccaagagaatcctcagtccacaggcatgggtgccaatctttgagtcgaacaagccaaag gagcaaagaacaatcaaggacaaggactatgtcgaatggagagccaaggctgccaagatacctt tcgacacctacggttctgcctatgggtccgcacatcaaatgtccacctgtcgtatgtccggaaagg gtcctaaatacggcgccgttgataccgatggtagattgtttgaatgttcgaatgtctatgttgctgatg ctagtgttttgcctactgccagcggtgccaacccaatgatctccaccatgacgtttgctagacagatt gcgttaggtttggctgactctttgaagaccaaacccaagttgtag
SEQ ID NO:11	FAO-2b(fatty alcohol oxidase activity) C. Tropicalis A.A. Seq	MNTFLPDVLEYKHVDTLLLLCDGIIHETTVDQIRDAIAPDFPE DQYEEYLKTFTKPSETPGFREAVYDTINSTPTEAVHMCIVLTT ALDSRILAPTLTNSLTPIKDMTLKEREQLLAAWRDSPIAAKRR LFRLISSLTLTTFTRLASDLHLRAIHYPGRDLREKAYETQVVDP FRYSFMEKPKFDGTELYLPDIDVIIIGSGAGAGVMAHTLANDG YKTLVLEKGKYFSNSELNFNDADGMKELYQGKCALTTTNQQ MFILAGSTLGGGTTVNWSACLKTPFKVRKEWYDEFGLEFAA DEAYDKAQDYVWKQMGASTEGITHSLANAVVVEGGKKLGY KSKEIEQNNGGHPDHPCGFCYLGCKYGIKQGSVNNWFRDAA AHGSKFMQQVRVVQILHNKGVAYGILCEDVETGVKFTISGPK

SEQ ID NO:	Description	Sequence
		KFVVSAGSLNTPTVLTNSGFKNKHIGKNLTLHPVSTVFGDFGR DVQADHFHKSIMTSLCYEVADLDGKGHGCRIETILNAPFIQAS LLPWRGSDEVRRDLLRYNNMVAMLLITRDTTSGSVSADPKKP DALIVDYDINKFDKNAILQAFLITSDMLYIEGAKRILSPQAWV PIFESNKPKEQRTIKDKDYVEWRAKAAKIPFDTYGSAYGSAH QMST
SEQ ID NO:12	FAO-18(fatty alcohol oxidase activity) C. Tropicalis Nuc. Seq	atggctccatttttgcccgaccaggtcgactacaaacacgtcgacacccttatgttattatgtgacgg gatcatccacgaaaccaccgtggacgaaatcaaagacgtcattgcccctgacttccccgccgaca aatacgaggagtacgtcaggacattcaccaaaccctccgaaaccccagggttcagggaaaccgt ctacaacaccgtcaacgcaaacaccatggatgcaatccaccagttcattatcttgaccaatgttttgg gatcaagggtcttggcaccagctttgaccaactcgttgactcctatcaaggacatgagcttggaag accgtgaaaagttgttagcctcgtggcgtgactcccctattgctgctaaaaggaagttgttcaggttg gtttctacgcttaccttggtcacgttcacgagattggccaatgagttgcatttgaaagccattcattatc caggaagagaagaccgtgaaaaggcttatgaaacccaggagattgacccttttaagtaccagttttt ggaaaaaccgaagttttacggcgctgagttgtacttgccagatattgatgtgatcattattggatctg ggggccggtgctggtgtcgtggcccacactttgaccaacgacggcttcaagagtttggttttggaaa agggcagatactttagcaactccgagttgaactttgatgacaaggacggggttcaagaattatacc aaagtggaggtactttgaccaccgtcaaccagcagttgtttgttcttgctggttccacttttggtggtg gtaccactgtcaattggtcggcctgtcttaaaacgccattcaaggtgcgtaaggaatggtatgatga gtttggcgttgactttgctgccgatgaagcctacgacaaagcacaggattatgtttggcagcaaatg ggagcttctaccgaaggcatcacccactctttggctaacgagattattattgaaggtggcaagaaat taggttacaaggccaaggtattagaccaaaacagcggtggtcatcctcatcacagatgcggtttct gttatttgggttgtaagcacggtatcaagcagggctctgttaataactggtttagagacgcagctgcc cacggttctcagttcatgcaacaggttagagttttgcaaatccttaacaagaagggcatcgcttatgg tatcttgtgtgaggatgttgtaaccggtgccaagttcaccattactggccccaaaaagtttgttgttgc cgccggcgccttaaacactccatctgtgttggtcaactccggattcaagaacaagaacatcggtaa gaacttaactttgcatccagtttctgtcgtgtttggtgattttggcaaagacgttcaagcagatcacttc cacaactccatcatgactgctctttgttcagaagccgctgatttagacggcaagggtcatggatgca gaattgaaaccatcttgaacgctccattcatccaggcttcattcttaccatggagaggtagtaacgag gctagacgagacttgttgcgttacaacaacatggtggccatgttacttcttagtcgtgataccaccag

SEQ ID NO:	Description	Sequence
		tggttccgtttcgtcccatccaactaaacctgaagcattagttgtcgagtacgacgtgaacaagtttg acagaaactccatcttgcaggcattgttggtcactgctgacttgttgtacattcaaggtgccaagaga atccttagtccccaaccatgggtgccaatttttgaatccgacaagccaaaggataagagatcaatca aggacgaggactatgtcgaatggagagccaaggttgccaagattccttttgacacctacggctcg ccttatggttcggcgcatcaaatgtcttcttgtcgtatgtcaggtaagggtcctaaatacggtgctgtt gataccgatggtagattgtttgaatgttcgaatgtttatgttgctgacgctagtcttttgccaactgcta gcggtgctaatcctatggtcaccaccatgactcttgcaagacatgttgcgttaggtttggcagactcc ttgaagaccaaggccaagttgtag
SEQ ID NO:l 3	FAO-1(fatty alcohol oxidase activity) C. Tropicalis A.A. Seq	MAPFLPDQVDYKHVDTLMLLCDGIIHETTVDEIKDVIAPDFPA DKYEEYVRTFTKPSETPGFRETVYNTVNANTMDAIHQFIILTN VLGSRVLAPALTNSLTPIKDMSLEDREKLLASWRDSPIAAKRK LFRLVSTLTLVTFTRLANELHLKAIHYPGREDREKAYETQEID PFKYQFLEKPKFYGAELYLPDIDVIIIGSGAGAGVVAHTLTND GFKSLVLEKGRYFSNSELNFDDKDGVQELYQSGGTLTTVNQQ LFVLAGSTFGGGTTVNWSACLKTPFKVRKEWYDEFGVDFAA DEAYDKAQDYVWQQMGASTEGITHSLANEIIIEGGKKLGYKA KVLDQNSGGHPHHRCGFCYLGCKHGIKQGSVNNWFRDAAA HGSQFMQQVRVLQILNKKGIAYGILCEDVVTGAKFTITGPKKF VVAAGALNTPSVLVNSGFKNKNIGKNLTLHPVSVVFGDFGKD VQADHFHNSIMTALCSEAADLDGKGHGCRIETILNAPFIQASF LPWRGSNEARRDLLRYNNMVAMLLLSRDTTSGSVSSHPTKPE ALVVEYDVNKFDRNSILQALLVTADLLYIQGAKRILSPQPWV PIFESDKPKDKRSIKDEDYVEWRAKVAKIPFDTYGSPYGSAHQ MSSCRMSGKGPKYGAVDTDGRLFECSNVYVADASLLPTASG ANPMVTTMTLARHVALGLADSLKTKAKL
SEQ ID NO:l 4	cytochrome P450 A12 (CYP52A12) Nuc. Seq	atggccacacaagaaatcatcgattctgtacttccgtacttgaccaaatggtacactgtgattactgc agcagtattagtcttccttatctccacaaacatcaagaactacgtcaaggcaaagaaattgaaatgtg tcgatccaccatacttgaaggatgccggtctcactggtattctgtctttgatcgccgccatcaaggcc aagaacgacggtagattggctaactttgccgatgaagttttcgacgagtacccaaaccacaccttct acttgtctgttgccggtgctttgaagattgtcatgactgttgacccagaaaacatcaaggctgtcttgg

SEQ ID NO:	Description	Sequence
		ccacccaattcactgacttctccttgggtaccagacacgcccactttgctcctttgttgggtgacggt atcttcaccttggacggagaaggttggaagcactccagagctatgttgagaccacagtttgctaga gaccagattggacacgttaaagccttggaaccacacatccaaatcatggctaagcagatcaagttg aaccagggaaagactttcgatatccaagaattgttctttagatttaccgtcgacaccgctactgagtt cttgtttggtgaatccgttcactccttgtacgatgaaaaattgggcatcccaactccaaacgaaatcc caggaagagaaaactttgccgctgctttcaacgtttcccaacactacttggccaccagaagttactc ccagacttttttacttttgaccaaccctaaggaattcagagactgtaacgccaaggtccaccacttgg ccaagtactttgtcaacaaggccttgaactttactcctgaagaactcgaagagaaatccaagtccgg ttacgttttcttgtacgaattggttaagcaaaccagagatccaaaggtcttgcaagatcaattgttgaa cattatggttgccggaagagacaccactgccggtttgttgtcctttgctttgtttgaattggctagaca cccagagatgtggtccaagttgagagaagaaatcgaagttaactttggtgttggtgaagactcccg cgttgaagaaattaccttcgaagccttgaagagatgtgaatacttgaaggctatccttaacgaaacct tgcgtatgtacccatctgttcctgtcaactttagaaccgccaccagagacaccactttgccaagagg tggtggtgctaacggtaccgacccaatctacattcctaaaggctccactgttgcttacgttgtctaca agacccaccgtttggaagaatactacggtaaggacgctaacgacttcagaccagaaagatggttt gaaccatctactaagaagttgggctgggcttatgttccattcaacggtggtccaagagtctgcttgg gtcaacaattcgccttgactgaagcttcttatgtgatcactagattggcccagatgtttgaaactgtct catctgatccaggtctcgaataccctccaccaaagtgtattcacttgaccatgagtcacaacgatgg tgtctttgtcaagatgtaa
SEQ ID NO:15	cytochrome P450 A13 (CYP52A13) Nuc. Seq	atgactgtacacgatattatcgccacatacttcaccaaatggtacgtgatagtaccactcgctttgatt gcttatagagtcctcgactacttctatggcagatacttgatgtacaagcttggtgctaaaccatttttcc agaaacagacagacggctgtttcggattcaaagctccgcttgaattgttgaagaagaagagcgac ggtaccctcatagacttcacactccagcgtatccacgatctcgatcgtcccgatatcccaactttcac attcccggtcttttccatcaaccttgtcaataccttgagccggagaacatcaaggccatcttggcca ctcagttcaacgatttctccttgggtaccagacactcgcactttgctcctttgttgggtgatggtatctt acgttggatggcgccggctggaagcacagcagatctatgttgagaccacagtttgccagagaaca gatttcccacgtcaagttgttggagccacacgttcaggtgttcttcaaacacgtcagaaaggcacag ggcaagacttttgacatccaggaattgtttttcagattgaccgtcgactccgccaccgagtttttgtttg gtgaatccgttgagtccttgagagatgaatctatcggcatgtccatcaatgcgcttgactttgacggc aaggctggctttgctgatgcttttaactattcgcagaattatttggcttcgagagcggttatgcaacaa

SEQ ID NO:	Description	Sequence
		ttgtactgggtgttgaacgggaaaaagtttaaggagtgcaacgctaaagtgcacaagtttgctgact actacgtcaacaaggctttggacttgacgcctgaacaattggaaaagcaggatggttatgtgtttttg tacgaattggtcaagcaaaccagagacaagcaagtgttgagagaccaattgttgaacatcatggtt gctggtagagacaccaccgccggtttgttgtcgtttgttttctttgaattggccagaaacccagaagtt accaacaagttgagagaagaaattgaggacaagtttggactcggtgagaatgctagtgttgaaga catttcctttgagtcgttgaagtcctgtgaatacttgaaggctgttctcaacgaaaccttgagattgtac ccatccgtgccacagaatttcagagttgccaccaagaacactaccctcccaagaggtggtggtaa ggacgggttgtctcctgttttggtgagaaagggtcagaccgttatttacggtgtctacgcagcccac agaaacccagctgtttacggtaaggacgctcttgagtttagaccagagagatggtttgagccagag acaaagaagcttggctgggccttcctcccattcaacggtggtccaagaatctgtttgggacagcag tttgccttgacagaagcttcgtatgtcactgtcaggttgctccaggagtttgcacacttgtctatggac ccagacaccgaatatccacctaagaaaatgtcgcatttgaccatgtcgcttttcgacggtgccaata ttgagatgtattag
SEQ ID NO:l 6	cytochrome P450 A14 (CYP52A14) Nuc. Seq	atgactgcacaggatattatcgccacatacatcaccaaatggtacgtgatagtaccactcgctttgat tgcttatagggtcctcgactacttttacggcagatacttgatgtacaagcttggtgctaaaccgtttttc cagaaacaaacagacggttatttcggattcaaagctccacttgaattgttaaaaaagaagagtgac ggtaccctcatagacttcactctcgagcgtatccaagcgctcaatcgtccagatatcccaacttttac attcccaatcttttccatcaaccttatcagcacccttgagccggagaacatcaaggctatcttggcca cccagttcaacgatttctccttgggcaccagacactcgcactttgctcctttgttgggcgatggtatct ttaccttggacggtgccggctggaagcacagcagatctatgttgagaccacagtttgccagagaa cagatttcccacgtcaagttgttggagccacacatgcaggtgttcttcaagcacgtcagaaaggca cagggcaagacttttgacatccaagaattgtttttcagattgaccgtcgactccgccactgagttttg tttggtgaatccgttgagtccttgagagatgaatctattgggatgtccatcaatgcacttgactttgac ggcaaggctggctttgctgatgcttttaactactcgcagaactatttggcttcgagagcggttatgca acaattgtactgggtgttgaacgggaaaaagtttaaggagtgcaacgctaaagtgcacaagtttgct gactattacgtcagcaaggctttggacttgacacctgaacaattggaaaagcaggatggttatgtgt tcttgtacgagttggtcaagcaaaccagagacaggcaagtgttgagagaccagttgttgaacatca tggttgccggtagagacaccaccgccggtttgttgtcgtttgttttctttgaattggccagaaaccca gaggtgaccaacaagttgagagaagaaatcgaggacaagtttggtcttggtgagaatgctcgtgtt gaagacatttcctttgagtcgttgaagtcatgtgaatacttgaaggctgttctcaacgaaactttgaga

SEQ ID NO:	Description	Sequence
		ttgtacccatccgtgccacagaatttcagagttgccaccaaaaacactaccttccaaggggaggt ggtaaggacgggttatctcctgttttggtcagaaagggtcaaaccgttatgtacggtgtctacgctg cccacagaaacccagctgtctacggtaaggacgcccttgagtttagaccagagaggtggtttgag ccagagacaaagaagcttggctgggccttccttccattcaacggtggtccaagaatttgcttggga cagcagtttgccttgacagaagcttcgtatgtcactgtcagattgctccaagagtttggacacttgtct atggaccccaacaccgaatatccacctaggaaaatgtcgcatttgaccatgtccctttcgacggtg ccaacattgagatgtattag
SEQ ID NO:17	cytochrome P450 A15 (CYP52A15) Nuc. Seq	atgtcgtcttctccatcgtttgcccaagaggttctcgctaccactagtccttacatcgagtactttcttg acaactacaccagatggtactacttcatacctttggtgcttctttcgttgaactttataagtttgctccac acaaggtacttggaacgcaggttccacgccaagccactcggtaactttgtcagggaccctacgttt ggtatcgctactccgttgcttttgatctacttgaagtcgaaaggtacggtcatgaagtttgcttggggc ctctggaacaacaagtacatcgtcagagacccaaagtacaagacaactgggctcaggattgttgg cctcccattgattgaaaccatggacccagagaacatcaaggctgttttggctactcagttcaatgatt tctctttgggaaccagacacgatttcttgtactccttgttgggtgacggtattttcaccttggacggtgc tggctggaaacatagtagaactatgttgagaccacagtttgctagagaacaggtttctcacgtcaag ttgttggagccacacgttcaggtgttcttcaagcacgttagaaagcaccgcggtcaaacgttcgac atccaagaattgttcttcaggttgaccgtcgactccgccaccgagttcttgtttggtgagtctgctgaa tccttgagggacgaatctattggattgaccccaaccaccaaggatttcgatggcagaagagatttc gctgacgctttcaactattcgcagacttaccaggcctacagatttttgttgcaacaaatgtactggatc ttgaatggctcggaattcagaaagtcgattgctgtcgtgcacaagtttgctgaccactatgtgcaaaa ggctttggagttgaccgacgatgacttgcagaaacaagacggctatgtgttcttgtacgagttggct aagcaaaccagagacccaaaggtcttgagagaccagttattgaacattttggttgccggtagagac acgaccgccggtttgttgtcatttgttttctacgagttgtcaagaaaccctgaggtgtttgctaagttga gagaggaggtggaaaacagatttggactcggtgaagaagctcgtgttgaagagatctcgtttgagt ccttgaagtcttgtgagtacttgaaggctgtcatcaatgaaaccttgagattgtacccatcggttccac acaactttagagttgctaccagaaacactaccctcccaagaggtggtggtgaagatggatactcgc caattgtcgtcaagaagggtcaagttgtcatgtacactgttattgctacccacagagacccaagtatc tacggtgccgacgctgacgtcttcagaccagaaagatggtttgaaccagaaactagaaagttggg ctgggcatacgttccattcaatggtggtccaagaatctgtttgggtcaacagtttgccttgaccgaag cttcatacgtcactgtcagattgctccaggagtttgcacacttgtctatggacccagacaccgaatat

SEQ ID NO:	Description	Sequence
		ccaccaaaattgcagaacaccttgaccttgtcgctctttgatggtgctgatgttagaatgtactaa
SEQ ID NO:18	cytochrome P450 A16 (CYP52A16) Nuc. Seq	atgtcgtcttctccatcgtttgctcaggaggttctcgctaccactagtccttacatcgagtactttcttga caactacaccagatggtactacttcatccctttggtgcttctttcgttgaacttcatcagcttgctccac acaaagtacttggaacgcaggttccacgccaagccgctcggtaacgtcgtgttggatcctacgttt ggtatcgctactccgttgatcttgatctacttaaagtcgaaaggtacagtcatgaagtttgcctggag cttctggaacaacaagtacattgtcaaagacccaaagtacaagaccactggccttagaattgtcgg cctcccattgattgaaaccatagacccagagaacatcaaagctgtgttggctactcagttcaacgat ttctccttgggaactagacacgatttcttgtactccttgttgggcgatggtatttttaccttggacggtg ctggctggaaacacagtagaactatgttgagaccacagtttgctagagaacaggtttcccacgtca agttgttggaaccacacgttcaggtgttcttcaagcacgttagaaaacaccgcggtcagacttttga catccaagaattgttcttcagattgaccgtcgactccgccaccgagttcttgtttggtgagtctgctga atccttgagagacgactctgttggtttgaccccaaccaccaaggatttcgaaggcagaggagattt cgctgacgctttcaactactcgcagacttaccaggcctacagatttttgttgcaacaaatgtactgga ttttgaatggcgcggaattcagaaagtcgattgccatcgtgcacaagtttgctgaccactatgtgcaa aaggctttggagttgaccgacgatgacttgcagaaacaagacggctatgtgttcttgtacgagttgg ctaagcaaactagagacccaaaggtcttgagagaccagttgttgaacattttggttgccggtagag acacgaccgccggtttgttgtcgtttgtgttctacgagttgtcgagaaaccctgaagtgtttgccaag ttgagagaggaggtggaaaacagatttggactcggcgaagaggctcgtgttgaagagatctctttt gagtccttgaagtcctgtgagtacttgaaggctgtcatcaatgaagccttgagattgtacccatctgtt ccacacaacttcagagttgccaccagaaacactacccttccaagaggcggtggtaaagacggat gctcgccaattgttgtcaagaagggtcaagttgtcatgtacactgtcattggtacccacagagaccc aagtatctacggtgccgacgccgacgtcttcagaccagaaagatggttcgagccagaaactagaa agttgggctgggcatatgttccattcaatggtggtccaagaatctgtttgggtcagcagtttgccttga ctgaagcttcatacgtcactgtcagattgctccaagagtttggaaacttgtccctggatccaaacgct gagtacccaccaaaattgcagaacaccttgaccttgtcactctttgatggtgctgacgttagaatgtt ctaa
SEQ ID NO:19	cytochrome P450 A17 (CYP52A17) Nuc. Seq	atgattgaacaactcctagaatattggtatgtcgttgtgccagtgttgtacatcatcaaacaactccttg catacacaaagactcgcgtcttgatgaaaaagttgggtgctgctccagtcacaaacaagttgtacg acaacgctttcggtatcgtcaatggatggaaggctctccagttcaagaaagagggcagggctcaa gagtacaacgattacaagtttgaccactccaagaacccaagcgtgggcacctacgtcagtattcttt

SEQ ID NO:	Description	Sequence
		tcggcaccaggatcgtcgtgaccaaagatccagagaatatcaaagctattttggcaacccagtttg gtgatttttctttgggcaagaggcacactctttttaagcctttgttaggtgatgggatcttcacattgga cggcgaaggctggaagcacagcagagccatgttgagaccacagtttgccagagaacaagttgct catgtgacgtcgttggaaccacacttccagttgttgaagaagcatattcttaagcacaagggtgaat actttgatatccaggaattgttctttagatttaccgttgattcggccacggagttcttatttggtgagtcc gtgcactccttaaaggacgaatctattggtatcaaccaagacgatatagattttgctggtagaaagg actttgctgagtcgttcaacaaagcccaggaatacttggctattagaaccttggtgcagacgttctac tggttggtcaacaacaaggagtttagagactgtaccaagctggtgcacaagttcaccaactactatg ttcagaaagctttggatgctagcccagaagagcttgaaaagcaaagtgggtatgtgttcttgtacga gcttgtcaagcagacaagagaccccaatgtgttgcgtgaccagtctttgaacatcttgttggccgga agagacaccactgctgggttgttgtcgtttgctgtctttgagttggccagacacccagagatctggg ccaagttgagagaggaaattgaacaacagtttggtcttggagaagactctcgtgttgaagagattac ctttgagagcttgaagagatgtgagtacttgaaagcgttccttaatgaaaccttgcgtatttacccaa gtgtcccaagaaacttcagaatcgccaccaagaacacgacattgccaaggggcggtggttcaga cggtacctcgccaatcttgatccaaaagggagaagctgtgtcgtatggtatcaactctactcatttgg accctgtctattacggccctgatgctgctgagttcagaccagagagatggtttgagccatcaaccaa aaagctcggctgggcttacttgccattcaacggtggtccaagaatctgtttgggtcagcagtttgcct tgacggaagctggctatgtgttggttagattggtgcaagagttctcccacgttaggctggacccaga cgaggtgtacccgccaaagaggttgaccaacttgaccatgtgtttgcaggatggtgctattgtcaa gtttgactag
SEQ ID NO:20	cytochrome P450 A18 (CYP52A18) Nuc. Seq	atgattgaacaaatcctagaatattggtatattgttgtgcctgtgttgtacatcatcaaacaactcattg cctacagcaagactcgcgtcttgatgaaacagttgggtgctgctccaatcacaaaccagttgtacg acaacgttttcggtatcgtcaacggatggaaggctctccagttcaagaaagagggcagagctcaa gagtacaacgatcacaagtttgacagctccaagaacccaagcgtcggcacctatgtcagtattcttt ttggcaccaagattgtcgtgaccaaggatccagagaatatcaaagctattttggcaacccagtttgg cgatttttctttgggcaagagacacgctctttttaaacctttgttaggtgatgggatcttcaccttggac ggcgaaggctggaagcatagcagatccatgttaagaccacagtttgccagagaacaagttgctca tgtgacgtcgttggaaccacacttccagttgttgaagaagcatatccttaaacacaagggtgagtac tttgatatccaggaattgttctttagatttactgtcgactcggccacggagttcttatttggtgagtccgt gcactccttaaaggacgaaactatcggtatcaaccaagacgatatagattttgctggtagaaagga

SEQ ID NO:	Description	Sequence
		ctttgctgagtcgttcaacaaagcccaggagtatttgtctattagaattttggtgcagaccttctactgg ttgatcaacaacaaggagtttagagactgtaccaagctggtgcacaagtttaccaactactatgttca gaaagctttggatgctaccccagaggaacttgaaaagcaaggcgggtatgtgttcttgtatgagctt gtcaagcagacgagagaccccaaggtgttgcgtgaccagtctttgaacatcttgttggcaggaag agacaccactgctgggttgttgtcctttgctgtgtttgagttggccagaaacccacacatctgggcc aagttgagagaggaaattgaacagcagtttggtcttggagaagactctcgtgttgaagagattacct ttgagagcttgaagagatgtgagtacttgaaagcgttccttaacgaaaccttgcgtgtttacccaagt gtcccaagaaacttcagaatcgccaccaagaatacaacattgccaaggggtggtggtccagacg gtacccagccaatcttgatccaaaagggagaaggtgtgtcgtatggtatcaactctacccacttaga tcctgtctattatggccctgatgctgctgagttcagaccagagagatggtttgagccatcaaccaga aagctcggctgggcttacttgccattcaacggtgggccacgaatctgtttgggtcagcagtttgcctt gaccgaagctggttacgttttggtcagattggtgcaagagttctcccacattaggctggacccagat gaagtgtatccaccaaagaggttgaccaacttgaccatgtgtttgcaggatggtgctattgtcaagtt tgactag
SEQ ID NO:21	cytochrome P450 A19 (CYP52A19) Nuc. Seq	atgctcgatcagatcttacattactggtacattgtcttgccattgttggccattatcaaccagatcgtgg ctcatgtcaggaccaattatttgatgaagaaattgggtgctaagccattcacacacgtccaacgtga cgggtggttgggcttcaaattcggccgtgaattcctcaaagcaaaaagtgctgggagactggttga tttaatcatctcccgtttccacgataatgaggacactttctccagctatgcttttggcaaccatgtggtg ttcaccagggaccccgagaatatcaaggcgcttttggcaacccagtttggtgatttttcattgggca gcagggtcaagttcttcaaaccattattggggtacggtatcttcacattggacgccgaaggctggaa gcacagcagagccatgttgagaccacagtttgccagagaacaagttgctcatgtgacgtcgttgg aaccacacttccagttgttgaagaagcatatccttaaacacaagggtgagtactttgatatccagga attgttctttagatttactgtcgactcggccacggagttcttatttggtgagtccgtgcactccttaaag gacgaggaaattggctacgacacgaaagacatgtctgaagaaagacgcagatttgccgacgcgt tcaacaagtcgcaagtctacgtggccaccagagttgctttacagaacttgtactggttggtcaacaa caaagagttcaaggagtgcaatgacattgtccacaagtttaccaactactatgttcagaaagccttg gatgctaccccagaggaacttgaaaagcaaggcgggtatgtgttcttgtatgagcttgtcaagcag acgagagaccccaaggtgttgcgtgaccagtctttgaacatcttgttggcaggaagagacaccact gctgggttgttgtcctttgctgtgtttgagttggccagaaacccacacatctgggccaagttgagag aggaaattgaacagcagtttggtcttggagaagactctcgtgttgaagagattacctttgagagcttg

SEQ ID NO:	Description	Sequence
		aagagatgtgagtacttgaaggccgtgttgaacgaaactttgagattacacccaagtgtcccaaga aacgcaagatttgcgattaaagacacgactttaccaagaggcggtggccccaacggcaaggatc ctatcttgatcaggaaggatgaggtggtgcagtactccatctcggcaactcagacaaatcctgctta ttatggcgccgatgctgctgattttagaccggaaagatggtttgaaccatcaactagaaacttggga tgggctttcttgccattcaacggtggtccaagaatctgtttgggacaacagtttgctttgactgaagcc ggttacgttttggttagacttgttcaggagtttccaaacttgtcacaagaccccgaaaccaagtaccc accacctagattggcacacttgacgatgtgcttgtttgacggtgcacacgtcaagatgtcatag
SEQ ID NO:22	cytochrome P450 A20 (CYP52A20) Nuc. Seq	atgctcgaccagatcttccattactggtacattgtcttgccattgttggtcattatcaagcagatcgtgg ctcatgccaggaccaattatttgatgaagaagttgggcgctaagccattcacacatgtccaactaga cgggtggtttggcttcaaatttggccgtgaattcctcaaagctaaaagtgctgggaggcaggttgat ttaatcatctcccgtttccacgataatgaggacactttctccagctatgcttttggcaaccatgtggtgt tcaccagggaccccgagaatatcaaggcgcttttggcaacccagtttggtgatttttcattgggaag cagggtcaaattcttcaaaccattgttggggtacggtatcttcaccttggacggcgaaggctggaa gcacagcagagccatgttgagaccacagtttgccagagagcaagttgctcatgtgacgtcgttgg aaccacatttccagttgttgaagaagcatattcttaagcacaagggtgaatactttgatatccaggaa ttgttctttagatttaccgttgattcagcgacggagttcttatttggtgagtccgtgcactccttaaggg acgaggaaattggctacgatacgaaggacatggctgaagaaagacgcaaatttgccgacgcgtt caacaagtcgcaagtctatttgtccaccagagttgctttacagacattgtactggttggtcaacaaca aagagttcaaggagtgcaacgacattgtccacaagttcaccaactactatgttcagaaagccttgg atgctaccccagaggaacttgaaaaacaaggcgggtatgtgttcttgtacgagcttgccaagcaga cgaaagaccccaatgtgttgcgtgaccagtctttgaacatcttgttggctggaagggacaccactg ctgggttgttgtcctttgctgtgtttgagttggccaggaacccacacatctgggccaagttgagaga ggaaattgaatcacactttgggctgggtgaggactctcgtgttgaagagattacctttgagagcttg aagagatgtgagtacttgaaagccgtgttgaacgaaacgttgagattacacccaagtgtcccaaga aacgcaagatttgcgattaaagacacgactttaccaagaggcggtggccccaacggcaaggatc ctatcttgatcagaaagaatgaggtggtgcaatactccatctcggcaactcagacaaatcctgcttat tatggcgccgatgctgctgattttagaccggaaagatggtttgagccatcaactagaaacttgggat gggcttacttgccattcaacggtggtccaagaatctgcttgggacaacagtttgctttgaccgaagc cggttacgttttggttagacttgttcaggaattccctagcttgtcacaggaccccgaaactgagtacc caccacctagattggcacacttgacgatgtgcttgtttgacggggcatacgtcaagatgcaatag

SEQ ID NO:	Description	Sequence
SEQ ID NO:23	cytochrome P450 D2 (CYP52D2) Nuc. Seq	atggctatatctagtttgctatcgtgggatgtgatctgtgtcgtcttcatttgcgtttgtgtttatttcgggt atgaatattgttatactaaatacttgatgcacaaacatggcgctcgagaaatcgagaatgtgatcaac gatgggttctttgggttccgcttacctttgctactcatgcgagccagcaatgagggccgacttatcga gttcagtgtcaagagattcgagtcggcgccacatccacagaacaagacattggtcaaccgggcat tgagcgttcctgtgatactcaccaaggacccagtgaatatcaaagcgatgctatcgacccagtttga tgacttttcccttgggttgagactacaccagtttgcgccgttgttggggaaaggcatctttacttttgga cggcccagagtggaagcagagccgatctatgttgcgtccgcaatttgccaaagatcgggtttctca tatcctggatctagaaccgcattttgtgttgcttcggaagcacattgatggccacaatggagactact tcgacatccaggagctctacttccggttctcgatggatgtggcgacggggtttttgtttggcgagtct gtggggtcgttgaaagacgaagatgcgaggttcctggaagcattcaatgagtcgcagaagtatttg gcaactagggcaacgttgcacgagttgtactttctttgtgacgggtttaggtttcgccagtacaacaa ggttgtgcgaaagttctgcagccagtgtgtccacaaggcgttagatgttgcaccggaagacacca gcgagtacgtgtttctccgcgagttggtcaaacacactcgagatcccgttgttttacaagaccaagc gttgaacgtcttgcttgctggacgcgacaccaccgcgtcgttattatcgtttgcaacatttgagctag cccggaatgaccacatgtggaggaagctacgagaggaggttatcctgacgatgggaccgtccag tgatgaaataaccgtggccgggttgaagagttgccgttacctcaaagcaatcctaaacgaaactctt cgactatacccaagtgtgcctaggaacgcgagatttgctacgaggaatacgacgcttcctcgtggc ggaggtccagatggatcgtttccgattttgataagaaagggccagccagtggggtatttcatttgtg ctacacacttgaatgagaaggtatatgggaatgatagccatgtgtttcgaccggagagatgggctg cgttagagggcaagagtttgggctggtcgtatcttccattcaacggcggcccgagaagctgccttg gtcagcagtttgcaatccttgaagcttcgtatgttttggctcgattgacacagtgctacacgacgata cagcttagaactaccgagtacccaccaaagaaactcgttcatctcacgatgagtcttctcaacggg gtgtacatccgaactagaacttga
SEQ ID NO:24	cytochrome P450:NADPH P450 reductase (Bacillus megaterium) nucleotide Nuc. Seq	atgacaattaaagaaatgcctcagccaaaaacgtttggagagcttaaaaatttaccgttattaaacac agataaaccggttcaagctttgatgaaaattgcggatgaattaggagaaatctttaaattcgaggcg cctggtcgtgtaacgcgctacttatcaagtcagcgtctaattaaagaagcatgcgatgaatcacgct ttgataaaaacttaagtcaagcgcttaaatttgtacgtgattttgcaggagacgggttatttacaagct ggacgcatgaaaaaaattggaaaaaagcgcataatatcttacttccaagcttcagtcagcaggcaa tgaaaggctatcatgcgatgatggtcgatatcgccgtgcagcttgttcaaaagtgggagcgtctaa atgcagatgagcatattgaagtaccggaagacatgacacgtttaacgcttgatacaattggtctttgc

SEQ ID NO:	Description	Sequence
		ggctttaactatcgctttaacagcttttaccgagatcagcctcatccatttattacaagtatggtccgtg cactggatgaagcaatgaacaagctgcagcgagcaaatccagacgacccagcttatgatgaaaa caagcgccagtttcaagaagatatcaaggtgatgaacgacctagtagataaaattattgcagatcg caaagcaagcggtgaacaaagcgatgatttattaacgcatatgctaaacggaaaagatccagaaa cgggtgagccgcttgatgacgagaacattcgctatcaaattattacattcttaattgcgggacacga aacaacaagtggtcttttatcatttgcgctgtatttcttagtgaaaaatccacatgtattacaaaaagca gcagaagaagcagcacgagttctagtagatcctgttccaagctacaaacaagtcaaacagcttaa atatgtcggcatggtcttaaacgaagcgctgcgcttatggccaactgctcctgcgttttccctatatg caaaagaagatacggtgcttggaggagaatatcctttagaaaaaggcgacgaactaatggttctga ttcctcagcttcaccgtgataaaacaatttggggagacgatgtggaagagttccgtccagagcgttt tgaaaatccaagtgcgattccgcagcatgcgtttaaaccgtttggaaacggtcagcgtgcgtgtatc ggtcagcagttcgctcttcatgaagcaacgctggtacttggtatgatgctaaaacactttgactttga agatcatacaaactacgagctggatattaaagaaactttaacgttaaaacctgaaggctttgtggtaa aagcaaaatcgaaaaaaattccgcttggcggtattccttcacctagcactgaacagtctgctaaaaa agtacgcaaaaaggcagaaaacgctcataatacgccgctgcttgtgctatacggttcaaatatggg aacagctgaaggaacggcgcgtgatttagcagatattgcaatgagcaaaggatttgcaccgcagg tcgcaacgcttgattcacacgccggaaatcttccgcgcgaaggagctgtattaattgtaacggcgt cttataacggtcatccgcctgataacgcaaagcaatttgtcgactggttagaccaagcgtctgctga tgaagtaaaaggcgttcgctactccgtatttggatgcggcgataaaaactgggctactacgtatcaa aaagtgcctgcttttatcgatgaaacgcttgccgctaaaggggcagaaaacatcgctgaccgcggt gaagcagatgcaagcgacgactttgaaggcacatatgaagaatggcgtgaacatatgtggagtg acgtagcagcctactttaacctcgacattgaaaacagtgaagataataaatctactctttcacttcaat ttgtcgacagcgccgcggatatgccgcttgcgaaaatgcacggtgcgttttcaacgaacgtcgtag caagcaaagaacttcaacagccaggcagtgcacgaagcacgcgacatcttgaaattgaacttcca aaagaagcttcttatcaagaaggagatcatttaggtgttattcctcgcaactatgaaggaatagtaaa ccgtgtaacagcaaggttcggcctagatgcatcacagcaaatccgtctggaagcagaagaagaa aaattagctcatttgccactcgctaaaacagtatccgtagaagagcttctgcaatacgtggagcttca agatcctgttacgcgcacgcagcttcgcgcaatggctgctaaaacggtctgcccgccgcataaag tagagcttgaagccttgcttgaaaagcaagcctacaaagaacaagtgctggcaaaacgtttaacaa tgcttgaactgcttgaaaaatacccggcgtgtgaaatgaaattcagcgaatttatcgcccttctgcca

SEQ ID NO:	Description	Sequence
		agcatacgcccgcgctattactcgatttcttcatcacctcgtgtcgatgaaaaacaagcaagcatca cggtcagcgttgtctcaggagaagcgtggagcggatatggagaatataaaggaattgcgtcgaac tatcttgccgagctgcaagaaggagatacgattacgtgctttatttccacaccgcagtcagaatttac gctgccaaaagaccctgaaacgccgcttatcatggtcggaccgggaacaggcgtcgcgccgttt agaggctttgtgcaggcgcgcaaacagctaaaagaacaaggacagtcacttggagaagcacatt tatacttcggctgccgttcacctcatgaagactatctgtatcaagaagagcttgaaaacgcccaaag cgaaggcatcattacgcttcataccgctttttctcgcatgccaaatcagccgaaaacatacgttcagc acgtaatggaacaagacggcaagaaattgattgaacttcttgatcaaggagcgcacttctatatttg cggagacggaagccaaatggcacctgccgttgaagcaacgcttatgaaaagctatgctgacgttc accaagtgagtgaagcagacgctcgcttatggctgcagcagctagaagaaaaaggccgatacgc aaaagacgtgtgggctgggtaa
SEQ ID NO:25	NADPH cytochrome P450 reductase, CPR (C. tropicalis strain ATCC750) Nuc. Seq	atggcattagataagttagatttatatgttattataacattggtggttgcaattgcagcttattttgcaaag aatcagtttcttgaccaacaacaagataccgggttccttaatactgatagtggagatggtaattcaag agatatcttacaagctttgaagaagaacaataaaaatacgttattattatttggatcccaaacaggtac agcagaagattatgccaacaaattgtcaagagaattgcattcaagatttggtttgaaaaccatggttg ctgatttcgctgattatgatttcgaaaacttcggagatattactgaagatatcttggttttctttattgttgc tacttatggtgaaggtgaaccaaccgataatgctgacgaatttcacacttggttgactgaagaagct gacaccttgagtactttgaaatatactgtttttggtttgggtaattcaacttatgaattcttcaatgctattg gtagaaaatttgacagattgttgggagaaaaaggtggtgacagatttgctgaatacggtgaaggtg acgatggtactggtactttagatgaagatttcttggcctggaaggataacgtgtttgattccttaaaga atgatttgaattttgaagaaaaagagttgaaatacgaaccaaatgttaaattgactgaaagagatgat ttatctggcaatgatccagatgtctccttgggtgaaccaaatgtcaaatacattaaatctgaaggtgtt gacttaactaaaggtccatttgatcatactcatccatttttggctagaattgttaaaactaaagaattgtt tacttctgaagacagacattgtgttcatgttgaatttgatatttctgaatcaaacttgaaatataccaccg gtgatcatcttgcaatctggccatctaactctgatgaaaacattaagcaatttgccaaatgttttggttt agaagacaaacttgatactgttattgaattgaaagctttggattccacttattccatcccattccctaat ccaatcacttatggagctgttattagacaccatttggaaatttcaggtcctgtttctagacaatttttctta tctattgctggatttgcccctgatgaagaaactaaaaagtcatttactagaattggtggtgataagcaa gaatttgctagtaaagtcacccgtagaaaattcaacattgccgatgctttattatttgcttccaacaaca gaccatggtccgatgttccattcgaattccttattgaaaatgtccaacacttaactcctcgttattactc

SEQ ID NO:	Description	Sequence
		catttcttcttcctcattaagtgaaaagcaaaccattaatgttactgctgttgttgaagccgaagaaga agctgatggaagaccagttactggtgttgtcaccaacttgttgaagaatattgaaattgaacaaaac aaaactggtgaaaccccaatggttcattatgatttgaatggtccaagaggcaaatttagcaagttca gattgccagttcacgttagaagatctaatttcaaattaccaaagaatagcactaccccagttattttgat tggtccaggtaccggtgttgcaccattgagaggttttgttagagaaagagttcaacaagttaaaaat ggtgttaatgttggtaagactgtattgttttatggatgtagaaattccgaacaagatttcttgtacaaac aagaatggagtgaatatgcctcagtattgggagaaaatttcgaaatgtttaatgccttctcaagacaa gatccaactaagaaagtttatgttcaagataagattttagaaaatagtgctcttgttgatgagttattatc tagtggagcaattatttatgtttgtggtgatgccagtagaatggctagagatgttcaagctgcaattgc caagattgttgccaaaagtagagatatccacgaagataaagctgctgaattggttaaatcttggaaa gttcaaaatagataccaagaagatgtctggtaa
SEQ ID NO:26	NADPH cytochrome P450 reductase A, CPRA (Candida strain ATCC20336) Nuc. Seq	atggctttagacaagttagatttgtatgtcatcataacattggtggtcgctgtagccgcctattttgcta agaaccagttccttgatcagccccaggacaccgggttcctcaacacggacagcggaagcaactc cagagacgtcttgctgacattgaagaagaataataaaaacacgttgttgttgtttgggtcccagacg ggtacggcagaagattacgccaacaaattgtccagagaattgcactccagatttggcttgaaaacg atggttgcagatttcgctgattacgattgggataacttcggagatatcaccgaagacatcttggtgttt ttcattgttgccacctatggtgagggtgaacctaccgataatgccgacgagttccacacctggttga ctgaagaagctgacactttgagtaccttgaaatacaccgtgttcgggttgggtaactccacgtacga gttcttcaatgccattggtagaaagtttgacagattgttgagcgagaaaggtggtgacaggtttgctg aatacgctgaaggtgatgacggtactggcaccttggacgaagatttcatggcctggaaggacaat gtctttgacgccttgaagaatgatttgaactttgaagaaaaggaattgaagtacgaaccaaacgtga aattgactgagagagacgacttgtctgctgctgactcccaagtttccttgggtgagccaaacaaga agtacatcaactccgagggcatcgacttgaccaagggtccattcgaccacacccacccatacttg gccagaatcaccgagacgagagagttgttcagctccaaggacagacactgtatccacgttgaattt gacatttctgaatcgaacttgaaatacaccaccggtgaccatctagctatctggccatccaactccg acgaaaacattaagcaatttgccaagtgtttcggattggaagataaactcgacactgttattgaattg aaggcgttggactccacttacaccatcccattcccaaccccaattacctacggtgctgtcattagac accatttagaaatctccggtccagtctcgagacaattctttttgtcaattgctgggtttgctcctgatga agaaacaaagaaggcttttaccagacttggtggtgacaagcaagaattcgccgccaaggtcaccc gcagaaagttcaacattgccgatgccttgttatattcctccaacaacgctccatggtccgatgttcctt

SEQ ID NO:	Description	Sequence
		ttgaattccttattgaaaacgttccacacttgactccacgttactactccatttcgtcttcgtcattgagt gaaaagcaactcatcaacgttactgcagttgttgaagccgaagaagaagctgatggcagaccagt cactggtgttgtcaccaacttgttgaagaacgttgaaattgtgcaaaacaagactggcgaaaagcc acttgtccactacgatttgagcggcccaagaggcaagttcaacaagttcaagttgccagtgcatgt gagaagatccaactttaagttgccaaagaactccaccaccccagttatcttgattggtccaggtact ggtgttgccccattgagaggttttgtcagagaaagagttcaacaagtcaagaatggtgtcaatgttg gcaagactttgttgttttatggttgcagaaactccaacgaggacttttttgtacaagcaagaatgggcc gagtacgcttctgttttgggtgaaaactttgagatgttcaatgccttctccagacaagacccatccaa gaaggtttacgtccaggataagattttagaaaacagccaacttgtgcacgagttgttgactgaaggt gccattatctacgtctgtggtgatgccagtagaatggctagagacgtgcagaccacaatttccaag attgttgctaaaagcagagaaattagtgaagacaaggctgctgaattggtcaagtcctggaaggtc caaaatagataccaagaagatgtttggtag
SEQ ID NO:27	NADPH cytochrome P450 reductase B, CPRB (Candida strain ATCC20336) Nuc. Seq	atggctttagacaagttagatttgtatgtcatcataacattggtggtcgctgtggccgcctattttgcta agaaccagttccttgatcagccccaggacaccgggttcctcaacacggacagcggaagcaactc cagagacgtcttgctgacattgaagaagaataataaaaacacgttgttgttgtttgggtcccagacc ggtacggcagaagattacgccaacaaattgtcaagagaattgcactccagatttggcttgaaaacc atggttgcagatttcgctgattacgattgggataacttcggagatatcaccgaagatatcttggtgtttt tcatcgttgccacctacggtgagggtgaacctaccgacaatgccgacgagttccacacctggttga ctgaagaagctgacactttgagtactttgagatataccgtgttcgggttgggtaactccacctacga gttcttcaatgctattggtagaaagtttgacagattgttgagtgagaaaggtggtgacagatttgctga atatgctgaaggtgacgacggcactggcaccttggacgaagatttcatggcctggaaggataatgt ctttgacgccttgaagaatgacttgaactttgaagaaaaggaattgaagtacgaaccaaacgtgaa attgactgagagagatgacttgtctgctgccgactcccaagtttccttgggtgagccaaacaagaa gtacatcaactccgagggcatcgacttgaccaagggtccattcgaccacacccacccatacttgg ccaggatcaccgagaccagagagttgttcagctccaaggaaagacactgtattcacgttgaatttg acatttctgaatcgaacttgaaatacaccaccggtgaccatctagccatctggccatccaactccga cgaaaacatcaagcaatttgccaagtgtttcggattggaagataaactcgacactgttattgaattga aggcattggactccacttacaccattccattcccaactccaattacttacggtgctgtcattagacacc atttagaaatctccggtccagtctcgagacaattctttttgtcgattgctgggtttgctcctgatgaaga aacaaagaagactttcaccagacttggtggtgacaaacaagaattcgccaccaaggttacccgca

SEQ ID NO:	Description	Sequence
		gaaagttcaacattgccgatgccttgttatattcctccaacaacactccatggtccgatgttccttttga gttccttattgaaaacatccaacacttgactccacgttactactccatttcttcttcgtcgttgagtgaaa aacaactcatcaatgttactgcagtcgttgaggccgaagaagaagccgatggcagaccagtcact ggtgttgttaccaacttgttgaagaacattgaaattgcgcaaaacaagactggcgaaaagccacttg ttcactacgatttgagcggcccaagaggcaagttcaacaagttcaagttgccagtgcacgtgagaa gatccaactttaagttgccaaagaactccaccaccccagttatcttgattggtccaggtactggtgtt gccccattgagaggtttcgttagagaaagagttcaacaagtcaagaatggtgtcaatgttggcaag actttgttgttttatggttgcagaaactccaacgaggacttttgtacaagcaagaatgggccgagta cgcttctgttttgggtgaaaactttgagatgttcaatgccttctctagacaagacccatccaagaaggt ttacgtccaggataagattttagaaaacagccaacttgtgcacgaattgttgaccgaaggtgccatt atctacgtctgtggtgacgccagtagaatggccagagacgtccagaccacgatctccaagattgtt gccaaaagcagagaaatcagtgaagacaaggccgctgaattggtcaagtcctggaaagtccaaa atagataccaagaagatgtttggtag
SEQ ID NO:28	cytochrome P450:NADPH P450 reductase (Bacillus megaterium) amino acid [P450 activity shown in italics, P450 reductase activity shown in normal font] A.A. Seq	*mtikempqpktfgelknlpUntdkpvqalmkiadelgeifkfeapgrvtrylssqrlikeacdes rfdknlsqalkfvrdfagdglftswtheknwkkahniUpsfsqqamkgyhammvdiavqlvq kwerlnadehievpedmtrltldtiglcgfnyrfnsfyrdqphpfitsmvrasdeamnksqran pddpaydenkrqfqedikvmndlvdkiiadrkasgeqsddUthmlngkdpetgeplddenir yqiitfliaghettsgllsfasyflvknphvlqkaaeeaarvlvdpvpsykqvkqlkyvgmvlnea srlwptapafslyakedtvlggeyplekgdelmvsipqlhrdktiwgddveefrperfenpsaip qhafkpfgngqracigqqfalheatsvlgmmlkhfdfedhtnyesdiketltlkpegfvvkaksk* ^z/?/ggz/?5/?5teg5a^^vrkkaenahntpslvlygsnmgtaegtardladiamskgfapqvat ldshagnlpregavlivtasynghppdnakqfvdwldqasadevkgvrysvfgcgdknwat tyqkvpafidetlaakgaeniadrgeadasddfegtyeewrehmwsdvaayfnldiensedn kstlslqfvdsaadmplakmhgafstnwaskelqqpgsarstrhleielpkeasyqegdhlgv iprnyegivnrvtarfgldasqqirseaeeeklahlplaktvsveelsqyvelqdpvtrtqlrama aktvcpphkveleallekqaykeqvsakrltmleslekypacemkfsefialspsirpryysiss sprvdekqasitvswsgeawsgygeykgiasnylaesqegdtitcfistpqseftspkdpetpl imvgpgtgvapfrgfvqarkqlkeqgqslgeahlyfgcrsphedysyqeelenaqsegiitlht afsrmpnqpktyvqhvmeqdgkkliellddqgahfyicgdgsqmapaveatlmksyadvhq vseadarlwsqqleekgryakdvwag*

SEQ ID NO:	Description	Sequence
SEQ ID NO:29	acyl CoA oxidase, POX4 (Candida strain ATCC20336) nucleotide	ATGACTTTTACAAAGAAAAACGTTAGTGTATCACAAGGTC CTGACCCTAGATCATCCATCCAAAAGGAAAGAGACAGCTC CAAATGGAACCCTCAACAAATGAACTACTTCTTGGAAGGC TCCGTCGAAAGAAGTGAGTTGATGAAGGCTTTGGCCCAAC AAATGGAAAGAGACCCAATCTTGTTCACAGACGGCTCCTA CTACGACTTGACCAAGGACCAACAAAGAGAATTGACCGCC GTCAAGATCAACAGAATCGCCAGATACAGAGAACAAGAAT CCATCGACACTTTCAACAAGAGATTGTCCTTGATTGGTATC TTTGACCCACAGGTCGGTACCAGAATTGGTGTCAACCTCGG TTTGTTCCTTTCTTGTATCAGAGGTAACGGTACCACTTCCCA ATTGAACTACTGGGCTAACGAAAAGGAAACCGCTGACGTT AAAGGTATCTACGGTTGTTTCGGTATGACCGAATTGGCCCA CGGTTCCAACGTTGCTGGTTTGGAAACCACCGCCACATTTG ACAAGGAATCTGACGAGTTTGTCATCAACACCCCACACATT GGTGCCACCAAGTGGTGGATTGGTGGTGCTGCTCACTCCGC CACCCACTGTTCTGTCTACGCCAGATTGATTGTTGACGGTC AAGATTACGGTGTCAAGACTTTTGTTGTCCCATTGAGAGAC TCCAACCACGACCTCATGCCAGGTGTCACTGTTGGTGACAT TGGTGCCAAGATGGGTAGAGATGGTATCGATAACGGTTGG ATCCAATTCTCCAACGTCAGAATCCCAAGATTCTTTATGTT GCAAAAGTTCTGTAAGGTTTCTGCTGAAGGTGAAGTCACCT TGCCACCTTTGGAACAATTGTCTTACTCCGCCTTGTTGGGT GGTAGAGTCATGATGGTTTTGGACTCCTACAGAATGTTGGC TAGAATGTCCACCATTGCCTTGAGATACGCCATTGGTAGAA GACAATTCAAGGGTGACAATGTCGATCCAAAAGATCCAAA CGCTTTGGAAACCCAATTGATAGATTACCCATTGCACCAAA AGAGATTGTTCCCATACTTGGCTGCTGCCTACGTCATCTCC GCTGGTGCCCTCAAGGTTGAAGACACCATCCATAACACCTT GGCTGAATTGGACGCTGCCGTTGAAAAGAACGACACCAAG

SEQ ID NO:	Description	Sequence
		GCTATCTTTAAGTCTATTGACGACATGAAGTCATTGTTTGT TGACTCTGGTTCCTTGAAGTCCACTGCCACTTGGTTGGGTG CTGAAGCCATTGACCAATGTAGACAAGCCTGTGGTGGTCA CGGTTACTCGTCCTACAACGGCTTCGGTAAAGCCTACAACG ATTGGGTTGTCCAATGTACTTGGGAAGGTGACAACAATGTC TTGGCCATGAGTGTTGGTAAGCCAATTGTCAAGCAAGTTAT CAGCATTGAAGATGCCGGCAAGACCGTCAGAGGTTCCACC GCTTTCTTGAACCAATTGAAGGACTACACTGGTTCCAACAG CTCCAAGGTTGTTTTGAACACTGTTGCTGACTTGGACGACA TCAAGACTGTCATCAAGGCTATTGAAGTTGCCATCATCAGA TTGTCCCAAGAAGCTGCTTCTATTGTCAAGAAGGAATCTTT CGACTATGTCGGCGCTGAATTGGTTCAACTCTCCAAGTTGA AGGCTCACCACTACTTGTTGACTGAATACATCAGAAGAATT GACACCTTTGACCAAAAGGACTTGGTTCCATACTTGATCAC CCTCGGTAAGTTGTACGCTGCCACTATTGTCTTGGACAGAT TTGCCGGTGTCTTCTTGACTTTCAACGTTGCCTCCACCGAA GCCATCACTGCTTTGGCCTCTGTGCAAATTCCAAAGTTGTG TGCTGAAGTCAGACCAAACGTTGTTGCTTACACCGACTCCT TCCAACAATCCGACATGATTGTCAATTCTGCTATTGGTAGA TACGATGGTGACATCTATGAGAACTACTTTGACTTGGTCAA GTTGCAGAACCCACCATCCAAGACCAAGGCTCCTTACTCTG ATGCTTTGGAAGCCATGTTGAACAGACCAACCTTGGACGA AAGAGAAAGATTTGAAAAGTCTGATGAAACCGCTGCTATC TTGTCCAAGTAA
SEQ ID NO:30	acyl CoA oxidase, POX4 (Candida strain ATCC20336) amino acid	MTFTKKNVSVSQGPDPRSSIQKERDSSKWNPQQMNYFLEGSV ERSELMKALAQQMERDPILFTDGSYYDLTKDQQRELTAVKIN RIARYREQESIDTFNKRLSLIGIFDPQVGTRIGVNLGLFLSCIRG NGTTSQLNYWANEKETADVKGIYGCFGMTELAHGSNVAGLE TTATFDKESDEFVINTPHIGATKWWIGGAAHSATHCSVYARLI VDGQDYGVKTFVVPLRDSNHDLMPGVTVGDIGAKMGRDGID

SEQ ID NO:	Description	Sequence
		NGWIQFSNVPvIPPvFFMLQKFCKVSAEGEVTLPPLEQLSYSALL GGRVMMVLDSYRMLARMSTIALRYAIGRPvQFKGDNVDPKDP NALETQLIDYPLHQKRLFPYLAAAYVISAGALKVEDTIHNTLA ELDAAVEKNDTKAIFKSIDDMKSLFVDSGSLKSTATWLGAEA IDQCRQACGGHGYSSYNGFGKAYNDWVVQCTWEGDNNVLA MSVGKPIVKQVISIEDAGKTVRGSTAFLNQLKDYTGSNSSKV VLNTVADLDDIKTVIKAIEVAIIRLSQEAASIVKKESFDYVGAE LVQLSKLKAHHYLLTEYIRRIDTFDQKDLVPYLITLGKLYAAT IVLDRFAGVFLTFNVASTEAITALASVQIPKLCAEVRPNVVAY TDSFQQSDMIVNSAIGRYDGDIYENYFDLVKLQNPPSKTKAPY SDALEAMLNRPTLDERERFEKSDETAAILSK*
SEQ ID NO:31	acyl CoA oxidase, POX5 (Candida strain ATCC20336) nucleotide	ATGCCTACCGAACTTCAAAAAGAAAGAGAACTCACCAAGT TCAACCCAAAGGAGTTGAACTACTTCTTGGAAGGTTCCCAA GAAAGATCCGAGATCATCAGCAACATGGTCGAACAAATGC AAAAAGACCCTATCTTGAAGGTCGACGCTTCATACTACAA CTTGACCAAAGACCAACAAAGAGAAGTCACCGCCAAGAAG ATTGCCAGACTCTCCAGATACTTTGAGCACGAGTACCCAGA CCAACAGGCCCAGAGATTGTCGATCCTCGGTGTCTTTGACC CACAAGTCTTCACCAGAATCGGTGTCAACTTGGGTTTGTTT GTTTCCTGTGTCCGTGGTAACGGTACCAACTCCCAGTTCTT CTACTGGACCATAAATAAGGGTATCGACAAGTTGAGAGGT ATCTATGGTTGTTTTGGTATGACTGAGTTGGCCCACGGTTC CAACGTCCAAGGTATTGAAACCACCGCCACTTTTGACGAA GACACTGACGAGTTTGTCATCAACACCCCACACATTGGTGC CACCAAGTGGTGGATCGGTGGTGCTGCGCACTCCGCCACC CACTGCTCCGTCTACGCCAGATTGAAGGTCAAAGGAAAGG ACTACGGTGTCAAGACCTTTGTTGTCCCATTGAGAGACTCC AACCACGACCTCGAGCCAGGTGTGACTGTTGGTGACATTG GTGCCAAGATGGGTAGAGACGGTATCGATAACGGTTGGAT CCAGTTCTCCAACGTCAGAATCCCAAGATTCTTTATGTTGC

SEQ ID NO:	Description	Sequence
		AAAAGTACTGTAAGGTTTCCCGTCTGGGTGAAGTCACCATG CCACCATCTGAACAATTGTCTTACTCGGCTTTGATTGGTGG TAGAGTCACCATGATGATGGACTCCTACAGAATGACCAGT AGATTCATCACCATTGCCTTGAGATACGCCATCCACAGAAG ACAATTCAAGAAGAAGGACACCGATACCATTGAAACCAAG TTGATTGACTACCCATTGCATCAAAAGAGATTGTTCCCATT CTTGGCTGCCGCTTACTTGTTCTCCCAAGGTGCCTTGTACTT AGAACAAACCATGAACGCAACCAACGACAAGTTGGACGA AGCTGTCAGTGCTGGTGAAAAGGAAGCCATTGACGCTGCC ATTGTCGAATCCAAGAAATTGTTCGTCGCTTCCGGTTGTTT GAAGTCCACCTGTACCTGGTTGACTGCTGAAGCCATTGACG AAGCTCGTCAAGCTTGTGGTGGTCACGGTTACTCGTCTTAC AACGGTTTCGGTAAAGCCTACTCCGACTGGGTTGTCCAATG TACCTGGGAAGGTGACAACAACATCTTGGCCATGAACGTT GCCAAGCCAATGGTTAGAGACTTGTTGAAGGAGCCAGAAC AAAAGGGATTGGTTCTCTCCAGCGTTGCCGACTTGGACGAC CCAGCCAAGTTGGTTAAGGCTTTCGACCACGCCCTTTCCGG CTTGGCCAGAGACATTGGTGCTGTTGCTGAAGACAAGGGT TTCGACATTACCGGTCCAAGTTTGGTTTTGGTTTCCAAGTT GAACGCTCACAGATTCTTGATTGACGGTTTCTTCAAGCGTA TCACCCCAGAATGGTCTGAAGTCTTGAGACCTTTGGGTTTC TTGTATGCCGACTGGATCTTGACCAACTTTGGTGCCACCTT CTTGCAGTACGGTATCATTACCCCAGATGTCAGCAGAAAG ATTTCCTCCGAGCACTTCCCAGCCTTGTGTGCCAAGGTTAG ACCAAACGTTGTTGGTTTGACTGATGGTTTCAACTTGACTG ACATGATGACCAATGCTGCTATTGGTAGATATGATGGTAAC GTCTACGAACACTACTTCGAAACTGTCAAGGCTTTGAACCC ACCAGAAAACACCAAGGCTCCATACTCCAAGGCTTTGGAA GACATGTTGAACCGTCCAGACCTTGAAGTCAGAGAAAGAG GTGAAAAGTCCGAAGAAGCTGCTGAAATCTTGTCCAGTTA

SEQ ID NO:	Description	Sequence
		A
SEQ ID NO:32	acyl CoA oxidase, POX5 (Candida strain ATCC20336) amino acid	MPTELQKERELTKFNPKELNYFLEGSQERSEIISNMVEQMQKD PILKVDASYYNLTKDQQREVTAKKIARLSPvYFEHEYPDQQAQ RLSILGVFDPQVFTRIGVNLGLFVSCVRGNGTNSQFFYWTINK GIDKLRGIYGCFGMTELAHGSNVQGIETTATFDEDTDEFVINT PHIGATKWWIGGAAHSATHCSVYARLKVKGKDYGVKTFVVP LRDSNHDLEPGVTVGDIGAKMGRDGIDNGWIQFSNVRIPRFF MLQKYCKVSRSGEVTMPPSEQLSYSALIGGRVTMMMDSYRM TSRFITIALRYAIHRRQFKKKDTDTIETKLIDYPLHQKRLFPFLA AAYLFSQGALYLEQTMNATNDKLDEAVSAGEKEAIDAAIVES KKLFVASGCLKSTCTWLTAEAIDEARQACGGHGYSSYNGFG KAYSDWVVQCTWEGDNNILAMNVAKPMVRDLLKEPEQKGL VLSSVADLDDPAKLVKAFDHALSGLARDIGAVAEDKGFDITG PSLVLVSKLNAHRFLIDGFFKRITPEWSEVLRPLGFLYADWILT NFGATFLQYGIITPDVSRKISSEHFPALCAKVRPNVVGLTDGF NLTDMMTNAAIGRYDGNVYEHYFETVKALNPPENTKAPYSK ALEDMLNRPDLEVRERGEKSEEAAEILSS*
SEQ ID NO:33	Acyl-CoA Hydrolase (ACHA) Nucleotide Seq	atgatcagaaccgtccgttatcaatccctcaagaggttcagacctctggctttgtctcctgtttttcgtc cacgctacaactcccagaaggccaatttccaccgtccagaccaccctgggtccgacgagccagc tgaagccgccgacgccgccgccacgatcctcgccgagttgcgagacaagcagacgaacccga acaaggccacctggctcgatgcgttaacggagcgggagaagttgcgtgccgagggcaagacga ttgacagtttcagctacgttgaccccaagacgaccgtcgtgggggagaagacacgcagtgactcg ttctcgttcttgttgttgccgttcaaggacgacaagtggttgtgtgacgcgtacatcaatgcgtttggc cggttgcgtgtagcgcagttgttccaggacttggacgccttggcggggcgcatcgcgtacaggca ctgttccccagcggagcccgtgaatgtcacggcgagcgtggatagggtgtacatggtgaagaaa gtggacgagattaacaattacaatttcgtgttggcggggtccgtgacgtggaccgggagatcgtcg atggagatcacggtgaaagggtatgcttttgaagacgccgtgccggatataacgaacgaggagtc cttgccggcagagaatgtgtttttggctgctaatttcaccttcgtggcacggaacccacttacacaca agtcctttgctattaacagattgttgcccgtgactgagaaggactgggtcgactatcgccgtgctga gtcccacaacgccaagaagaagttgatggcaaagaacaagaagatcttggagcctaccgcgga

SEQ ID NO:	Description	Sequence
		agagtccaagttgatctacgacatgtggagatcgtccaagtccttacagaacatcgagagggcca acgatgggatcgcgttcatgaaggacacgaccatgaagtccaccttgttcatgcagccccagtac cgtaacagacactcatacatgattttcggagggtacttgttaagacaaactttcgaattggcctactgt accgcggcaacgttttccctggccgggccccgtttcgtcagcttggactccaccacgttcaagaac cccgtgcccgtggggtcggtgctcaccatggactcgtcgatctcgtacacggagcacgtgcacg agggagtggaggagattgacgcggactcaccgttcaacttcagcttgcctgccacgaacaagatc tcgaagaaccccgaggcgttcttgtcggaacccggcacgttgattcaagtcaaggtcgacacata catccaggagttagagcagagtgtgaagaagcccgcgggtacgttcatctactcgttctatgttgat aaagaaagcgttactgttgatggaaaggcgtcgttttgttcagttatcccgcagacgtactccgaga tgatgacttatgtgggcgggagaagaagagcccaggatactgctaactacgtggagactttgccg tttagtggaagcggcaattaa
SEQ ID NO:34	Acyl-CoA Hydrolase (ACHA) Amino Acid Seq	MIRTVRYQSLKRFRPSALSPVFRPRYNSQKANFHRPDHPGSDE PAEAADAAATILAELRDKQTNPNKATWLDALTEREKLRAEG KTIDSFSYVDPKTTVVGEKTRSDSFSFLLLPFKDDKWLCDAYI NAFGRLRVAQLFQDLDALAGRIAYRHCSPAEPVNVTASVDRV YMVKKVDEINNYNFVLAGSVTWTGRSSMEITVKGYAFEDAV PDITNEESLPAENVFLAANFTFVARNPLTHKSFAINRLLPVTEK DWVDYRRAESHNAKKKLMAKNKKILEPTAEESKLIYDMWRS SKSLQNIERANDGIAFMKDTTMKSTLFMQPQYRNRHSYMIFG GYLLRQTFELAYCTAATFSSAGPRFVSLDSTTFKNPVPVGSVL TMDSSISYTEHVHEGVEEIDADSPFNFSLPATNKISKNPEAFLS EPGTLIQVKVDTYIQELEQSVKKPAGTFIYSFYVDKESVTVDG KASFCSVIPQTYSEMMTYVGGRRRAQDTANYVETLPFSGSGN
SEQ ID NO:35	Acyl-CoA Hydrolase (ACHB) Nucleotide Seq	atgatcagaaccgtccgttatcaatccttcaagaggttcaaacctctgactttatcccccgttttccgtc cacgctacaactcccagaaggccaatttccaccgtccagaccacgctgggtccgacgagccagc cgaagccgccgacgccgctgccacgatcctcgccgagttgcgagacaagcagacgaacccga acaaggccacctggctcgatgcgttaacggagcgggagaagttgcgtgccgagggcaagacaa tcgacagcttcagctacgttgaccccaagacaaccgtcgtgggggagaagacacgcagcgactc gttctcgttcttgttgttgccgttcaaggacgacaagtggttgtgtgacgcgtacatcaatgcgtttgg ccggttgcgtgtagcgcagttgttccaggacttggacgccttggcgggccgcatcgcgtacaggc

SEQ ID NO:	Description	Sequence
		actgttcccccgctgagcccgtgaatgtcacggcgagcgtggatagagtgtatatggtgaagaaa gtggacgagattaataattacaatttcgtgttggcggggtccgtgacgtggaccgggagatcgtcg atggagatcacggtcaaagggtatgcttttgaagacgccgtgccggagataactaacgaggagtc cttgccggcagagaatgtgttcttggctgttaatttcaccttcgtggcacgtaacccactcacacaca agtccttcgctattaacagattgttgcccgtgactgagaaggactgggtcgattatcgccgtgctga gtcccacaacgccaagaagaagttgatggcaaagaacaagaagatcttggagcctaccccggaa gagtccaagttgatctacgacatgtggagatcgtccaagtccttacagaacatcgagaaggccaa cgacgggatcgcgttcatgaaggacacgataatgaagtccaccttgttcatgcagccccagtacc gtaacagacactcatacatgattttcggtgggtatttgttaagacaaactttcgaattggcctattgtac cgcagcaacgttttccctggcgggaccccgtttcgtcagcttggactccaccacgttcaagaaccc cgtgcccgtggggtcggtgctcaccatggactcgtcgatctcgtacacggagcacgtccacgatg gcgttgaggagattgacgccgactccccgttcaacttcagcttgcctgccacgaacaagatctcga agaaccccgaggcgttcttgtcggagcccggcacgttgatccaagtcaaggtcgacacgtacatc caggagttagagcaaagtgtgaagaagcctgcgggaacgttcatctactcgttctatgttgataaag agagcgttactgtggatggaaaggcgtcgttttgttcagttatcccgcagacgtactccgagatgat gacttatgtgggcgggagaagaagagcccaggatactgctaattacgtggagactttgccgtttag tggaagcggcaattaa
SEQ ID NO:36	Acyl-CoA Hydrolase (ACHB)	MIRTVRYQSFKRFKPLTLSPVFRPRYNSQKANFHRPDHAGSDE PAEAADAAATILAELRDKQTNPNKATWLDALTEREKLRAEG KTIDSFSYVDPKTTVVGEKTRSDSFSFLLLPFKDDKWLCDAYI NAFGRLRVAQLFQDLDALAGRIAYRHCSPAEPVNVTASVDRV YMVKKVDEINNYNFVLAGSVTWTGRSSMEITVKGYAFEDAV PEITNEESLPAENVFLAVNFTFVARNPLTHKSFAINRLLPVTEK DWVDYRRAESHNAKKKLMAKNKKILEPTPEESKLIYDMWRS SKSLQNIEKANDGIAFMKDTIMKSTLFMQPQYRNRHSYMIFG GYLLRQTFELAYCTAATFSLAGPRFVSLDSTTFKNPVPVGSVL TMDSSISYTEHVHDGVEEIDADSPFNFSLPATNKISKNPEAFLS EPGTLIQVKVDTYIQELEQSVKKPAGTFIYSFYVDKESVTVDG KASFCSVIPQTYSEMMTYVGGRRRAQDTANYVETLPFSGSGN
SEQ	*E. coli* Acyl-	atggccgatacattgctcatcttgggtgactctttgtctgcagggtatcggatgtccgcatctgccgc

SEQ ID NO:	Description	Sequence
ID NO:37	CoA Thioesterase (TESA) gene without signal peptide sequence optimized for *C. tropicalis* Nucleotide Seq	atggcctgcactcctcaatgacaaatggcaaagcaagacatcggtcgtgaatgcatctatctctgg cgatacctcgcagcaggggttggcccgtctcccagccttgttgaagcaacatcaaccacgttgggt cttggtcgaattgggcggcaatgatggtctcagaggttttcaacctcaacagaccgagcagacatt gcgtcaaatcctccaagacgtgaaggcagcaaacgccgaacctctcttgatgcagataagattgc ctgccaactatggtcgtagatacaatgaagccttttctgcaatctacccgaagcttgcaaaggagttt gacgtcccattgttgccgtttttgatggaagaggtgtaccttaagcctcagtggatgcaagacgatg gtatccatccgaaccgtgatgcacaaccattcatcgcagattggatggccaaacaactccaacctttt ggtcaatcatgatagctaa
SEQ ID NO:38	*E. coli* Acyl-CoA Thioesterase (TESA) without signal peptide Amino Acid Seq	MADTLLILGDSLSAGYRMSASAAWPALLNDKWQSKTSVVNA SISGDTSQQGLARLPALLKQHQPRWVLVELGGNDGLRGFQPQ QTEQTLRQILQDVKAANAEPLLMQIRLPANYGRRYNEAFSAI YPKLAKEFDVPLLPFLMEEVYLKPQWMQDDGIHPNRDAQPFI ADWMAKQLQPLVNHDS
SEQ ID NO:39	Acyl-CoA Synthetase (ACSl) Nuc. Seq	atgggtgccccttaacagtcgccgttggcgaagcaaaaccaggcgaaaccgctccaagaagaa aagccgctcaaaaaatggcctctgtcgaacgcccaacagactcaaaggcaaccactttgccaga cttcattgaagagtgttttgccagaaacggcaccagagatgccatggcctggagagacttggtcga aatccacgtcgaaaccaaacaggttaccaaaatcattgacggcgaacagaaaaaggtcgataag gactggatctactacgaaatgggtccttacaactacatatcctaccccaagttgttgacgttggtcaa gaactactccaagggtttgttggagttgggcttggccccagatcaagaatccaagttgatgatctttg ccagtacctcccacaagtggatgcagaccttcttagcctccagtttccaaggtatccccgttgtcac cgcctacgacaccttgggtgagtcgggcttgacccactccttggtgcaaaccgaatccgatgccgt gttcaccgacaaccaattgttgtcctccttgattcgtcctttggagaaggccacctccgtcaagtatgt catccacggggaaaagattgaccctaacgacaagagacagggcggcaaaatctaccaggatgc ggaaaaggccaaggagaagattttacaaattagaccagatattaaatttatttctttcgacgaggttgt tgcattgggtgaacaatcgtccaaagaattgcatttcccaaaaccagaagacccaatctgtatcatg tacacctcgggttccaccggtgctccaaagggtgtggttatcaccaatgccaacattgttgccgcc gtgggtggtatctccaccaatgctactagagacttggttagaactgtcgacagagtgattgcatttttt

SEQ ID NO:	Description	Sequence
		gccattggcccacattttcgagttggcctttgagttggttaccttctggtgggggctccattgggtta cgccaatgtcaagactttgaccgaagcctcctgcagaaactgtcagccagacttgattgaattcaaa ccaaccatcatggttggtgttgctgccgtttgggaatcggtcagaaagggtgtcttgtctaaattgaa acaggcttctccaatccaacaaaagatcttctgggctgcattcaatgccaagtctactttgaaccgtt atggcttgccaggcggtgggttgtttgacgctgtcttcaagaaggttaaagccgccactggtggcc aattgcgttatgtgttgaatggtgggtccccaatctctgttgatgcccaagtgtttatctccaccttgctt gcgccaatgttgttgggttacggtttgactgaaacctgtgccaataccaccattgtcgaacacacgc gcttccagattggtactttgggtaccttggttggatctgtcactgccaagttggttgatgttgctgatgc tggatactacgccaagaacaaccagggtgaaatctggttgaaaggcggtccagttgtcaaggaat actacaagaacgaagaagaaaccaaggctgcattcaccgaagatggctggttcaagactggtgat attggtgaatggaccgccgacggtggtttgaacatcattgaccgtaagaagaacttggtcaagactt tgaatggtgaatacattgctttggagaaattggaaagtatttacagatccaaccacttgattttgaactt gtgtgtttacgctgaccaaaccaaggtcaagccaattgctattgtcttgccaattgaagccaacttga agtctatgttgaaggacgaaaagattatcccagatgctgattcacaagaattgagcagcttggttca caacaagaaggttgcccaagctgtcttgagacacttgctccaaaccggtaaacaacaaggtttgaa aggtattgaattgttgcagaatgttgtcttgttggatgacgagtggaccccacagaatggttttgttac ttctgcccaaaagttgcagagaaagaagattttagaaagttgtaaaaaagaagttgaagaggcata caagtcgtcttag
SEQ ID NO:40	Acyl-CoA Synthetase (ACS1) A. A. Seq	MGAPLTVAVGEAKPGETAPRRKAAQKMASVERPTDSKATTL PDFIEECFARNGTRDAMAWRDLVEIHVETKQVTKIIDGEQKK VDKDWIYYEMGPYNYISYPKLLTLVKNYSKGLLELGLAPDQE SKLMIFASTSHKWMQTFLASSFQGIPVVTAYDTLGESGLTHSL VQTESDAVFTDNQLLSSLIRPLEKATSVKYVIHGEKIDPNDKR QGGKIYQDAEKAKEKILQIRPDIKFISFDEVVALGEQSSKELHF PKPEDPICIMYTSGSTGAPKGVVITNANIVAAVGGISTNATRDL VRTVDRVIAFLPLAHIFELAFELVTFWWGAPLGYANVKTLTE ASCRNCQPDLIEFKPTIMVGVAAVWESVRKGVLSKLKQASPI QQKIFWAAFNAKSTLNRYGLPGGGLFDAVFKKVKAATGGQL RYVLNGGSPISVDAQVFISTLLAPMLLGYGLTETCANTTIVEH TRFQIGTLGTLVGSVTAKLVDVADAGYYAKNNQGEIWLKGG

SEQ ID NO:	Description	Sequence
		PVVKEYYKNEEETKAAFTEDGWFKTGDIGEWTADGGLNIIDR KKNLVKTLNGEYIALEKLESIYRSNHLILNLCVYADQTKVKPI AIVLPIEANLKSMLKDEKIIPDADSQELSSLVHNKKVAQAVLR HLLQTGKQQGLKGIELLQNVVLLDDEWTPQNGFVTSAQKLQ RKKILESCKKEVEEAYKSS
SEQ ID NO:41	Long-chain Acyl-CoA Synthetase (FAT1) Nuc. Seq	atgtcaggattagaaatagccgctgctgccatccttggtagtcagttattggaagccaaatatttaatt gccgacgacgtgctgttagccaagacagtcgctgtcaatgccctcccatacttgtggaaagccag cagaggtaaggcatcatactggtacttttcgagcagtccgtgttcaagaacccaaacaacaaagc gttggcgttcccaagaccaagaaagaatgcccccacccccaagaccgacgccgagggattcca gatctacgacgatcagtttgacctagaagaatacacctacaaggaattgtacgacatggttttgaag tactcatacatcttgaagaacgagtacggcgtcactgccaacgacaccatcggtgtttcttgtatgaa caagccgcttttcattgtcttgtggttggcattgtggaacattggtgccttgcctgcgttcttgaacttc aacaccaaggacaagccattgatccactgtcttaagattgtcaacgcttcgcaagttttcgttgaccc ggactgtgattccccaatcagagataccgaggctcagatcagagaggaattgccacatgtgcaaa taaactacattgacgagtttgccttgtttgacagattgagactcaagtcgactccaaaacacagagc cgaggacaagaccagaagaccaaccgatactgactcctccgcttgtgcattgatttacacctcggg taccaccggtttgccaaaagccggtatcatgtcctggagaaaagccttcatggcctcggttttctttg gccacatcatgaagattgactcgaaatcgaacgtcttgaccgccatgcccttgtaccactccaccg cggccatgttggggttgtgtcctactttgattgtcggtggctgtgtctccgtgtcccagaaattctccg ctacttcgttctggacccaggccagattatgtggtgccacccacgtgcaatacgtcggtgaggtctg tcgttacttgttgaactccaagcctcatccagaccaagacagacacaatgtcagaattgcctacggt aacgggttgcgtccagatatatggtctgagttcaagcgcagattccacattgaaggtatcggtgagt tctacgccgccaccgagtcccctatcgccaccaccaacttgcagtacggtgagtacggtgtcggc gcctgtcgtaagtacgggtccctcatcagcttgttattgtctacccagcagaaattggccaagatgg acccagaagacgagagtgaaatctacaaggaccccaagaccgggttctgtaccgaggccgctta caacgagccaggtgagttgttgatgagaatcttgaaccctaacgacgtgcagaaatccttccaggg ttattatggtaacaagtccgccaccaacagcaaaatcctcaccaatgttttcaaaaaaggtgacgcg tggtacagatccggtgacttgttgaagatggacgaggacaaattgttgtactttgtcgacagattag gtgacactttccgttggaagtccgaaaacgtctccgccaccgaggtcgagaacgaattgatgggc tccaaggccttgaagcagtccgtcgttgtcggtgtcaaggtgccaaaccacgaaggtagagcctg

SEQ ID NO:	Description	Sequence
		ttttgccgtctgtgaagccaaggacgagttgagccatgaagaaatcttgaaattgattcactctcacg tgaccaagtctttgcctgtgtatgctcaacctgcgttcatcaagattggcaccattgaggcttcgcac aaccacaaggttcctaagaaccaattcaagaaccaaaagttgccaaagggtgaagacggcaagg atttgatctactggttgaatggcgacaagtaccaggagttgactgaagacgattggtctttgatttgta ccggtaaagccaaattg
SEQ ID NO:42	Long-chain Acyl-CoA Synthetase (FAT1) A.A. Seq	MSGLEIAAAAILGSQLLEAKYLIADDVSLAKTVAVNALPYLW KASRGKASYWYFFEQSVFKNPNNKALAFPRPRKNAPTPKTDA EGFQIYDDQFDLEEYTYKELYDMVLKYSYILKNEYGVTANDT IGVSCMNKPLFIVLWLALWNIGALPAFLNFNTKDKPLIHCLKI VNASQVFVDPDCDSPIRDTEAQIREELPHVQINYIDEFALFDRL RLKSTPKHRAEDKTRRPTDTDSSACALIYTSGTTGLPKAGIMS WRKAFMASVFFGHIMKIDSKSNVLTAMPLYHSTAAMLGLCP TLIVGGCVSVSQKFSATSFWTQARLCGATHVQYVGEVCRYLL NSKPHPDQDRHNVRIAYGNGLRPDIWSEFKRRFHIEGIGEFYA ATESPIATTNLQYGEYGVGACRKYGSLISLLLSTQQKLAKMD PEDESEIYKDPKTGFCTEAAYNEPGELLMRILNPNDVQKSFQG YYGNKSATNSKILTNVFKKGDAWYRSGDLLKMDEDKLLYFV DRLGDTFRWKSENVSATEVENELMGSKALKQSVVVGVKVPN HEGRACFAVCEAKDELSHEEILKLIHSHVTKSLPVYAQPAFIKI GTIEASHNHKVPKNQFKNQKLPKGEDGKDLIYWLNGDKYQE LTEDDWSLICTGKAKL
SEQ ID NO:43	Acyl-CoA Sterol acyl transferase (ARE1) Nuc. Seq	atgtccgacgacgagatagcaggaatagtcattgaaatcgacgatgacgtgaaatccacgtcttcg ttccaggaagaactagtcgaggttgaaatgtccaactcgtccattaacgaatcccagaccgatgag tcgtaccgtcctgaagaaacctcattgcattacaggaggaagtcccacaggaccccgtcagagga gtcgttcctagagatcaccaagaacgtgaatgatccggatctagtttccaagattgagaacctaagg ggcaaagtaagccaacgggaagacaggttgaggaagcactaccttcacacctcccaggacgtc aagttcttgtcccggttcaacgacatcaagttcaagctgaactccgcgacgattctagattcggatg cgttttacaagagtgaatactttggagtcttgaccatcttctgggtggttatcgcactctacatattgtc aacgttgtcagatgtttactttggcatggccaagcccttactggactggatcatcataggaatgttca agcaggacttggtgaaagttgcactcgttgatcttgccatgtacctatcctcgtattttccttatttcttg

SEQ ID NO:	Description	Sequence
		caggttgcatgcaaacggggtgatgtatcttggcatggtcttggatgggcaatacaggggtttac agcttggtgtttttgacgttctggacggtagttccgcaggagttggccatggatcttccttggattgca cgaattttcttgatcttgcattgcttggtgtttattatgaagatgcagtcgtatgggcattacaatggata cctttgggatgtgtatcaggaaggattggcctctgaggctgatctcagggacctttctgagtatgatg aagatttccccctggatcacgtggaggttctagaacagagcttgtggtttgccaaacacgagttgga gtttcaatcgaatggaactgctgagaggaaggaccaccatcaccatgtattcgacgaaaaggatgt caacaaaccaatacgtgtcttgcaagaagagggaattatcaagtttccggcaaacatcaacttcaa ggattatttcgagtacagtatgttcccaacgctagtctacacgttgagcttcccccgaactcgacaga ttagatggacgtatgtgttgcagaaggttttgggaacatttgccttagtgtttgccatgattatcgtcgc cgaagagagtttctgccccttgatgcaagaagttgatcagtacacaaaattgccaaccaaccaaag gttcccaaaatacttcgtcgttctttcccacttgatattaccgctcggcaagcagtacttgctctcattc atcctcatctggaatgaaattctcaacggcatagcggagttaagcaggtttggcgaccggcatttct acggcgcttggtggtcgagcgtcgattacatggactattcaagaaaatggaacaccatcgtgcacc gattcctccgtcggcacgtttacaattcgagcattcacatcctcggtatttccaggacgcaagccgc gatagttacacttttgctttctgccacaatccacgaactcgttatgtacgtcctatttggcaaattacga gggtacctattccttacgatgcttgtccagatccccatgaccgtcacctccaagttcaacaaccgtgt ttggggcaacatcatgttctggttgacgtatttatctggccccagcttggttagtgcgttgtatttactct tctag
SEQ ID NO:44	Acyl-CoA Sterol acyl transferase (ARE1) A.A. Seq	MSDDEIAGIVIEIDDDVKSTSSFQEELVEVEMSNSSINESQTDE SYRPEETSLHYRRKSHRTPSEESFLEITKNVNDPDLVSKIENLR GKVSQREDRLRKHYLHTSQDVKFLSRFNDIKFKLNSATILDSD AFYKSEYFGVLTIFWVVIALYILSTLSDVYFGMAKPLLDWIIIG MFKQDLVKVALVDLAMYLSSYFPYFLQVACKRGDVSWHGL GWAIQGVYSLVFLTFWTVVPQELAMDLPWIARIFLILHCLVFI MKMQSYGHYNGYLWDVYQEGLASEADLRDLSEYDEDFPLD HVEVLEQSLWFAKHELEFQSNGTAERKDHHHHVFDEKDVNK PIRVLQEEGIIKFPANINFKDYFEYSMFPTLVYTLSFPRTRQIRW TYVLQKVLGTFALVFAMIIVAEESFCPLMQEVDQYTKLPTNQ RFPKYFVVLSHLILPLGKQYLLSFILIWNEILNGIAELSRFGDRH FYGAWWSSVDYMDYSRKWNTIVHRFLRRHVYNSSIHILGISR

SEQ ID NO:	Description	Sequence
		TQAAIVTLLLSATIHELVMYVLFGKLRGYLFLTMLVQIPMTVT SKFNNRVWGNIMFWLTYLSGPSLVSALYLLF
SEQ ID NO:45	Acyl-CoA Sterol acyl transferase (ARE2) Nuc. Seq	atgtccgacgacgagatagcaggaatagtcattgaaatcgacgatgacgtgaaatctacgtcttcg ttccaggaagacctagtcgaggttgagatgtccaactcgtccattaacgaatcccagacggatgag ttgtcgtaccgtcctgaagaaatctcattgcattcgagaaggaagtcccacaagaccccgtcagat gagtcgttcctagagatcaccaagaacgtgaatgatccggatctagtctccaagattgagaacttaa ggggcaaagtaagccaacgggaagacaggttgaggaaacactacctccacacatcccaggacg tcaagttcttgtctcggttcaacgacatcaagttcaagctgaactccgcgacgattctagattcggat gcgttttacaagagcgagcactttggagtcttgactatcttctgggtggttatcggactctacataatg tcaacgttgtcagacatgtattttggcatggccaagcccttactggactggataatcataggaatgtt caagaaggatttgatgcaagttgcactcgttgatcttgtcatgtacttatcctcgtattttccttatttcct acaggttgcatgcaagaccggagctatatcttggcatggtcttggatgggccatacaggggttta cagcttggtgtttttaactttctgggcggtacttccgctggagctggccatggatcttccttggattgc acgagttttcttgatcttgcattgcttggtgtttattatgaagatgcaatcatatggacattacaatggat acctttgggatgtatatcaggaaggattggtctcggaagctgatctcacggctgtttctgagtatgat gatgatttccccctggatcacggggaggttctagaacagagcttgtggttcgccaaacacgagttg gagtttcaatctaatggaactacggagaggaaggatcaccatcatcatgtattcgacgaaaaggat gtcaacaaaccaatgcgtgtcttgcaagaagagggaattatcaaatttccggcaaacatcaatttca aggattatttcgagtacagtatgttccccacgctagtctacacattgaacttccccagaattcgacata ttagatgggcgtatgtgttgcagaaagttttgggaacatttgccttagtgtttgccatgattatcgtcgc cgaagagagtttctgtcccttgatgcaagaagttgaacagtacacaagattgccaaccaaccaaag gttctcaaagtacttcgtcgttctttcccacttgatattgcccctcggcaaacagtacttgctctcgttta tcctcatttggaacgaaattctcaacgggatagcggagttaagcaggtttggggatcgccatttcta cggcgcctggtggtcaagcgtcgactacatggactattcaagaaaatggaacacgatcgtgcacc gattcctccgccggcacgtttacaattcgaccattcgcatcctcggtatttccaggacccaagccgc gataattacacttttgctttcagccacaatccacgaactcgttatgtacatcctatttggaaaattacga gggtacctattccttacgatgcttgtccagatccccatgacagtcaccgccaagttcaacaaccgttt gtggggcaacatcatgttctggttgacgtatttatctggccccagcttggttagtgcgttgtatttactc ttctga
SEQ	Acyl-CoA	MSDDEIAGIVIEIDDDVKSTSSFQEDLVEVEMSNSSINESQTDE

SEQ ID NO:	Description	Sequence
ID NO:46	Sterol acyl transferase (ARE2) A.A. Seq	LSYRPEEISLHSRRKSHKTPSDESFLEITKNVNDPDLVSKIENLR GKVSQPvEDRLRKHYLHTSQDVKFLSRFNDIKFKSNSATILDSD AFYKSEHFGVLTIFWVVIGLYIMSTLSDMYFGMAKPLSDWIII GMFKKDLMQVALVDLVMYLSSYFPYFLQVACKTGAISWHGL GWAIQGVYSLVFLTFWAVLPSESAMDLPWIARVFLILHCLVFI MKMQSYGHYNGYLWDVYQEGLVSEADLTAVSEYDDDFPSD HGEVLEQSLWFAKHELEFQSNGTTERKDHHHHVFDEKDVNK PMRVLQEEGIIKFPANINFKDYFEYSMFPTLVYTLNFPRIRHIR WAYVLQKVLGTFALVFAMIIVAEESFCPLMQEVEQYTRLPTN QRFSKYFVVLSHLILPLGKQYLLSFILIWNEILNGIAELSRFGDR HFYGAWWSSVDYMDYSRKWNTIVHRFLRRHVYNSTIRILGIS RTQAAIITLLLSATIHELVMYILFGKLRGYLFLTMLVQIPMTVT AKFNNRLWGNIMFWLTYLSGPSLVSALYLLF
SEQ ID NO:47	Diacylglycerol acyltransferase (DGA1) Nuc. Seq	atgactcaggactataaagacgatagtcctacgtccactgagttggacactaacatagaagaggtg gaaagcactgcaaccctagagtcggaactcagacagagaaaacagaccacggaaactccagca tcaaccccaccaccacctccacaacaacagcaggcgcataagaaagccctgaagaatggcaag aggaagagaccatttataaacgtggcgccgctcaacaccccgttggctcacaggctcgagacttt ggctgttgtttggcactgtgtcagtatcccgttctttatgtttttgttcttgcttacggtctccatggggtt gcttgggtggttctttatcattttgccatatttcatttggtggtacggtttcgacttgcacactccatcga atggtaaagttgtctatcgtgtgcgcaactcgttcaagaatttcatcatttgggactggtttgtcaagta tttcccgattgaagtgcacaagacggtcgagttggatcctacttttagcgaattgcctgtggaagag agcggcgacagttcggacgacgacgaacaagacttggtgtctgagcacagcagaactttggttg atcaaatcttcaagtttttcgggttgaagaaacgcttgaatgacacctccctgggcaaaccagagac attcaagaatgtgcctacgggtccaaggtatatttttgggtaccacccacacggagtgatttctatgg gggcagtggggttgtttgccaacaacgccttgaggaacgaaccatatacgccaatttccaaatggt taaaaccattcttccacgacagctccaagggcgagagattgttccctggtattggcaatatcttccca ttgacgcttaccacacagtttgcgctcccattttaccgtgactacttgatggctttggggatcactagt gcatcggctaaaaacattagaagcttgatcaacaatggagacaactctgtgtgtctcgtcgttggcg gtgcacaagaatcgttgttgaacaatatgattgccaagcacgccagagtcgggtacggttacaaag agagcctagatattcatggcgaccagtccgaagaagaagaagaagaagaggatgataccaagc

SEQ ID NO:	Description	Sequence
		agctagagaacccaagtcctaaacgtgaagtgcaattggtcttgaacaaacgtaaaggttttgtgaa gttggctatcgaactaggaaatgtttccttggtgcctatttttgcattcggagaagctgatgtttacaga ttggcccagccagcaccaggctcgttcttgtacaagttccagcaatggatgaaggcaacttttcaat tcaccatcccattgtttagtgctcgaggcgtgttcatctatgatttcggattgttgccattcagaaaccc aataaacatttgcgtcggtagacccgtctacattccgcacaacgtcttgcaagaatacaagcaaaa gcacccagaggagtttgccgaagaggaacctgccagtaccccgatgaagaagtctggatctttca ccgatatgttcaaagctggtgaaaagaagcccaagacttcaagtatcaagactaaaatcccacctg cattactagacaagtaccacaagctatacgtcgacgagttgaagaaggtctatgaagagaacaag gaaaggtttggctacggtgatgttgaattaaacattgtagaatag
SEQ ID NO:48	Diacylglycerol acyltransferase (DGA1) A.A. Seq	MTQDYKDDSPTSTELDTNIEEVESTATLESELRQRKQTTETPA STPPPPPQQQQAHKKASKNGKRKRPFINVAPLNTPLAHRLETL AVVWHCVSIPFFMFLFLLTVSMGLLGWFFIILPYFIWWYGFDL HTPSNGKVVYRVRNSFKNFIIWDWFVKYFPIEVHKTVELDPTF SELPVEESGDSSDDDEQDLVSEHSRTLVDQIFKFFGLKKRLND TSSGKPETFKNVPTGPRYIFGYHPHGVISMGAVGLFANNALR NEPYTPISKWLKPFFHDSSKGERLFPGIGNIFPLTLTTQFALPFY RDYLMALGITSASAKNIRSLINNGDNSVCLVVGGAQESLLNN MIAKHARVGYGYKESLDIHGDQSEEEEEEEDDTKQLENPSPK REVQLVLNKRKGFVKLAIELGNVSLVPIFAFGEADVYRLAQP APGSFLYKFQQWMKATFQFTIPLFSARGVFIYDFGLLPFRNPIN ICVGRPVYIPHNVLQEYKQKHPEEFAEEEPASTPMKKSGSFTD MFKAGEKKPKTSSIKTKIPPALLDKYHKLYVDELKKVYEENK ERFGYGDVELNIVE
SEQ ID NO:49	Diacylglycerol acyltransferase (LROl) Nuc. Seq	atgtcgtctttaaagaacagaaaatccgcaagcgtcgccacaagcgatacagaagactcagaaac agaggcagtatcctcctcaattgatcccaacggcaccatattgcgaccagtcctacatgacgaacc ccaccacagccatcaccaccacaacataactagaccagtattggaggacgatggcagcatcctg gtgtccagaagatcgtcgatctccaaatccgacgacctgcaggcaaagcaaaagaagaagaaac ccaagaagaagatcttggagtctcgtcgggtcatgtttatctttggtaccctcattgggttaatctttgc gtgggcgtttaccacagacacgcatcctttcaatggcgacttggagaagtttatcaactttgaccag ctcaacgggatctttgacgactggaagaactggaaggatatcttgcccaacagcatccagacgta

SEQ ID NO:	Description	Sequence
		cttgcaggaatcgggcaagggcgaagataacgacgggttgcatggtctggccgattccttctccg tcgggctccgcttgaaagcccagaagaacttcactgacaaccacaatgtcgtgttggttcctggtgt ggtgagcacggggttggaatcgtggggaacaaccaccaccggtgattgtccatctatcggatactt caggaagagattgtggggatcattttatatgttaaggacaatgattttggagaaaacgtgctggttga agcatatccagttggacgagaagacggggttggatcctcccaatattaaggtccgtgcggcgcag ggtttcgaagcggcagatttctttatggctgggtactggatctggaacaagatcttgcagaacttgg cggttattgggtacggaccaaataacatggtgagtgctagttatgactggagattggcttacattgac ttggagagaagagatggatattttcgaaacttaaagcgcagattgagttgaataacaagttgaaca acaagaagactgtgttgattggccactcgatggggacccagattattttctacttttgaaatgggtcg aagccaccgggaaaccatactatggcaatggcggaccaaactgggtgaatgatcatattgagtcg attattgacatcagtgggtcgactttgggtacccccaagagtattcctgtgttgatctctggggaaat gaaagacaccgttcaattgaacgcgttggcggtttacgggttggagcaatttttcagcaggcgtga aagagtcgatatgttgcgtacatttggtggcgttgccagtatgttacccaaggggggagacaagat atggggcaacttgacgcatgcgccagatgatccaatttccacattcagtgatgacgaagttacgga cagccacgaacctaaagatcgttcttttggtacgtttatccaattcaagaaccaaactagcgacgcta agccatacagggagatcaccatggctgaaggtatcgatgaattgttggacaaatcaccagactggt attccaagagagtccgtgagaactactcttacggcattacagacagcaaggcgcaattagagaag aacaacaatgaccacctgaagtggtcgaacccattagaagctgccttgcctaaagcacccgacat gaagatctattgtttctacggagttggaaatcctaccgaaagggcatacaagtatgtgactgccgat aaaaaagccacgaaattggactacataatagacgccgacgatgccaatggagtcatattaggaga cggagacggcactgtttcgttattaacccactcgatgtgccatgagtgggccaagggagacaagt cgagatacaacccagccaactcgaaggttaccattgttgaaatcaagcacgagccagacagattt gatttacgaggcggcgccaagactgcggaacatgttgatattttggggagtgccgagttgaacga gttgattttgactgtggttagcgggaacggggacgagattgagaatagatatgtcagcaacttaaaa gaaatagtagaggccataaatttataa
SEQ ID NO:50	Diacylglycerol acyltransferase (LROl) A.A. Seq	MSSLKNRKSASVATSDTEDSETEAVSSSIDPNGTILRPVLHDEP HHSHHHHNITRPVLEDDGSISVSRRSSISKSDDSQAKQKKKKP KKKILESRRVMFIFGTLIGLIFAWAFTTDTHPFNGDLEKFINFD QLNGIFDDWKNWKDILPNSIQTYLQESGKGEDNDGLHGSADS FSVGLRLKAQKNFTDNHNVVLVPGVVSTGLESWGTTTTGDC

SEQ ID NO:	Description	Sequence
		PSIGYFRKRLWGSFYMLRTMILEKTCWLKHIQLDEKTGLDPP NIKVPvAAQGFEAADFFMAGYWIWNKILQNLAVIGYGPNNMV SASYDWRLAYIDLERRDGYFSKLKAQIELNNKLNNKKTVLIG HSMGTQIIFYFLKWVEATGKPYYGNGGPNWVNDHIESIIDISG STLGTPKSIPVLISGEMKDTVQLNALAVYGLEQFFSRRERVDM LRTFGGVASMLPKGGDKIWGNLTHAPDDPISTFSDDEVTDSH EPKDRSFGTFIQFKNQTSDAKPYREITMAEGIDELLDKSPDWY SKRVRENYSYGITDSKAQLEKNNNDHSKWSNPLEAALPKAPD MKIYCFYGVGNPTERAYKYVTADKKATKLDYIIDADDANGVI LGDGDGTVSLLTHSMCHEWAKGDKSRYNPANSKVTIVEIKH EPDRFDLRGGAKTAEHVDILGSAELNELILTVVSGNGDEIENR YVSNLKEIVEAINL

SEQ ID NO:	Description	Sequence
SEQ ID NO: 3674	ECU, ATCC20336 (Nucleic Acid Seq.)	atgtccgacgaggaatcagatatcttatacgaggtcagagacagaaccgccatcatcaccttgaac atccccaagagattgaacgcattgaacggtgctcaatacttgaagttgggtaagttcttggagagag ccaacaacgaagaggacaccgtcttgacattgatccaggccctgggcagattcttctccgccggtg ccaatttcgccgacaacgatatggccaaggtcgaaatgtccaagttgttcagtcacgagtactggtt ggaaagattcgtcgccagaaacatctggttgaccaacttgttcaacgaccacaagaagatcttggct gctgctgtcaatggtccagttatcggtttgagcactggtttgttgttgttggtcgatttggtctacgtcca cgacttgaacaagttctacctcttggccccatttgccaacttgggtttggttgccgaaggtgcttcctct gccactttgttcaacagattgggctggtcaaaggcttctgaggccttgttgttggccaagccaatcgg cggccaagactgttacaacgccggtttcatcaacaagcactacgacggtaagttttcctccactgaa gagttcaacgaacacgtctacaaggagttgacggaagcttttgaaaacttgcacgatgactccatttt gcagaacaagcaattgttgaagttgtccagagaccaggccatcaactag
SEQ	ECU,	MSDEESDILYEVRDRTAIITLNIPKRLNALNGAQYLKLGKFLER

SEQ ID NO:	Description	Sequence
ID NO: 3675	ATCC20336 (Amino Acid seq.)	ANNEEDTVLTLIQASGRFFSAGANFADNDMAKVEMSKLFSHE YWLERFVARNIWLTNLFNDHKKILAAAVNGPVIGLSTGLLLLV DLVYVHDLNKFYLLAPFANLGLVAEGASSATLFNRLGWSKAS EALLLAKPIGGQDCYNAGFINKHYDGKFSSTEEFNEHVYKELT EAFENLHDDSILQNKQLLKLSRDQAIN*
SEQ ID NO: 3676	ECI2, ATCC20336 (Nucleic Acid Seq.)	atgtccgacgaccttatcacctacgaagtcaaagaccgagctgccgtgatcaccttgaacaacccc aagaagctcaacgccttatcgatcccgcagtacgacaccatctgcaagctcttagaacgagccaac gccgaagaagacaccgtcatcaccttgctccagtccacgggccgggtgttctctgccggggccaa cgccgactccatcgtggggcaggatgccgagctcgagacctggttgaacatgtcggtggccaag cagacgttcttggtgcagacgttcctcgcacacaagaagatccttgccgtcgccttgaacggcccc gtgattggcttatcggcggcgttcgtggcgctctgcgacttggtctacgtgcacaacgccgcaaag acgttcttcttgaccccgttcgccaacatcgggatccttgccgagggcggcacctcagccacgttgc ccatgcgcgtggggtggtccagggccgcggaagcgttgttgttgtcaaagaggatttcgggagat gacttgcagagagcggggttcttcaataaggactacaaggggcagttcaagtccgcggaggagttt aacgaggtcgtcttgaaggagttgcttgacgccacggaaaacttgcatgaggactcgatcatccag aacaaggagttgttgaaggctattttcaagccaaagatcagtgaggtcaactcgcaggaggtgtca agaggtgtgtacaagtggacctctggggtgccaatggatagatttaaaaaattgcttaatggtgagtt gaaacataaattatag
SEQ ID NO: 3677	ECI2, ATCC20336 (Amino Acid seq.)	MSDDLITYEVKDRAAVITLNNPKKLNALSIPQYDTICKLLERA NAEEDTVITLLQSTGRVFSAGANADSIVGQDAELETWLNMSV AKQTFLVQTFLAHKKILAVALNGPVIGLSAAFVALCDLVYVH NAAKTFFLTPFANIGILAEGGTSATLPMRVGWSRAAEALLLSK RISGDDLQRAGFFNKDYKGQFKSAEEFNEVVLKELLDATENL HEDSIIQNKELLKAIFKPKISEVNSQEVSRGVYKWTSGVPMDRF KKLLNGELKHKL*
SEQ ID NO: 3678	>gil50550800lr eflXM_502873. 1| Yarrowia lipolytica YALIOD15708	ATGTTGTCCATTCGATCCATTACCCGATCTCTCCCCATTGGC AGCCGAATCTGCCAGCAGAGTGCCATGAAGGCCTCTACTGT GCGACCTCTCGCCTTGAGAGCTTACTCCACCCGACCCCCTGT CACTCACTTCTCCGAGGAGGAGGAGATGTTTCGTGACATGG TTAGCAAGTTTGCTGATGAGGTGATTGCTCCCAAGGTCCGT

SEQ ID NO:	Description	Sequence
	P (YALIOD15708 g) mRNA, complete cds (similar to uniprotlP45954 Homo sapiens Acyl-CoA dehydrogenase short/branched chain specific mitochondrial precursor)	GAGATGGACGAGGCCGAGCAGATGGACAAGACAATCATCC AGGACATGTTCGACAATGGCCTTATGGGCATCGAGACTCCC GAGGAGTTCGGTGGTGCAGGTGCCAACTTCACCTCTGCTAT CATCGTCGTTGAGGAGCTTGCCAAGGTGGACCCCTCAGTGT CTGTGATGAACGATGTCCACAATACCCTCGTCAACACCTGC ATCCGATCCTGGGGATCCGACGCACTCCGAAACAAGTATCT CCCCCAGCTTGCTGCCCAGAAGGTCGGATCTTTCGCTCTTTC TGAGCCCTCTTCCGGATCTGATGCCTTCGCCATGAAGTCTCG AGCCACAAAGACTGACGATGGATACATTTTGAACGGTTCCA AGATGTGGATCACCAACGCTGCCGAGGCTGAGCTTTTCATT GTTTTTGCTAATCTCGATCCCAGCAAGGGCTACAAGGGTAT TACTGCCTTTGTTGTCGAGAAGGACATGGGAGTGCAGATTG CTAAGAAGGAGCAGAAGCTGGGTATCCGAGCCTCTTCTACC TGCGTTCTCAACTTCGAGGACGTTTTCATTCCTAAGGAGAA CCTTCTTGGCGAGGAGGGCAAGGGCTACAAGATTGCTATCG AGTGCTTGAACGAGGGCCGAATCGGAATTGCGGCCCAGAT GCTTGGCCTTGCTGGTGGAGCTTTCAAGAAGGCTACCGGCT ATGCTTTCAACGACAGAAAGCAGTTCGGCCAGTACATCGGT GAGTTCCAGGGTATGCAGCACCAGATTGGCCAGGCCGCTAC TGAGATCGAGGCTGCTCGACTCCTGGTCTACAACGCTGCCC GACTCAAGGAGGCTGGCGTTCCTTTCACAAAGGAGGCTGCT ATGGCAAAGCTCTATGCTTCCCAGGTTGCAGGAAACGTCGC ATCCAAGGCTGTCGAATGGATGGGTGGTGTCGGATTCACTC GAGAGGAGACTCTGGAGAAGTTCTTCCGAGATTCTAAGATC GGTGCCATTTACGAGGGAACTTCCAACATCCAGCTGCAGAC TATTGCCAAGATCATCCAGAAGGAGTCTGCCTAA
SEQ ID NO: 3679	>gil49648741|e mb|CAG81061. 1| YALIOD15708	MLSIRSITRSLPIGSRICQQSAMKASTVRPLALPvAYSTPvPPVTHF SEEEEMFRDMVSKFADEVIAPKVREMDEAEQMDKTIIQDMFD NGLMGIETPEEFGGAGANFTSAIIVVEELAKVDPSVSVMNDVH NTLVNTCIRSWGSDALRNKYLPQLAAQKVGSFALSEPSSGSDA

SEQ ID NO:	Description	Sequence
	p [Yarrowia lipolytica CLIB122](similar to uniprot|P45954 Homo sapiens Acyl-CoA dehydrogenase short/branched chain specific mitochondrial precursor)	FAMKSRATKTDDGYILNGSKMWITNAAEAELFIVFANLDPSK GYKGITAFVVEKDMGVQIAKKEQKLGIRASSTCandidaLNFEDV FIPKENLLGEEGKGYKIAIECLNEGRIGIAAQMLGLAGGAFKKA TGYAFNDRKQFGQYIGEFQGMQHQIGQAATEIEAARLLVYNA ARLKEAGVPFTKEAAMAKLYASQVAGNVASKAVEWMGGVG FTREETLEKFFRDSKIGAIYEGTSNIQLQTIAKIIQKESA
SEQ ID NO: 3680	>gi|210075528|ref]XM_501919.2| Yarrowia lipolytica YALI0C 16797 P (YALI0C 16797g) mRNA, complete cds (similar to uniprot|Q96VP9 Glomus intraradices Probable acyl-CoA dehydrogenase)	ATGAGCGAGCAGTACACCCCCGAACAAGTTGCGGAGCACA ACTCTCCCGAATCTCTGTGGATCATCATTGACGGTAACGTTT TCGACCTCACTGAATTCCAGAAAGAACACCCCGGCGGAAA AAAGATTCTCAAACGAGTCGCAGGAAAAGACGCTACCAAG TTTTTCCACAAATACCACGACGCCCCCAAGATTATGCGAAA GGTTGGACACAAGTTCAAGATCGGAACCCTTAAAGACGCTG AAGCAAACCCCACTCGAGCCATGATTGCCCCTAACAAGACC ACCGCCCTCGAGCCCTACGGAGACCTTGTCCCCTACGCCGA CCCCAACTGGTACCACGGCTACCACAACCCCTACTACAAGG AGTCCCACGCCAAGCTGCGTGACGAGGTCCGACAGTGGGTT GAGGAGAAGATTGAGCCCTTCGTTGAGGAATGGGATGAGG AGAAGGAGGTTCCCAAGGAGATCTTCCAGGAGATGGGCAA GCGAGGTTACCTTGCCGGCTCTCTCGGCACCCCCTACAAGG AGCTGGCCAAGTACACCAACGTCAAGCCCGCCTCTGTGCCC ATTGAGGAGTACGACATGTTCCACGAGCTCATCATCACCGA CGAGATCATGCGAGCTGGCTCCGGAGGTCTCACCTGGAACC TGCTTGGTGGCTACTGTATTGGTCTGCCTCCCGTGATCAAGT TCGCCAAGGAGCCCCTTAAGGAGCGAATCCTCCCCGGCCTG

SEQ ID NO:	Description	Sequence
		CTCGACGGTTCCAAGCGAATCTGTCTGTGTATCACTGAGCC CGACGCTGGCTCCGATGTTGCCAACATCACCACTACCGCCG AGAAGACCCCCGACGGAAAGTTCTACATTGTCAACGGTATC AAGAAGTGGATCACCAACGGTATCTGGGCTGACTACTTCAC TGTTGCCGTCCGAACCGGTGGCCCCGGCTCTGGCATGAACG GTATCTCTGTTCTGCTGCTCGAGCGAGGCATGGAGGGTCTT GAGACCCGACGAATGAACACTCAGGGTATGCTGTCTTCCGG CTCTACCTGGGTCACCATGGAGGATGTCAAGGTCCCCGTGG AGAACCTGCTCGGCAAGGAGAACAAGGGTTTCAAGGTCAT CATGACCAACTTCAACCACGAGCGAGTTGGTATCATCATCC AGGCAAACCGAGCTTCTCGAGTTTGCTACGAGGAGGCCTGC AAGTACGCCCACAAGCGAAAGACTTTCGGCAAGCCTCTGAT TGAGCACCCCGTCATCCGAGCCAAGCTCGCCAACATGGCCA TTCGAATCGAGTCCACCCACGCCTGGCTCGAGAACCTGGTC TTCCAGTGCCAGATGTTCCCCGAGGAGGAGGCCATGCTTCG ACTTGGTGGTGCCATTGCTGGTTGCAAGGCCCAGGCCACCC AGACCCTCGAGCTGTGTGCCCGAGAGGCTTCCCAGATCTTT GGTGGTCTTTCCTACACCCGAGGCGGTCTCGGAGGTAAGGT TGAGCGACTGTACCGAGAGGTCCGAGCCTACGCCATCCCCG GTGGATCCGAGGAGATTATGCTGGATCTGGCCATGCGACAG GCCCTCAAGGTCCACAAGGCTGTTGGCGCCAAGCTTTAA
SEQ ID NO: 3681	>gi|1994252921|emb|CAG82239.2|YALIOC16797p [Yarrowia lipolytica CLIB122](similar to uniprot|Q96VP	MSEQYTPEQVAEHNSPESLWIIIDGNVFDLTEFQKEHPGGKKIL KRVAGKDATKFFHKYHDAPKIMRKVGHKFKIGTLKDAEANPT RAMIAPNKTTALEPYGDLVPYADPNWYHGYHNPYYKESHAK LRDEVRQWVEEKIEPFVEEWDEEKEVPKEIFQEMGKRGYLAG SLGTPYKELAKYTNVKPASVPIEEYDMFHELIITDEIMRAGSGG LTWNLLGGYCIGLPPVIKFAKEPLKERILPGLLDGSKRICLCITE PDAGSDVANITTTAEKTPDGKFYIVNGIKKWITNGIWADYFTV AVRTGGPGSGMNGISVLLLERGMEGLETRRMNTQGMLSSGST WVTMEDVKVPVENLLGKENKGFKVIMTNFNHERVGIIIQANR

SEQ ID NO:	Description	Sequence
	9 Glomus intraradices Probable acyl-CoA dehydrogenase)	ASRVCYEEACKYAHKRKTFGKPLIEHPVIRAKLANMAIRIEST HAWLENLVFQCQMFPEEEAMLRLGGAIAGCKAQATQTLELC AREASQIFGGLSYTRGGLGGKVERLYREVPvAYAIPGGSEEIML DLAMRQALKVHKAVGAKL
SEQ ID NO: 3682	>gil50556785|ref|XM_505801.1| Yarrowia lipolytica YALI0F23749p (YALI0F23749g) mRNA, complete cds (highly similar to uniprot|Q7S579 Neurospora crassa NCU02291.1 hypothetical protein probable Glutaryl-CoA dehydrogenase)	ATGCTTACCAGAATCTCCCGTTTGGCACCTGCTGCCCGAGG CTTTGCTACCTCCTCCGTCAACCGATCCACAGCCGCCATGG ACTGGCAGGATCCCTTCCAGCTGGACTCTCTTCTCACCGAG GACGAGATTGCCGTGGCTGAGGCTGCTCGAGACTTCTGCCA GACAGAGCTCTACCCCAAGGTACTTGAGGGCTACCGAACCG AGGAGTTCCCCCGAAGCATCATGAAGCAGATGGGTGAGGTT GGTCTGCTCGGAACAACCGTCAAGAGCCACGGATGCCCCGG CATGTCTTCTGTCGCTTACGGTCTCGTGGCCCGAGAGGTCG AGAGGGTCGACTCCGGCTACCGATCTGCCATGTCTGTGCAG TCGTCGCTGGTCATGCACCCCATTGAACAGTTTGGATCCCA GGAGCAGAAGGACCGGTTCCTGCCCAAATTGGCCTCCGGCG AGATGATCGGCTGCTTCGGTCTCACCGAGCCTAACCACGGT TCCGACCCTGGATCCATGGAGACCGTCGCCAAGATGCACCC TACTAAGAAGGGCGTCATTGTGCTCAATGGAGCCAAGAACT GGATCACTAACTCTCCTATTGCCGATCTCATGGTTGTGTGGG CCAAGTTGGACGGTAAGATCCGAGGCTTCCTTGTCGAGCGA TCTCAGGTCGCCTCCGGCCTCGCTACTCCCGCCATCAAGAA CAAGACCGCTCTGCGAGCCTCCATCACCGGTATGATCCAGA TGGACGACGTTGAGATCCCTGTGGAGAACATGTTCCCCGAG GTGACCGGTCTCAAGGGCCCCTTCACCTGCCTCAACTCTGC CCGATACGGTATCGCCTGGGGAACCATGGGCGCTCTGTCCG AGTCCATCAAGCTCGCTCGAGAGTACTCTCTGGACCGAAAG CAGTTTAAGGGCCAGCCTCTGGCCAAGTACCAGCTCATCCA GAAGAAGCTCGCTGACGCTCTGACCGATGCCACCTACGGAC AGGTCGCTGCCATTCAGGTCGGCCGGCTCAAGGATGCCGGC

SEQ ID NO:	Description	Sequence
		AATTGTCCTCCCGAGCTCATCTCCATGATTAAGAGACAGAA CTGTGACCGAGCCCTCGCTGGCGCTCGAAACCTGATGGAGA TCTTTGGCGGTAACGCTGCCTCTGACGAGTACCACATTGGC CGAATTGCCGCCAACCTGTGGGTTGTCCAGACCTATGAGGG CCAGTCTGATATCCATGCTCTCATCCTGGGAAGAGCCATGA CCGGCGTCCAGGCTTTTGCTTAA
SEQ ID NO: 3683	>gi|50556786|ref|XP_505801.1 | YALI0F23749p [Yarrowia lipolytica](highly similar to uniprot|Q7S579 Neurospora crassa NCU02291.1 hypothetical protein probable Glutaryl-CoA dehydrogenase)	MLTRISRLAPAARGFATSSVNRSTAAMDWQDPFQLDSLLTEDE IAVAEAARDFCQTELYPKVLEGYRTEEFPRSIMKQMGEVGLLG TTVKSHGCPGMSSVAYGLVAREVERVDSGYRSAMSVQSSLV MHPIEQFGSQEQKDRFLPKLASGEMIGCFGLTEPNHGSDPGSM ETVAKMHPTKKGVIVLNGAKNWITNSPIADLMVVWAKLDGKI RGFLVERSQVASGLATPAIKNKTALRASITGMIQMDDVEIPVE NMFPEVTGLKGPFTCLNSARYGIAWGTMGALSESIKLAREYSL DRKQFKGQPLAKYQLIQKKLADALTDATYGQVAAIQVGRLK DAGNCPPELISMIKRQNCDRALAGARNLMEIFGGNAASDEYHI GRIAANLWVVQTYEGQSDIHALILGRAMTGVQAFA
SEQ ID NO: 3684	>gi|255723091|ref|XM_002546434.11 Candida tropicalis MYA-3404 conserved hypothetical protein, mRNA	ATGTCAGTCAAAGAAGATATCCCAGCTGTTTTTCTTTCCCAA ATTTCTCCTCGTGGTCTTGAAGCTATCCAGAAAACCAAAGA CTTTGTCAATGACTACTGTATTCCAGCCGATGAAATCTACTT CAAACAGGTCTCTACTGATCCTGCCAAAAGATGGAAAACAA TCCCACCTATTATTGAGACATTGAAATCCAAAGCCAAAGAA CTTGGTTTATGGAATATGTTTTTATCCAAACATTATAAGGAA GGTCCACAATATACCAACTTAGAATATGGTTTGATGGCTAG ATATTTGGGTCGTGCACACACTGCTCCTGAAGCTACTAATA

SEQ ID NO:	Description	Sequence
	(similar to C. albicans ACD99)	CTGCTGCTCCAGATACTGGTAACATGGAATTACTTGCTAAA TACGGTACTCCATACCAAAAAGAAAAGTACTTACAACCATT GTTAGATGGAAAGATCAGATCTGCTTTCTTGATGACCGAAA AGGGCACATCATCTTCCAATGCATTAAATATCTCCACTAGT GCCAAAAAGAATGCCAGTGGTAACTATGTTCTTGATGGTGT AAAGTGGTTTGCTTCAGGTGCTGGTGATCCAAGGTGTTCTG TTTGGTTGGTCATGTGTAAAACTGAAGACAATAAGAAGAAC CCATATGCAAACCACACCGTGTTGGTTCTTGATGCCAAGAG AGCATTGGCTAGCGGCAGGGCCAAATTAGTCAGACCTTTGC ATGTTATTGGATATGATGATGCTCCTCATGGTCATTGTGAAA TTTCTTTTGAAAACTACGAAGTTCCTGCTGACGAAATGCCA AATGCTGTTTTGGCCGGTATTGGAAGAGGATTTGAGTTGAT TCAGTCTAGATTAGGACCTGGTAGAATTCATCATTGTATGA GAGCTATTGGTACAGGTGAAATTGCATTGTTGATCATTGCT CATAGAGCTAACCACAGAATGATTTTTGGAAAACCAATGAA AGACAGAGAAGGATTTTTGTCTAAGTTCGGTCAGAGCAGAA TTGATATTACCAGATGCTTGTTATTGGTATTAAATGCTGCTC ATAAAATTGATATTTCCAACGCAAAGGCTGCTCAGAAAGAG ATTGCCATGGCTAAGATTGAAACACCAAGAACCATCTCTGA TATCCTTGACTGGGGTATCCAAGTTTTTGGCGCAGAAGGGG TCTCACAAGACACAGACTTAGCTAGAATGTATGCTCTCAAC AGAACCTTGAGAATTGCTGATGGTCCTGATGAAGCTCACTT GGCACAATTGGCAAGAAATGAGGCCAAAAAATTCCCAGAG GTCGATATCTTCTTTGAACATGTTGCTAGTCAACGTAATAAA TTATAG
SEQ ID NO: 3685	>gi|240130997|gb|EER30559.1| conserved hypothetical protein	MSVKEDIPAVFLSQISPRGLEAIQKTKDFVNDYCIPADEIYFKQ VSTDPAKRWKTIPPIIETLKSKAKELGLWNMFLSKHYKEGPQY TNLEYGLMARYLGRAHTAPEATNTAAPDTGNMELLAKYGTP YQKEKYLQPLLDGKIRSAFLMTEKGTSSSNALNISTSAKKNAS GNYVLDGVKWFASGAGDPRCSVWLVMCKTEDNKKNPYANH

SEQ ID NO:	Description	Sequence
	[Candida tropicalis MYA-3404] (similar to C. albicans ACD99)	TVLVLDAKRALASGRAKLVRPLHVIGYDDAPHGHCEISFENYE VPADEMPNAVLAGIGRGFELIQSRLGPGPvIHHCMRAIGTGEIAL LIIAHRANHRMIFGKPMKDREGFLSKFGQSRIDITRCLLLVLNA AHKIDISNAKAAQKEIAMAKIETPRTISDILDWGIQVFGAEGVS QDTDLARMYALNRTLRIADGPDEAHLAQLARNEAKKFPEVDI FFEHVASQRNKL
SEQ ID NO: 3686	XP_716423.1| probable acyl-CoA dehydrogenase [Candida albicans SC5314]	MSVKEDIPAVFLEKVSPRGLEAIQKTKDFVNDYCLPADQIYFE QLSDIPSERWKSVPPVIETLKKKAKELGLWNMFLSKHYKEGPQ YTNLEYGLMARYLGRSYTAPEATNTAAPDTGNMELFAKYGT TYQKDRYLKPLLNGEIRSAFLMTEKGVSSSNALNISTSAVKNS NGNYVLNGVKWFASGAGDPRCSVWLVMCKTDNNKQNPYQN HTVLIIDAKKALATGKAKLIRPLQVIGFDDAPHGHCEIQFQDYE VPADEMPNVVMAGVGRGFELIQSRLGPGRIHHCMRAIGAGEF ALLRIAHRANHRLIFGKPMNQREGFLSRYGQSKIDIERCLLLVL NAAHKIDISNAKEAQKEIAMAKIETPRTISDILDWGIQVYGAEG MSQDTELARMYAHNRTLRIADGPDEAHLAQLARNEAKKFAK VDDFFANMETQRSKL
SEQ ID NO: 3687	XM_71 1330.1| Candida albicans SC5314 probable acyl-CoA dehydrogenase (ACD99) mPvNA, complete cds	ATGTCAGTTAAAGAAGACATTCCTGCTGTTTTCCTTGAAAA GGTTTCTCCTCGTGGTCTCGAAGCCATCCAGAAAACCAAAG ATTTCGTTAACGATTATTGTCTTCCAGCCGATCAAATTTATT TTGAACAACTTTCAGACATCCCATCAGAAAGATGGAAGAGT GTTCCTCCTGTCATTGAGACATTGAAGAAGAAAGCCAAGGA ACTTGGTTTATGGAACATGTTTTTGTCAAAGCATTATAAGG AAGGTCCACAATATACAAACTTAGAGTATGGATTGATGGCC AGATACTTGGGTCGTTCATACACAGCACCAGAGGCTACCAA CACAGCTGCTCCAGATACCGGTAATATGGAATTGTTTGCCA AATATGGAACCACTTATCAGAAAGATAGATACTTGAAACCC TTGTTAAATGGGGAAATTAGATCAGCATTCTTGATGACAGA AAAAGGTGTTTCATCATCTAATGCTCTCAATATTTCTACAAG TGCTGTCAAGAATTCGAATGGAAATTACGTGCTCAATGGTG

SEQ ID NO:	Description	Sequence
		TCAAATGGTTTGCTTCAGGTGCAGGAGATCCAAGATGTTCC GTCTGGTTGGTGATGTGCAAGACAGACAACAACAAGCAAA ATCCATATCAAAACCACACAGTTTTGATCATCGATGCCAAA AAGGCTTTGGCTACTGGAAAAGCCAAATTGATCAGACCATT GCAGGTCATTGGTTTTGATGATGCTCCTCATGGACATTGTGA GATTCAATTTCAAGATTACGAAGTTCCTGCCGATGAAATGC CTAATGTTGTTATGGCTGGTGTTGGTAGAGGATTTGAGTTG ATTCAATCCAGATTGGGTCCAGGTAGAATCCACCATTGTAT GAGAGCTATTGGTGCTGGTGAATTTGCATTATTGAGAATTG CTCACAGAGCAAATCACAGATTGATTTTTGGTAAACCTATG AACCAGAGGGAAGGATTCTTATCCAGATACGGACAAAGCA AAATCGACATTGAAAGATGTTTATTGTTGGTGTTGAATGCT GCTCACAAAATTGATATTTCCAATGCCAAAGAAGCACAAAA GGAAATTGCTATGGCTAAGATTGAGACCCCGAGAACTATCT CTGATATTCTCGATTGGGGTATTCAAGTTTATGGTGCTGAGG GTATGTCACAAGATACCGAGTTGGCCAGAATGTATGCTCAT AACAGAACATTGAGAATAGCTGATGGACCTGATGAAGCTC ATTTGGCCCAATTGGCTAGAAATGAAGCTAAAAAGTTTGCA AAAGTTGACGACTTTTTCGCCAACATGGAAACTCAACGTAG CAAATTATAA
SEQ ID NO: 3688	acyl-CoA dehydrogenase NM domain-like protein [Candida tenuis ATCC 10573] (similar to C. albicans ACD99)	ATGACAGACCTTGACATTCCAGCAGTATTCCTTGATAAGAT CTCACCACGTGGCCTCGAGGCGATCCGCAAGACCTACGACT TTGTGCATAACTACTGTATTCCTGCGGATGCTCTCTACTTTG ACCAAATTTCCCAGGATCCCGAACAAAGGTGGAAAACCACT CCTGAAGTCACTGAAAAATTGAAACAAAAGGCCAAACAAT TAGGTTTGTGGAACATGTTCCTCTCTAAGCACTATACCGATG GACCTGGCTACACAAACTTGGAGTATGGCCTTATGGCGCAA TTCTTGGGCCGGTCGTTCGTGGCACCCGAGGCCACCAATAC AGGTGCACCCGATACAGGTAACATGGAGATTCTCGCCAAGT TCGGCTCGGCCTATCACCGGGAGCAGTACCTCCTTCCATTG

SEQ ID NO:	Description	Sequence
		CTCCGCGGTGAGATCCGCTCGGCGTTCTTGATGACAGAAAA AGGCACTTCTTCATCCAATGCCTTGAACATCTCATGCTCGGC CCAGAAGAATTCACACGGCAACTACGTTCTCAATGGAGTCA AGTGGTTTGCCTCTGGTGCAGGTGATCCTCGGTGTCGCGTGT GGTTGGTGATGTGCAAAACCGAGTCTCTGGACAACATCTAC CGTAATCACAGTGTGTTGGTGTTGGATGCGAAAAAGGCTTT AGCTTCAGGAAAAGCCAAATTGATCCGACCACTCAGCGTGT TTGGCTATGACGATGCTCCTCATGGACACTGTGAGGTGGAG TTCAACGACTTTGAGGTGCCAGCCGAGGATATGGATAATTC TATCCTTGGTAAGGTGGGTATGGGATTTGAGATCATCCAGT CTCGTTTAGGCCCTGGGCGTATTCACCACTGTATGCGTCTTA TTGGTGCCGGAGAATATGCCTTAATGAGGGCGGTGCTGAGG GCTGCCGGCAGAGACATTTTCGGCAAGCCCATGGTGAAGAG AGAATCATTTCTCAATGCTTATGGAGAGCATAAGCTTTCAC TTCAGAAATGCCGTCTTTTGGTGCTTAATGCAGCTCATCAAA TCGATATTTCGAATGCTAAGACTGCCAAAAGAGATATAGCC ATGGCCAAAATCGAGACTCCCAGAGCAGTATTGAAGATTCT TGACTGGTGTATTCAGGTTTATGGGGCTGAAGGAGTGTCTC AAGACACAGAGCTTGCAAAGATGTATGCTCACGCTCGGACT TTGAGAATCGCAGATGGACCAGATGAAGCACACCTTGGAC AGCTTGCACGGGACGAGTCAAAGAAGTTTGCGGAGGTGGT GAAGTACTTTGAGGGACACAAGGCACGTCAAGACCAAGTC CTGAAGTTGTGA
SEQ ID NO: 3689	>gi|344233800|gb|EGV65670.1| acyl-CoA dehydrogenase NM domain-like protein [Candida tenuis	MTDLDIPAVFLDKISPRGLEAIRKTYDFVHNYCIPADALYFDQI SQDPEQRWKTTPEVTEKLKQKAKQLGLWNMFLSKHYTDGPG YTNLEYGLMAQFLGRSFVAPEATNTGAPDTGNMEILAKFGSA YHREQYLLPLLRGEIRSAFLMTEKGTSSSNALNISCSAQKNSHG NYVLNGVKWFASGAGDPRCRVWLVMCKTESSDNIYRNHSVL VLDAKKALASGKAKLIRPLSVFGYDDAPHGHCEVEFNDFEVP AEDMDNSILGKVGMGFEIIQSRLGPGRIHHCMRLIGAGEYALM

SEQ ID NO:	Description	Sequence
	ATCC 10573](similar to C. albicans ACD99)	RAVSRAAGRDIFGKPMVKRESFLNAYGEHKLSLQKCRLLVLN AAHQIDISNAKTAKRDIAMAKIETPRAVLKILDWCIQVYGAEG VSQDTELAKMYAHARTLRIADGPDEAHLGQLARDESKKFAEV VKYFEGHKARQDQVSKL
SEQ ID NO: 3690	>gil50309254lref|XM_454634.1| Kluyveromyces lactis NRRL Y-1140 hypothetical protein partial mRNA (similar to C. albicans ACD99)	ATGCCTAATGTCAGTGATAGACCGCGGACATATAAGAAACC TGCTTTAGAAGATGTTGATCCCATCACAAACTATATACCTG CCAGTGTTAGGGATAAATTTGATGAGAGGCAGATGGATCGG TTCAAGAAGTTGCGGAAATTTGTTGAGTTTGAATGTTTGCC ATTAGATACGGTGTATTTGCAAGAGAGTACCCTATTTGAGC ATGAAAGCGATTTAGAGACGTGCCCAGTCATTATTAATTTA AGGAAGAAATTGGAGGCATACCAGTTGCATAAAATGTTTGT TCCAATGGATCAACGTGGGTACGACCATAGTTTCAACGATA ATTGGGAAGTGGTGAGTATGGTTGAATTTGCTATGATCGCT TTCCTTGCTGGAAGATCTGTCATTGCCAGTTATTTGTTCCAT TTGGATGATTTGATCGATTTAGGAACTATACAAGTTTTGTTG AGAAATGGTTGTTCGAACCATGATTTGTGGGTACAAGTGAT AGATGAGTTAGTTTCTAATAATATGAAATCGTGTTTGATGG TAAGTGAAAGAGATGTGTCTGGTTCTGATGCGTTGAACGTT CAAACCACCTGTAAAATTGAAGGGGATGATCTAAACGAAG AGGAGGCTACTATGACACTTAACGGTACTAAATGGTTTATC AAAGATGCAGGAGACTCAGATATTTGGTTAGTTTTATGTGT CACTGAATTTGATGAGGGCAACATTTATAGAAAACATACAT TATGCCTTGTTAACAGGAATGATTTACCACCAAATTCAACA AGAATTGAACCTATAGAAACAAATGAAGCGATTGGTAAATT TTATGAAGTACAATTTAAAGATTGTAAAGTACCGTTAAATA TTATTGGTGAAAGAGGTGAAGGTTATCAAATTTTACAAATG AAATCCTCTGTTACAAAATTATTTCAATGCTTAAAACTTTGT GGTATGGGACAAGAATCCTTGAGACTTTCCAATAAGAGAGC TGCTGAAAGGAAAGTGTTTGGTTCCAAATTACAGAAGAGTG AGTATTTCAAATTTGATCTTGCTCATTGGAGGATTAAGATTG

SEQ ID NO:	Description	Sequence
		AAACCTGTAAGCTGCTTTGTTTCAACGCGGCAATCAAATGT GATTACGAAGGTGTAAAAGCGGCAAGAGAAGAAATTGGGA TGGTGAAAGCCGTGACACCAAAGGAAATCTCGTCACTGGTG GATTGGTCTATCCAGTTGCATGGATGTTACGGACTCTGTTCA ACACAAACACCCTTGTCACATATGTGGCAAGTGAGTCGATC GCTAAGAATTAATGATACGCCGGACGAATCATTAATATCAC AACTGGGGAGGTTGGAAATCAGTAATTATAACAAATTTCAA AAGACATACGATCAAGAATTAACGACGCTCGCTGGCAAAT GA
SEQ ID NO: 3691	>gi|49643769|emb|CAG99721.1| KLLA0E15181P [Kluyveromyces lactis] (similar to C. albicans ACD99)	MPNVSDRPRTYKKPALEDVDPITNYIPASVRDKFDERQMDRFK KLRKFVEFECLPLDTVYLQESTLFEHESDLETCPVIINLRKKLE AYQLHKMFVPMDQRGYDHSFNDNWEVVSMVEFAMIAFLAG RSVIASYLFHLDDLIDLGTIQVLLRNGCSNHDLWVQVIDELVS NNMKSCLMVSERDVSGSDALNVQTTCKIEGDDLNEEEATMTL NGTKWFIKDAGDSDIWLVLCandidaTEFDEGNIYRKHTLCLVN RNDLPPNSTRIEPIETNEAIGKFYEVQFKDCKVPLNIIGERGEGY QILQMKSSVTKLFQCLKLCGMGQESLRLSNKRAAERKVFGSK LQKSEYFKFDLAHWRIKIETCKLLCFNAAIKCDYEGVKAAREE IGMVKAVTPKEISSLVDWSIQLHGCYGLCSTQTPLSHMWQVSR SLRINDTPDESLISQLGRLEISNYNKFQKTYDQELTTLAGK
SEQ ID NO: 3692	>gi|301507715|gb|GU338397.1| Candida rugosa propionyl-CoA dehydrogenase mPvNA, complete cds (similar to C. albicans	ATGTCGATTAAGGACGACATCCCTGCCATCTTTTACGAAAA ACTTTCCCCCCGCGGGCTTGAGGCTATCGCCAAAACCAAGG AATTCGTCGACACTTACTGCTCCCCCGCCGACGAGATCTAC TTCCAACAGGTGAGAACTGACGACCGCCGGTGGAAGGAAA CGCCCCCCATCACCGAGCACTTGAAGAAGAAAGCTAAAGA GCTCGGGTTATGGAACATGTTCTTGCTGAAGCACTACGCCG AGGGCGCCGGCTACACCAACTTGGAGTATGGGCTTATGGCC CAGTACCTTGGCCGCAGTCACATCGCCCCTGAAGCTACCAA CACCAATGCTCCTGACACCGGCAACATGGAGATCCTTGCCA AGTACGGCAACGACTACCACAAGCAGCGCTACCTCCAGCCG

SEQ ID NO:	Description	Sequence
	ACD99)	CTTCTCGACGGTAAAATCCGCCTGGCGTTCTTAATGACGGA AAAGGGGACGTCGCTGTCCAACGCCCTTAACATCTCGTGCC TGGCAAAACTTAACCAAAATGGCAACTACGTCATCAACGGC GTCAAGTGGTTCGCCCTGGGTGCCGGCGACCCCCGGTGCAA GGTGTGGTTGACGATGTGCAAGACCAGCGACGACGACGCC AACCCATATTTCAACCACTCGTTGCTTGTGCTTGATGTCGAC AAGGCCCTCGCCCTGGGACAGGCTCGTGTTGTCCGCCCGTT GCACGTGTTTGGCTACGACGACGCTCCTCACGGTCACTGTG AAATTGAATTTAACAACTACGAAGTGTCCAAAGAGGAAAT GGCCAACGTCATCCTCGGCCAGGTGGGCCAAGGATTTGCCA TCATCCAGCTGAGATTGGGGCCGGGGCGCATCCACCACTGC ATGCGGATGATTGGCGTCGGCGAATTCGCCTTGATGAGAGT GGCTCAGCGGGCTAACCACCGTATCATCTTCGGTAAGCCCA TGGCCAAGCGCGAACTGTTTTTGAACGCCTACGCTCAGGCA AAGATCGACATCCAAAAGTGCCGCTTGTTTGTTCTTAATGC CGCCCACCACATCGACATTGCCGGAGCCAAAGCGGCGCAA GCCGACATCGCCATGGCCAAGATCGAGACCCCGAGAACCA TCCTTCGCATCTTGGACTGGGGGATCCAGATGTTTGGCGCC GAAGGGGTGTCTCAAGACACCGAGCTCTCGCGCATGTACGC GTTGGGGCGGACGTTACGCATTGCCGACGGCCCCGATGAAG CTCACTTGGGCCAATTGGCCCGTAAGGAGCTGAAGAAGTTC CCTTACGTCGATGAGTACTTTAAGCGGTTTGAAGAAAATAA GGCGAAGTTGGCCAAGTTGTAA
SEQ ID NO: 3693	>gi|301507716|gb|ADK77878.1| propionyl-CoA dehydrogenase [Candida rugosa] (similar	MSIKDDIPAIFYEKLSPRGLEAIAKTKEFVDTYCSPADEIYFQQV RTDDRRWKETPPITEHLKKKAKELGLWNMFLSKHYAEGAGY TNLEYGLMAQYLGRSHIAPEATNTNAPDTGNMEILAKYGNDY HKQRYLQPLLDGKIRSAFLMTEKGTSSSNALNISCSAKLNQNG NYVINGVKWFASGAGDPRCKVWLTMCKTSDDDANPYFNHSL LVLDVDKALASGQARVVRPLHVFGYDDAPHGHCEIEFNNYEV SKEEMANVILGQVGQGFAIIQSRLGPGRIHHCMRMIGVGEFAL

SEQ ID NO:	Description	Sequence
	to C. albicans ACD99)	MRVAQRANHRIIFGKPMAKRESFLNAYAQAKIDIQKCRLFVLN AAHHIDIAGAKAAQADIAMAKIETPRTILRILDWGIQMFGAEG VSQDTELSRMYALGRTLRIADGPDEAHLGQLARKESKKFPYV DEYFKRFEENKAKLAKL
SEQ ID NO: 3694	>gil380353348: 214809-216140 Candida orthopsilosis Co 90-125, chromosome 4 draft sequence (similar to C. albicans ACD99)	ATGTCAGTTAAAGACGATATCCCAGCTATCTTTTTAGATAA GGTTTCTCCAAGAGGTCTTGAAGCAATTCAAAAGACAAAGG ACTTTGTCGACCAATATTGTATCCCTGCTGATAAGATTTTCA AGGAGCAAATTTCGCAAGACCCAAAAATAAGATGGAAACA ATATCCAGCTATCATTGAACCATTGAAGAAAAAGGCTAGAG AGTTGGGTTTGTGGAACATGTTTTTGTCCAAGCATTACAAA GAGGGTCCTCAATTTACCAATTTGGAATACGGATTAATGGC TAGGTATTTGGGAAGATGTCACACTGGACCAGAAGCAACCA ACACCAGTGCCCCAGACACAGGTAATATGGAATTGTTTGCT AAATATGGTACAAAGGCGCAAAAGGATAAGTATTTAGTGC CCTTGATGGATGGTAAGATCAGATCGGCATTCTTGATGACC GAAAAGGGGATTTCATCGTCGAATGCATTAAACATTTCAAC CACTGCCATTAAGAATGCCCGTGGTAACTATGTGTTGAATG GAACAAAGTGGTTTGCCTCTGGTGCTGGAGATCCAAGAACT GCTGTTTGGTTGGTTATGTGCAAAACAGACAATGATGAAAG TAATATGTTCAGAAACCACTCCGTGTTAGTCATTGATGTCA AGCATGCATTAGCATCAGGTAAGGCTGAAGTTATCAGGCCT TTGAGTATTTTTGGCTACGATGATGCACCCCATGGTCATTGT GAAATCGTTTTCAAGGATTATGAAGTTTCATCTGAATTGAT GCCAGAAACGATTTTGGCCGGTGTCGGTAGGGGATTTGAAT TGATTCAATCCCGTTTGGGTCCAGGTAGAATCCATCATTGTA TGAGAGCCATAGGTGCTGGTGAATTTGCCTTGTTGCGTATT GCTCACAGAGCAAATCACAGAACCATCTTTGGTAGGCCAAT GAATAGAAGAGAAGGCTTCTTGATGCAGTATGCCAAGTACA GAATTGAAATTCAAAAATGTTTATTATTGGTTTTGAATGCTG CTCACAAGATTGACATCACTAATGCCAAACATGCACAAAGA

SEQ ID NO:	Description	Sequence
		GAAATTGCCATGGCTAAAATTGAGACTCCAAAAACAATTTG CGATATTCTCGACTGGGGTATTCAAGTCTTTGGAGCCGAAG GATTCTCTCAAGATACAGAATTGGCACAAATGTATGCTTGG AATAGAACTTTGAGAATCGCTGATGGTCCTGATGAAGCACA TTTGGCTCAATTGTCAAGAAGAGAAGCTGCCAAGTTTCCAG AAGTTGATGAGTTTTTCAAGAGTGTTGAATCAAGAGTTGAA GCTATTAGTAAGTTATAA
SEQ ID NO: 3695	>gil380353467|emb|CCG22977.1| hypothetical protein CORT_0D01290 [Candida orthopsilosis](similar to C. albicans ACD99)	MSVKDDIPAIFLDKVSPRGLEAIQKTKDFVDQYCIPADKIFKEQ ISQDPKIRWKQYPAIIEPLKKKARELGLWNMFLSKHYKEGPQF TNLEYGLMARYLGRCHTGPEATNTSAPDTGNMELFAKYGTK AQKDKYLVPLMDGKIRSAFLMTEKGISSSNALNISTTAIKNAR GNYVLNGTKWFASGAGDPRTAVWLVMCKTDNDESNMFRNH SVLVIDVKHALASGKAEVIRPLSIFGYDDAPHGHCEIVFKDYEV SSELMPETILAGVGRGFELIQSRLGPGRIHHCMRAIGAGEFALL RIAHRANHRTIFGRPMNRREGFLMQYAKYRIEIQKCLLLVLNA AHKIDITNAKHAQREIAMAKIETPKTICDILDWGIQVFGAEGFS QDTELAQMYAWNRTLRIADGPDEAHLAQLSRREAAKFPEVDE FFKSVESRVEAISKL
SEQ ID NO: 3696	>gil354545630:225012-226343 Candida parapsilosis strain CDC317 annotated contig 005809(similar to C. albicans ACD99)	ATGTCAGTTAAGGACGATATTCCAGCAATCTTTTTAGATAA GGTTTCCCCAAGAGGTCTTGAAGCTATTCAAAAGACAAAAG ACTTTGTTGAGCAATACTGTATTCCTGCCGATAAAGTTTTCA AGAAACAGATTTCGACAGACCCAGCGGTAAGATGGAAACA ATACCCTGCTATTATTGAACCATTGAAGAAAAAGGCTAGGG AATTGGGATTGTGGAACATGTTTTTGTCCAAGCATTACAAA GAGGGTCCTCAATTTACCAACTTGGAATATGGATTGATGGC TAGGTATCTAGGAAGATGCCACACTGGTCCTGAAGCCACTA ACACTAGTGCACCAGACACGGGTAATATGGAGTTGTTTGCA AAATATGGTACAAAGGCGCAAAAAGACAAATATTTGGTGC CCTTGATGGATGGTAAGATTAGATCAGCATTTTTGATGACT GAAAAGGGGATCTCATCGTCCAATGCGTTGAACATTTCCAC

SEQ ID NO:	Description	Sequence
		CACTGCAATTAAAAACTCACGTGGAAACTATGTCTTGAATG GTACCAAGTGGTTTGCATCAGGCGCTGGTGATCCTAGAACT GCCGTTTGGTTGGTTATGTGTAAGACTGCCAACGATGAAAA GAATGCATTTAAAAACCACTCAGTATTAGTGATTAATGTTA AGCATGCATTAGCATCAGGCAAGGCTGAAGTTATTAGACCT TTGGGAATTTTCGGATACGACGATGCTCCTCATGGACATTG TGAAATTGTTTTCAAAGATTATGAAGTTTCATCAGAGTTGAT GCCAGATACCATTTTGGCTGGTGTTGGTAAAGGATTCGAAT TGATTCAATCTAGATTGGGCCCGGGTAGAATCCATCATTGT ATGAGAGCTATTGGTGCTGGTGAATTTGCATTGTTGCGTATC GCCCACAGAGCTAATCACAGAATTATTTTTGGTAAACCAAT GAATAGAAGAGAAGGCTTTTTGATGCAGTATGCCAAGTACA GAATCGAGATTCAAAAATGTTTATTGTTGGTTTTAAATGCTG CCCACAAGATAGATATCACTAATGCCAAAGAAGCTCAAAG AGAAATTGCAATGGCCAAGATTGAAACTCCAAAAACCATTT GTGATATTCTTGATTGGGGTATTCAAGTGTTTGGAGCTGAG GGTTTCTCTCAGGATACAGAATTGGCGCAAATGTACGCTTG GAACAGAACTTTGAGAATTGCAGATGGACCAGATGAAGCA CATTTGGCTCAATTAGCAAGAAGAGAAGCCGCAAAGTTCCC TGACGTTGACGTGTTTTTTAAAGATGTTGATTCAAGAGTTGA GGCTGTTAGTAAATTATAA
SEQ ID NO: 3697	>gi|354545753|emb|CCE42481.1| hypothetical protein CPAR2_201240 [Candida parapsilosis](similar to C. albicans	MSVKDDIPAIFLDKVSPRGLEAIQKTKDFVEQYCIPADKVFKK QISTDPAVRWKQYPAIIEPLKKKARELGLWNMFLSKHYKEGP QFTNLEYGLMARYLGRCHTGPEATNTSAPDTGNMELFAKYGT KAQKDKYLVPLMDGKIRSAFLMTEKGISSSNALNISTTAIKNSR GNYVLNGTKWFASGAGDPRTAVWLVMCKTANDEKNAFKNH SVLVINVKHALASGKAEVIRPLGIFGYDDAPHGHCEIVFKDYE VSSELMPDTILAGVGKGFELIQSRLGPGRIHHCMRAIGAGEFAL LRIAHRANHRIIFGKPMNRREGFLMQYAKYRIEIQKCLLLVLN AAHKIDITNAKEAQREIAMAKIETPKTICDILDWGIQVFGAEGF

SEQ ID NO:	Description	Sequence
	ACD99)	SQDTELAQMYAWNRTLRIADGPDEAHLAQLARREAAKFPDV DVFFKDVDSRVEAVSKL
SEQ ID NO: 3698	>gil241959309| ref]XM_002422 **329.11** Candida dubliniensis CD36 acyl-CoA dehydrogenase, putative (CD36_34410) mRNA, complete cds	ATGTCAGTTAAAGAAGATATTCCTGCTATTTTCCTTGAAAA GATTTCCCCTCGTGGTCTTGATGCTATCCAGAAAACCAAAG ATTTCGTAAACGATTATTGTCTTCCAGCAGATCAGATCTATT TTGAGCAGCTCTCTGACATCCCTTCAGAAAGATGGAAAAGT GTTCCTCCTGTCATTGAGACATTGAAGAAGAAAGCCAAGGA ACTTGGTTTATGGAACATGTTTTTGTCAAAGCATTATAAAG AAGGTCCACAATACACAAACTTAGAGTATGGGTTGATGGCC AGATACTTGGGTCGTTCATACACTGCGCCAGAGGCTACCAA TACTGCTGCTCCAGATACCGGTAATATGGAATTGTTTGCCA AATATGGTACCACTTATCAGAAAGATAGATACTTGAAACCC TTGTTAAATGGGGAAATCAGATCGGCATTCTTGATGACCGA AAAGGGTGTTTCATCATCCAATGCTCTCAATATTTCTACAAG CGCTATCAAGAACTCTAATGGTAATTACGTGCTCAATGGTG TCAAATGGTTTGCTTCAGGGGCAGGAGATCCAAGATGCTCT GTATGGTTGGTAATGTGCAAGACCGACAACAATAAGCAAA ACCCTTATCAGAACCACACTGTTTTGATTATCGATGCAAAA AAGGCTTTGGCTACCGGAAAAGCCAAATTGATCAGACCATT GCAAGTCATTGGTTTTGATGATGCTCCCCATGGACATTGTG AAATCCAATTCAAAGACTACGAAGTTCCTGCTGATGAAATG CCTAATGTTGTAATGGCAGGTGTTGGTAGAGGATTTGAGTT GATTCAATCCAGATTGGGTCCAGGTAGAATCCACCATTGTA TGAGAGCTATTGGTTCTGGTGAATTTGCTTTATTAAGAATTG CTCATAGAGCAAATCACAGATTAATTTTTGGTAAGCCCATG AACCAAAGAGAGGGGTTCTTATCCAGATACGGACAAAGCA AAATAGATATTGAAAGATGTTTGTTATTGGTGTTGAATGCC GCTCACAAAATCGATATTTCCAATGCCAAAGAGGCACAAAG GGAAATTGCTATGGCCAAGATTGAAACCCCAAGAACTATTT CTGATATTCTCGATTGGGGTATTCAAGTTTATGGAGCTGAA

SEQ ID NO:	Description	Sequence
		GGTATGTCTCAAGACACTGAGTTGGCCAGAATGTATGCCCA TAACAGAACATTGAGAATAGCTGATGGACCTGATGAAGCTC ATTTGGCTCAATTGGCTAGAAACGAAGCTAAGAAGTTTCCA AAAGTTGACGCCTTCTTTACCAACATGGAAACACAACGTAG CAAATTATAA
SEQ ID NO: 3699	>gil241959310|ref]XP_002422374.11 acyl-CoA dehydrogenase, putative [Candida dubliniensis CD36]	MSVKEDIPAIFLEKISPRGLDAIQKTKDFVNDYCLPADQIYFEQ LSDIPSERWKSVPPVIETLKKKAKELGLWNMFLSKHYKEGPQY TNLEYGLMARYLGRSYTAPEATNTAAPDTGNMELFAKYGTT YQKDRYLKPLLNGEIRSAFLMTEKGVSSSNALNISTSAIKNSNG NYVLNGVKWFASGAGDPRCSVWLVMCKTDNNKQNPYQNHT VLIIDAKKALATGKAKLIRPLQVIGFDDAPHGHCEIQFKDYEVP ADEMPNVVMAGVGRGFELIQSRLGPGRIHHCMRAIGSGEFAL LRIAHRANHRLIFGKPMNQREGFLSRYGQSKIDIERCLLLVLNA AHKIDISNAKEAQREIAMAKIETPRTISDILDWGIQVYGAEGMS QDTELARMYAHNRTLRIADGPDEAHLAQLARNEAKKFPKVD AFFTNMETQRSKL
SEQ ID NO: 3700	>gil126138209|ref]XM_001385591.11 Scheffersomyces stipitis CBS 6054 acetyl-coenzyme-A dehydrogenase partial mRNA	ATGTCCGCCAAAGACGATATCCCTGCCATTTTCTTGGACAA GATCTCTCCCAGAGGTCTTGAGGCCATTGAGAAGACCAAAC GTTTCGTGGAAGACTACTGTTTGCCAGCTGACGATATCTACT TCAAGCAGATCAAGACCGATCCCGCAGTTAGATGGAAATAT ACTCCCGAAATCACGGAAAAGTTGAAGAAGAAAGCAAAGG AACTCGGGCTCTGGAACATGTTCTTGTCTAAGCACTACAAG GAAGGACCCCAGTTCACTAACTTGGAGTACGGGTTGATGGC TGAGTACTTGGGCAAATCCTTTGTTGCTCCAGAGGCTACCA ACACTGCAGCTCCAGATACCGGAAACATGGAACTTTTTGCC AAATACGGAACTCCATACCAAAAGGAGAAGTGGCTCAAGC CATTGTTGAACGGAGAAATCAGATCAGCTTTCTTGATGACA GAGAAGGGTGTTTCTTCATCGAATGCCTTGAACATTTCGAC TAGTGCCATTAAGAACGCCCAAGGCAACTACGTTCTTAACG GTGTCAAGTGGTTTGCTTCTGGAGCTGGAGATCCCAGATGT

SEQ ID NO:	Description	Sequence
		TCAGTCTGGCTTGTCATGTGTAAAACCACCGACGACTCCAG CAAGCCATACTTCAACCATTCTGTCTTGATTTTAGATCCCAA AGTCGCTATTGCTTCTGGAAAAGCCAGGGTGGTCAGACCTT TGCATGTGATTGGGTACGACGATGCGCCCCATGGCCATTGT GAAATCGAGTTCACCAACTACGAGGTTTCAGCTGAAGAAAT GAAGAACACCATTCTTGCTGGTGTTGGCCGTGGTTTTGAGC TCATCCAGTCCCGTTTGGGACCAGGCAGAATCCATCACTGT ATGAGACTGATTGGTTCTGGCGAGTTTGCTTTGCTCAAGAC AGCACACAGAGCCAACAACAGAATCATCTTTGGCAAGCCCT TGGCCAATAGAGAGTCCTTTATCACAGCTTTTGCTCAACAT AAGATCGACATTCAGAAGTGTCGTTTGTTGGTGTTGAACGC GGCCCACAAGATTGACATCACCAATGCCAAGGGTGCCCAG AAGGAAATTGCCATGGCAAAGATCGAGACTCCAAGGACAG TGTGCAAGATCATAGATTGGGGCATGCAAATGTTTGGTGCC GAAGGGTTATCTCAAGACACTGAGCTTGCCAGAATTTATGC CATGACCAGAATATTGAGAATTGCCGACGGTCCAGATGAAG CTCATTTGAACCAGTTAGGTAGAAACGAAGCAAAGAAATTC AACGAGGCTGATGCCTTCTTTGCTACCTATGAGGCAAGCAG AGCCAGATTGGAAAAATTGTAG
SEQ ID NO: 3701	>gi|150866135|ref|XP_001385628.2| acetyl-coenzyme-A dehydrogenase [Scheffersomyces stipitis CBS 6054]	MSAKDDIPAIFLDKISPRGLEAIEKTKRFVEDYCLPADDIYFKQI KTDPAVRWKYTPEITEKLKKKAKELGLWNMFLSKHYKEGPQF TNLEYGLMAEYLGKSFVAPEATNTAAPDTGNMELFAKYGTPY QKEKWLKPLLNGEIRSAFLMTEKGVSSSNALNISTSAIKNAQG NYVLNGVKWFASGAGDPRCSVWLVMCKTTDDSSKPYFNHSV LILDPKVAIASGKARVVRPLHVIGYDDAPHGHCEIEFTNYEVSA EEMKNTILAGVGRGFELIQSRLGPGRIHHCMRSIGSGEFALLKT AHRANNRIIFGKPLANRESFITAFAQHKIDIQKCRLLVLNAAHK IDITNAKGAQKEIAMAKIETPRTVCKIIDWGMQMFGAEGLSQD TELARIYAMTRILRIADGPDEAHLNQLGRNEAKKFNEADAFFA TYEASRARLEKL

SEQ ID NO:	Description	Sequence
SEQ ID NO: 3702	>gil 1464229291 ref]XM_001487 349.11 Meyerozyma guilliermondii ATCC 6260 hypothetical protein (PGUG_00776) partial mRNA	ATGTCTGTTAAAGAGGATATTCCGGCTATTTTTCTCGACAAG ATTTC GCC CAAAGG ATTGG ACGC GATC CA GAAATGT AAGG A TTTTGTCGAGCAATACTGTCTTCCGGCGGATAAAATATACCT AGAGCAGCTTAGCCCTGACCCCACAAAAAGATGGAAATCT ACCCCACAAATCACTGAAAAATTGAAGAAAAAAGCCCAAG AATTGGGACTTTGGAACATGTTCTTGTCAAAACACTATGCT GAGGGTGCAGGGTACACCAACTTGGAATATGGGCTCATGGC AGGTTATTTAGGGCGGTCGTTGGTGGCCCCAGAAGCAACCA ATACCAATGC ACCCGACACGGGC AAT ATGG AATTGCTTGC C AAATACGGCACTCAGTACCATAAAGAACGTTGGCTCAAGCC ATTGTTGAACGGAGAGATTCGGTCGGCTTTTTTGATGACGG AAAAGGGTACTTCTTCGTCTAATGCGTTGAACATTTCTGTTT CGGCCAAGAAAAATGCCAATGGGAATTGGGTATTGAATGG TATTAAGTGGTTCGCTTCTGGATCAGGAGACCCACGGTGTT CAGTTTGGTTGGTAATGTGCAAAACAGCCGAAACTAAGGCG ATTTATGAAAACCACTCGGTTCTCGTTATCGATGCCAAAAA GGCATTGGCTACAGGAAATGCCAAATTGATCCGGCCATTAC ATGTTTTTGGCTATGACGATGCTCCTCACGGACACTTTGAGG TGGAATTCAACAACTATGAGATTCCAAGTGAAGATATGCCC CATTCCATATTGGCTTCTGAAGGTAGAGGATTCGAGCTCAT TCAGTCGAGACTTGGTCCTGGTCGTATCCACCACTGTATGA GACTGATTGGTGCTGGAGAACAAGCGTTGTTGCGCGTGAGC CATCGTGCCAACAATCGGCTCATTTTCGGTACACTTATGGC AAAGAGAGAATCATTTATTACTGCATTTGCCCAGCACAAGA TCAACCTTCAGAAATGTAGATTGCTCGTTTTGAATGCTGCCC ACAAAATTG ACATC AGT AATGC CAAAC AGGC ACAAC GGGA GATTGCTATGGCCAAAATTGAGACTCCAAGAACCGTTGGTA GGGTACTTGACTGGGGTATCCAAATGTTTGGAGCAGAGGGA GTTTCGCAAGACACCGAATTGGCTCGTCTGTATGCTATCAA CCGGACACTCCAGATTGCTGATGGCCCCGACGAAGCTCATT

SEQ ID NO:	Description	Sequence
		TGAACCAATTGGGATTGAAAGAGGCCAAGAAATTTGCACTT GCAAGTGAATTCTTTGCTCAACAAGAAGAATACCGCAAACG ATTATCTAACCTCTAG
SEQ ID NO: 3703	>gil 1464229301 ref]XP_001487 **399.11** hypothetical protein PGUG_00776 [Meyerozyma guilliermondii ATCC 6260]	MSVKEDIPAIFLDKISPKGLDAIQKCKDFVEQYCLPADKIYLEQ LSPDPTKRWKSTPQITEKLKKKAQELGLWNMFLSKHYAEGAG YTNLEYGLMAGYLGRSLVAPEATNTNAPDTGNMELLAKYGT QYHKERWLKPLLNGEIRSAFLMTEKGTSSSNALNISVSAKKNA NGNWVLNGIKWFASGSGDPRCSVWLVMCKTAETKAIYENHS VLVIDAKKALATGNAKLIRPLHVFGYDDAPHGHFEVEFNNYEI PSEDMPHSILASEGRGFELIQSRLGPGRIHHCMRLIGAGEQALL RVSHRANNRLIFGTLMAKRESFITAFAQHKINLQKCRLLVLNA AHKIDISNAKQAQREIAMAKIETPRTVGRVLDWGIQMFGAEG VSQDTELARLYAINRTLQIADGPDEAHLNQLGLKEAKKFALAS EFFAQQEEYRKRLSNL
SEQ ID NO: 3704	OAA2835	CACACAGCTCTTCCATAATGTCCGACGAGGAATCAGA
SEQ ID NO: 3705	oAA2836	CACACAGCTCTTCCCTCTCTTCTATTCCTAGTTGATGGCCTG GTCTC
SEQ ID NO: 3706	OAA2837	CACACAGCTCTTCCATAATGTCCGACGACCTTATCAC
SEQ ID NO: 3707	OAA2838	CACACAGCTCTTCCCTCTCTTCTATTCCTATAATTTATGTTTC AACTCACC
SEQ ID NO: 3708	OAA3085	ATCGTTACCACCATCCCTACAAT
SEQ ID NO: 3709	oAA3086	CCGAAACAACCGTAGATACCTTTAAGCTACAACACTATACA CGATAATTCCC
SEQ ID NO: 3710	oAA3087	GGGAATTATCGTGTATAGTGTTGTAGCTTAAAGGTATCTAC GGTTGTTTCGG

SEQ ID NO:	Description		Sequence
SEQ ID NO: 371 1		oAA3088	CTTGGACATTTCGACCTTGGCGGTACCGAGCTCTGCGAATT
SEQ ID NO: 3712		oAA3089	AATTCGCAGAGCTCGGTACCGCCAAGGTCGAAATGTCCAAG
SEQ ID NO: 3713		oAA3090	GCTTGTTCTGCAAAATGGAGTCA
SEQ ID NO: 3714		OAA3212	GGGGGAGATCGTTACCACCA
SEQ ID NO: 3715		OAA3213	AATTCGCAGAGCTCGGTACCGCTGCTGCTGCTGCTGTTTT
SEQ ID NO: 3716		OAA3214	AAAACAGCAGCAGCAGCAGCGGTACCGAGCTCTGCGAATT
SEQ ID NO: 3717		OAA3215	TTCGTTGTTGGCTCTCTCCATTAAAGGTATCTACGGTTGTTT CGG
SEQ ID NO: 3718		OAA3216	CCGAAACAACCGTAGATACCTTTAATGGAGAGAGCCAACA ACGAA
SEQ ID NO: 3719		OAA3217	CAAAGGCATCGGTCAACTCC
SEQ ID NO: 3720	Candida strain ATCC20336 POX4		ATGACTTTTACAAAGAAAAACGTTAGTGTATCACAAGGTCC TGACCCTAGATCATCCATCCAAAAGGAAAGAGACAGCTCCA AATGGAACCCTCAACAAATGAACTACTTCTTGGAAGGCTCC GTCGAAAGAAGTGAGTTGATGAAGGCTTTGGCCCAACAAAT GGAAAGAGACCCAATCTTGTTCACAGACGGCTCCTACTACG ACTTGACCAAGGACCAACAAAGAGAATTGACCGCCGTCAA GATCAACAGAATCGCCAGATACAGAGAACAAGAATCCATC GACACTTTCAACAAGAGATTGTCCTTGATTGGTATCTTTGAC CCACAGGTCGGTACCAGAATTGGTGTCAACCTCGGTTTGTT CCTTTCTTGTATCAGAGGTAACGGTACCACTTCCCAATTGAA

SEQ ID NO:	Description	Sequence
		CTACTGGGCTAACGAAAAGGAAACCGCTGACGTTAAAGGT ATCTACGGTTGTTTCGGTATGACCGAATTGGCCCACGGTTCC AACGTTGCTGGTTTGGAAACCACCGCCACATTTGACAAGGA ATCTGACGAGTTTGTCATCAACACCCCACACATTGGTGCCA CCAAGTGGTGGATTGGTGGTGCTGCTCACTCCGCCACCCAC TGTTCTGTCTACGCCAGATTGATTGTTGACGGTCAAGATTAC GGTGTCAAGACTTTTGTTGTCCCATTGAGAGACTCCAACCA CGACCTCATGCCAGGTGTCACTGTTGGTGACATTGGTGCCA AGATGGGTAGAGATGGTATCGATAACGGTTGGATCCAATTC TCCAACGTCAGAATCCCAAGATTCTTTATGTTGCAAAAGTT CTGTAAGGTTTCTGCTGAAGGTGAAGTCACCTTGCCACCTTT GGAACAATTGTCTTACTCCGCCTTGTTGGGTGGTAGAGTCA TGATGGTTTTGGACTCCTACAGAATGTTGGCTAGAATGTCC ACCATTGCCTTGAGATACGCCATTGGTAGAAGACAATTCAA GGGTGACAATGTCGATCCAAAAGATCCAAACGCTTTGGAAA CCCAATTGATAGATTACCCATTGCACCAAAAGAGATTGTTC CCATACTTGGCTGCTGCCTACGTCATCTCCGCTGGTGCCCTC AAGGTTGAAGACACCATCCATAACACCTTGGCTGAATTGGA CGCTGCCGTTGAAAAGAACGACACCAAGGCTATCTTTAAGT CTATTGACGACATGAAGTCATTGTTTGTTGACTCTGGTTCCT TGAAGTCCACTGCCACTTGGTTGGGTGCTGAAGCCATTGAC CAATGTAGACAAGCCTGTGGTGGTCACGGTTACTCGTCCTA CAACGGCTTCGGTAAAGCCTACAACGATTGGGTTGTCCAAT GTACTTGGGAAGGTGACAACAATGTCTTGGCCATGAGTGTT GGTAAGCCAATTGTCAAGCAAGTTATCAGCATTGAAGATGC CGGCAAGACCGTCAGAGGTTCCACCGCTTTCTTGAACCAAT TGAAGGACTACACTGGTTCCAACAGCTCCAAGGTTGTTTTG AACACTGTTGCTGACTTGGACGACATCAAGACTGTCATCAA GGCTATTGAAGTTGCCATCATCAGATTGTCCCAAGAAGCTG CTTCTATTGTCAAGAAGGAATCTTTCGACTATGTCGGCGCTG

SEQ ID NO:	Description	Sequence
		AATTGGTTCAACTCTCCAAGTTGAAGGCTCACCACTACTTGT TGACTGAATACATCAGAAGAATTGACACCTTTGACCAAAAG GACTTGGTTCCATACTTGATCACCCTCGGTAAGTTGTACGCT GCCACTATTGTCTTGGACAGATTTGCCGGTGTCTTCTTGACT TTCAACGTTGCCTCCACCGAAGCCATCACTGCTTTGGCCTCT GTGCAAATTCCAAAGTTGTGTGCTGAAGTCAGACCAAACGT TGTTGCTTACACCGACTCCTTCCAACAATCCGACATGATTGT CAATTCTGCTATTGGTAGATACGATGGTGACATCTATGAGA ACTACTTTGACTTGGTCAAGTTGCAGAACCCACCATCCAAG ACCAAGGCTCCTTACTCTGATGCTTTGGAAGCCATGTTGAA CAGACCAACCTTGGACGAAAGAGAAAGATTTGAAAAGTCT GATGAAACCGCTGCTATCTTGTCCAAGTAA
SEQ ID NO: 3721	POX4	GTTCACTGCCATATGACTTTTACAAAGAAAAACGTTAGTGT ATCACAAGG
		Candida strain ATCC20336, Fwd. Primer, Ndel
SEQ ID NO: 3722	POX4	CTTCGAGATGCGGCCGCTTACTTGGACAAGATAGCAGCGGT TTCATC
		Candida strain ATCC20336, Rev. Primer, Notl
SEQ ID NO: 3723	Candida strain ATCC20336	ATGCCTACCGAACTTCAAAAAGAAAGAGAACTCACCAAGTT CAACCCAAAGGAGTTGAACTACTTCTTGGAAGGTTCCCAAG AAAGATCCGAGATCATCAGCAACATGGTCGAACAAATGCA AAAAGACCCTATCTTGAAGGTCGACGCTTCATACTACAACT TGACCAAAGACCAACAAAGAGAAGTCACCGCCAAGAAGAT TGCCAGACTCTCCAGATACTTTGAGCACGAGTACCCAGACC AACAGGCCCAGAGATTGTCGATCCTCGGTGTCTTTGACCCA CAAGTCTTCACCAGAATCGGTGTCAACTTGGGTTTGTTTGTT TCCTGTGTCCGTGGTAACGGTACCAACTCCCAGTTCTTCTAC
	POX5	

SEQ ID NO:	Description	Sequence
		TGGACCATAAATAAGGGTATCGACAAGTTGAGAGGTATCTA TGGTTGTTTTGGTATGACTGAGTTGGCCCACGGTTCCAACGT CCAAGGTATTGAAACCACCGCCACTTTTGACGAAGACACTG ACGAGTTTGTCATCAACACCCCACACATTGGTGCCACCAAG TGGTGGATCGGTGGTGCTGCGCACTCCGCCACCCACTGCTC CGTCTACGCCAGATTGAAGGTCAAAGGAAAGGACTACGGT GTCAAGACCTTTGTTGTCCCATTGAGAGACTCCAACCACGA CCTCGAGCCAGGTGTGACTGTTGGTGACATTGGTGCCAAGA TGGGTAGAGACGGTATCGATAACGGTTGGATCCAGTTCTCC AACGTCAGAATCCCAAGATTCTTTATGTTGCAAAAGTACTG TAAGGTTTCCCGTCTGGGTGAAGTCACCATGCCACCATCTG AACAATTGTCTTACTCGGCTTTGATTGGTGGTAGAGTCACC ATGATGATGGACTCCTACAGAATGACCAGTAGATTCATCAC CATTGCCTTGAGATACGCCATCCACAGAAGACAATTCAAGA AGAAGGACACCGATACCATTGAAACCAAGTTGATTGACTAC CCATTGCATCAAAAGAGATTGTTCCCATTCTTGGCTGCCGCT TACTTGTTCTCCCAAGGTGCCTTGTACTTAGAACAAACCATG AACGCAACCAACGACAAGTTGGACGAAGCTGTCAGTGCTG GTGAAAAGGAAGCCATTGACGCTGCCATTGTCGAATCCAAG AAATTGTTCGTCGCTTCCGGTTGTTTGAAGTCCACCTGTACC TGGTTGACTGCTGAAGCCATTGACGAAGCTCGTCAAGCTTG TGGTGGTCACGGTTACTCGTCTTACAACGGTTTCGGTAAAG CCTACTCCGACTGGGTTGTCCAATGTACCTGGGAAGGTGAC AACAACATCTTGGCCATGAACGTTGCCAAGCCAATGGTTAG AGACTTGTTGAAGGAGCCAGAACAAAAGGGATTGGTTCTCT CCAGCGTTGCCGACTTGGACGACCCAGCCAAGTTGGTTAAG GCTTTCGACCACGCCCTTTCCGGCTTGGCCAGAGACATTGG TGCTGTTGCTGAAGACAAGGGTTTCGACATTACCGGTCCAA GTTTGGTTTTGGTTTCCAAGTTGAACGCTCACAGATTCTTGA TTGACGGTTTCTTCAAGCGTATCACCCCAGAATGGTCTGAA

SEQ ID NO:	Description	Sequence
		GTCTTGAGACCTTTGGGTTTCTTGTATGCCGACTGGATCTTG ACCAACTTTGGTGCCACCTTCTTGCAGTACGGTATCATTACC CCAGATGTCAGCAGAAAGATTTCCTCCGAGCACTTCCCAGC CTTGTGTGCCAAGGTTAGACCAAACGTTGTTGGTTTGACTG ATGGTTTCAACTTGACTGACATGATGACCAATGCTGCTATT GGTAGATATGATGGTAACGTCTACGAACACTACTTCGAAAC TGTCAAGGCTTTGAACCCACCAGAAAACACCAAGGCTCCAT ACTCCAAGGCTTTGGAAGACATGTTGAACCGTCCAGACCTT GAAGTCAGAGAAAGAGGTGAAAAGTCCGAAGAAGCTGCTG AAATCTTGTCCAGTTAA
SEQ ID NO: 3724	POX5	GTTCACTGCCATATGCCTACCGAACTTCAAAAAGAAAGAGA ACTC
		Candida strain ATCC20336, Fwd. Primer, Ndel
SEQ ID NO: 3725	POX5	CTTCGAGATGCGGCCGCTTAACTGGACAAGATTTCAGCAGC TTCTTCG
		Candida strain ATCC20336, Rev. Primer, Notl
SEQ ID NO: 3726	Acol (AJ001299.1)	GTTCACTGCCATATGACAACCAACACATTCACCGATCCTC
		Yarrowia lipolytica, Fwd. Primer, Ndel
SEQ ID NO: 3727	Acol (AJ001299.1)	CTTCGAGATCTCGAGTCACTCATCGAGATCGCAAATTTCAT CGTC
		Yarrowia lipolytica, Rev. Primer, Xhol
SEQ ID NO:	Aco2 (XM_505264)	GTTCACTGCCATATGAACCCCAACAACACTGGCACC
		Yarrowia lipolytica, Fwd. Primer, Ndel

SEQ ID NO:	Description	Sequence
3728		
SEQ ID NO: 3729	Aco2 (XM_505264)	CTTCGAGATGCGGCCGCCTATTCCTCATCAAGCTCGCAAATGTCATC
		Yarrowia lipolytica, Rev. Primer, NotI
SEQ ID NO: 3730	Aco3 (XM_503244)	GTTCACTGCCATATGATCTCCCCCAACCTCACAGCTAAC
		Yarrowia lipolytica, Fwd. Primer, NdeI
SEQ ID NO: 3731	Aco3 (XM_503244)	CTTCGAGATGCGGCCGCCTATTCCTCGTCCAGCTCGCAAATG
		Yarrowia lipolytica, Rev. Primer, NotI
SEQ ID NO: 3732	Aco4 (XM_504475)	GTTCACTGCCATATGATCACCCCAAACCCCGCTAAC
		Yarrowia lipolytica, Fwd. Primer, NdeI
SEQ ID NO: 3733	Aco4 (XM_504475)	CTTCGAGATCTCGAGTTACTGAATATCCTCGGGCTCCATGG
		Yarrowia lipolytica, Rev. Primer, XhoI
SEQ ID NO: 3734	Aco5 (XM_502199)	GTTCACTGCCATATGAACAACAACCCCACCAACGTGATC
		Yarrowia lipolytica, Fwd. Primer, NdeI
SEQ ID NO:	Aco5 (XM_502199)	CTTCGAGATGCGGCCGCCTACTCGTCCAGGTCGCAAATCTC
		Yarrowia lipolytica, Rev. Primer, NotI

SEQ ID NO:	Description	Sequence
3735		
SEQ ID NO: 3736	Aco6 (XM_503632)	GTTCACTGCCATATGCTCTCTCAACAGTCCCTCAACAC
		Yarrowia lipolytica, Fwd. Primer, Ndel / Ncol
SEQ ID NO: 3737	Aco6 (XM_503632)	CTTCGAGATCTCGAGCTACTCATCCTCAAGAGAGCAAATTTCCTC
		Yarrowia lipolytica, Rev. Primer, Ncol / Xhol
SEQ ID NO: 3738	Acol (NM_001136902)	GTTCACTGCCATATGGACGCATCGGCGGAGGTGG
		Zea mays, Fwd. Primer, Ndel / Earl
SEQ ID NO: 3739	Acol (NM_001136902)	CTTCGAGATCTCGAGCTAGAGCCTGGAGAGCTTGAGCTGC
		Zea mays, Rev. Primer, Earl / Xhol
SEQ ID NO: 3740	Acolb (NM_001 17516 v)	GTTCACTGCCATATGGCGGAAGTGGACCACCTCGC
		Zea mays, Fwd. Primer, Ndel / BstXI
SEQ ID NO: 3741	Acolb (NM_001 17516 v)	CTTCGAGATCTCGAGCTAGAGCCTGGAGAGCTTGAGCTGC
		Zea mays, Rev. Primer, BstXI / Xhol
SEQ ID	Aco2 (NM_001 15855	GTTCACTGCCATATGGACCTCACCTCGCCGTCGCC

SEQ ID NO:	Description	Sequence
NO: 3742	2)	
		Zea mays, Fwd. Primer,
SEQ ID NO: 3743	Aco2 (NM_001 15855 2)	CTTGCGGCCGCTCAGTGGCTCCCGGTTGACAGTGCA
		Zea mays, Rev. Primer,
SEQ ID NO: 3744	Aco4 (NM_00 115683 4)	GTTCACTGC CATATGATGGC CGGGAAAC GAGTT ACGGG
		Zea mays, Fwd. Primer,
SEQ ID NO: 3745	Aco4 (NM_00 115683 4)	CTTCGAGATCTCGAGTCACAGCCGGGCTTTCGCTGG
		Zea mays, Rev. Primer,
SEQ ID NO: 3746	AC0X2 (XM_00138676 2)	GTTCACTGCCATATGATCCTGTTGCCCAAAGAGCTCC
		Scheffersomyces stipitis, Fwd. Primer, Ndel / Sail
SEQ ID NO: 3747	AC0X2 (XM_00138676 2)	GTTCACTGCGCGGCCGCCTAGCGGGACAATATCTTGGCAGC TTCG
		Scheffersomyces stipitis, Rev. Primer, Sail / Notl
SEQ ID NO: 3748	DEHA2D 17248 P (XM_459235)	GTTCACTGCCATATGGTTAGTGCTACTAATACAGTGAATTC AGG
		Debaryomyces hansenii, Fwd. Primer, Ndel

SEQ ID NO:	Description	Sequence
SEQ ID NO: 3749	DEHA2D 17248 P (XM_459235)	CTTCGAGATCTCGAGTTATTTGGATAAGATCTTAGCAGTTTC AGTAGACTTTTC
		Debaryomyces hansenii, Rev. Primer, Xhol
SEQ ID NO: 3750	ACX1 (NM_1 17778)	GTTCACTGCATTAATATGGAAGGAATTGATCACCTCGCCG
		Arabidopsis thaliana, Fwd. Primer, Asel
SEQ ID NO: 3751	ACX1 (NM_1 17778)	CTTCGAGATGTCGACTCAGAGCCTAGCGGTACGAAGTTGC
		Arabidopsis thaliana, Rev. Primer, Sail
SEQ ID NO: 3752	ACX2 (NM_00103706 8)	GTTCACTGCCATATGGAATCGCGGCGAGAGAAGAATCC
		Arabidopsis thaliana, Fwd. Primer, Ndel
SEQ ID NO: 3753	ACX2 (NM_00103706 8)	CTTCGAGATGTCGACTTATACAAGAAAACAAACCTTAGCTT TGTTAGGCGC
		Arabidopsis thaliana, Rev. Primer, Sail
SEQ ID NO: 3754	ACX2b (NMJ25910)	GTTCACTGCCATATGGAATCGCGGCGAGAGAAGAATCC
		Arabidopsis thaliana, Fwd. Primer, Ndel
SEQ ID NO: 3755	ACX2b (NMJ25910)	CTTCGAGATGTCGACTTAGAATCCAACAACTTGAGTATACT GGGAATAAG
		Arabidopsis thaliana, Rev. Primer, SslI

SEQ ID NO:	Description	Sequence
SEQ ID NO: 3756	ACX3 (NM_10051 1)	GTTCACTGCATTAATATGTCGGATAATCGTGCACTCCGACG
		Arabidopsis thaliana, Fwd. Primer, Asel
SEQ ID NO: 3757	ACX3 (NM_10051 1)	CTTCGAGATGTCGACCTAAACTGAAGACCAAGCATTGGCTTCG
		Arabidopsis thaliana, Rev. Primer, SALI
SEQ ID NO: 3758	ACX5 (NMJ29124)	GTTCACTGCCATATGGAGAGAGTTGATCACCTTGCCGATG
		Arabidopsis thaliana, Fwd. Primer, Ndel / Earl
SEQ ID NO: 3759	ACX5 (NMJ29124)	GTTCACTGCGCGGCCGCTTAGAGTTTGGCAGAGCGGAAGCGTTG
		Arabidopsis thaliana, Rev. Primer, Earl / Notl
SEQ ID NO: 3760	aco2 (XM_00352501 5)	GTTCACTGCCATATGCAAACTCCGAACTGTGAAGCA
		Glycine max, Fwd. Primer, Ndel
SEQ ID NO: 3761	aco2 (XM_00352501 5)	CTTGCGGCCGCTCAAAAACCGACGTATTGAGTGTAT
		Glycine max, Rev. Primer, Notl
SEQ ID NO: 3762	aoxA (XM_659264)	GTTCACTGCCATATGCCAAATCCACCTCCCGCCTGG
		Aspergillus nidulans, Fwd. Primer, Ndel

SEQ ID NO:	Description	Sequence
SEQ ID NO: 3763	aoxA (XM_659264)	CTTCGAGATCTCGAGTCACAGCTTGCTCTTAATCTCCCCCG
		Aspergillus nidulans, Rev. Primer, Xhol
SEQ ID NO: 3764	Acol (NM_0 17340)	GTTCACTGCCATATGAACCCAGACTTGAGAAAGGAAAGAG C
		Rattus norvegicus, Fwd. Primer, Ndel
SEQ ID NO: 3765	Acol (NM_0 17340)	CTTCGAGATCTCGAGCTACAACTTGGATTGCAATGGCTTCA AGTGC
		Rattus norvegicus, Rev. Primer, Xhol
SEQ ID NO: 3766	AcoII (1IS2_A)	GTTCACTGCCATATGAACCCAGACTTGAGAAAGGAAAGAG C
		Rattus norvegicus, Fwd. Primer, Ndel
SEQ ID NO: 3767	AcoII (1IS2_A)	CTTCGAGATCTCGAGCTACAACTTGGATTGCAATGGCTTCA AGTGC
		Rattus norvegicus, Rev. Primer, Xhol
SEQ ID NO: 3768	Aco (Cucsa.029560. 1)	GTTCACTGCCATATGGCTTCGCCGCGCGAGTC
		Cucumis sativus, Fwd. Primer, Ndel
SEQ ID NO: 3769	Aco (Cucsa.029560. 1)	CTTCGAGATCTCGAGTTAGAAGCCAACATACTGCGTATACT GCG
		Cucumis sativus, Rev. Primer, Xhol

SEQ ID NO:	Description	Sequence
SEQ ID NO: 3770	Aco (BAE47462)	GTTCACTGCCATATGACAGAAGTAGTGGACCGCGCATC
		Arthrobacter ureafaciens, Fwd. Primer, Ndel
SEQ ID NO: 3771	Aco (BAE47462)	CTTCGAGATCTCGAGCTAGCGGGACTTGCCGGCC
		Arthrobacter ureafaciens, Rev. Primer, Xhol
SEQ ID NO: 3772	Aco (YP_003571780)	GTTCACTGCCATATGCTCGATACCGACTCGCCACG
		Salinobacter rubber, Fwd. Primer, Ndel
SEQ ID NO: 3773	Aco (YP_003571780)	CTTCGAGATCTCGAGCTCGAGTCATTTCGGGCCGGG
		Salinobacter rubber, Rev. Primer, Xhol
SEQ ID NO: 3774	Aco (YP_290295.1)	GTTCACTGCCATATGCCCCAGCACGGCGATACAG
		Thermobifida fusca, Fwd. Primer, Ndel
SEQ ID NO: 3775	Aco (YP_290295.1)	CTTCGAGATCTCGAGCTCGAGTCACGTCTCCGCGC
		Thermobifida fusca, Rev. Primer, Xhol
SEQ ID NO: 3776	Aco (NC_009441.1)	GTTCACTGCCATATGAAACCAGCTAAACTTCAAGCCTTTACTCC
		Flavobacterium johnsoniae, Fwd. Primer, Ndel

SEQ ID NO:	Description	Sequence
SEQ ID NO: 3777	Aco (NC_009441.1)	CTTCGAGATCTCGAGCTAAACTGCAATTGGCGCTGCTAAACAG
		Flavobacterium johnsoniae, Rev. Primer, XhoI
SEQ ID NO: 3778	POX1 >gil50554589|ref|XP_504703.1| YALI0E32835p [Yarrowia lipolytica]	MAKERGKTQFTVRDVTNFLNGGEEETQIVEKIMSSIERDPVLSVTADYDCNLQQARKQTMERVAALSPYLVTDTEKLSLWRAQLHGMVDMSTRTRLSIHNNLFIGSIRGSGTPEQFKYWVKKGAVAVKQFYGCFAMTELGHGSNLKGLETTATYDQDSDQFIINTPHIGATKWWIGGAAHTSTHCVCFAKLIVHGKDYGTRNFVVPLRNVHDHSLKVGVSIGDIGKKMGRDGVDNGWIQFTNVRIPRQNMLMRYAKVSDTGVVTKPALDQLTYGALIRGRVSMIADSFHVSKRFLTIALRYACVRRQFGTSGDTKETKIIDYPYHQRRLLPLLAYCYAMKMGADEAQKTWIETTDRILALNPNDPAQKNDLEKAVTDTKELFAASAGMKAFTTWGCAKIIDECRQACGGHGYSGYNGFGQGYADWVVQCTWEGDNNVLCLSMGRGLVQSALQILAGKHVGASIQYVGDKSKISQNGQGTPREQLLSPEFLVEAFRTASRNNILRTTDKYQELVKTLNPDQAFEELSQQRFQCARIHTRQHLISSFYARIATAKDDIKPHLLKLANLFALWSIEEDTGIFLRENILTPGDIDLINSLVDELCVAVRDQVIGLTDAFGLSDFFINAPIGSYDGNVYEKYFAKVNQQNPATNPRPPYYESTLKPFLFREEEDDEICDLDE
SEQ ID NO: 3779	POX2 >gil50555712|ref|XP_505264.1| YALI0F10857p [Yarrowia lipolytica]	MNPNNTGTIEINGKEYNTFTEPPVAMAQERAKTSFPVREMTYFLDGGEKNTLKNEQIMEEIERDPLFNNDNYYDLNKEQIRELTMERVAKLSLFVRDQPEDDIKKRFALIGIADMGTYTRLGVHYGLFFGAVRGTGTAEQFGHWISKGAGDLRKFYGCFSMTELGHGSNLAGLETTAIYDEETDEFIINTPHIAATKWWIGGAAHTATHTVVFARLIVKGKDYGVKTFVVQLRNINDHSLKVGISIGDIGKKMGRDGIDNGWIQFTNVRIPRQNLLMKYTKVDREGNVTQPPLAQLTYGSLITGRVSMASDSHQVGKRFITIALRYACIRRQFSTTPGQPETKIIDYPYHQRRLLPLLAYVYALKMTADEVGALFSRTMLKMDDLKPDDKAGLNEVVSDVKELFSVSAGLKAFSTWACADVIDKTRQA

SEQ ID NO:	Description	Sequence
		CGGHGYSGYNGFGQAYADWVVQCTWEGDNNILTLSAGRALI QSAVALRKGEPVGNAVSYLKRYKDLANAKLNGPvSLTDPKVL VEAWEVAAGNIINRATDQYEKLIGEGLNADQAFEVLSQQRFQ AAKVHTRRHLIAAFFSRIDTEAGEAIKQPLLNLALLFALWSIEE DSGLFLREGFLEPKDIDTVTELVNKYCTTVREEVIGYTDAFNLS DYFINAPIGCYDGDAYRHYFQKVNEQNPARDPRPPYYASTLKP FLFREEEDDDICELDEE
SEQ ID NO: 3780	POX3 >gil50551539|ref|XP_503244.1 | YALI0D24750p [Yarrowia lipolytica]	MISPNLTANVEIDGKQYNTFTEPPKALAGERAKVKFPIKDMTE FLHGGEENVTMIERLMTELERDPVLNVSGDYDMPKEQLRETA VARIAALSGHWKKDTEKEALLRSQLHGIVDMGTRIRLGVHTG LFMGAIRGSGTKEQYDYWVRKGAADVKGFYGCFAMTELGHG SNVAGLETTATYIQDTDEFIINTPNTGATKWWIGGAAHSATHT ACFARLLVDGKDYGVKIFVVQLRDVSSHSLMPGIALGDIGKK MGRDAIDNGWIQFTNVRIPRQNMLMKYAKVSSTGKVSQPPLA QLTYGALIGGRVTMIADSFFVSQRFITIALRYACVRRQFGTTPG QPETKIIDYPYHQRRLLPLLAFTYAMKMAADQSQIQYDQTTDL LQTIDPKDKGALGKAIVDLKELFASSAGLKAFTTWTCANIIDQ CRQACGGHGYSGYNGFGQAYADWVVQCTWEGDNNVLCLSM GRGLIQSCLGHRKGKPLGSSVGYLANKGLEQATLSGRDLKDP KVLIEAWEKVANGAIQRATDKFVELTKGGLSPDQAFEELSQQ RFQCAKIHTRKHLVTAFYERINASAKADVKPYLINLANLFTLW SIEEDSGLFLREGFLQPKDIDQVTELVNHYCKEVRDQVAGYTD AFGLSDWFINAPIGNYDGDVYKHYFAKVNQQNPAQNPRPPYY ESTLRPFLFREDEDDDICELDEE
SEQ ID NO: 3781	POX4 >gil50554133|ref|XP_504475.1 | YALI0E27654p [Yarrowia	MITPNPANDIVHDGKLYDTFTEPPKLMAQERAQLDFDPRDITY FLDGSKEETELLESLMLMYERDPLFNNQNEYDESFETLRERSV KRIFQLSKSIAMDPEPMSFRKIGFLGILDMGTYARLGVHYALFC NSIRGQGTPDQLMYWLDQGAMVIKGFYGCFAMTEMGHGSNL SRLETIATFDKETDEFIINTPHVGATKWWIGGAAHTATHTLAF ARLQVDGKDYGVKSFVVPLRNLDDHSLRPGIATGDIGKKMGR

SEQ ID NO:	Description	Sequence
	lipolytica]	DAVDNGWIQFTNVRVPRNYMLMKHTKVLRDGTVKQPPLAQL TYGSLITGRVQMTTDSHNVSKKFLTIALRYATIRRQFSSTPGEP ETRLIDYLYHQRRLLPLMAYSYAMKLAGDHVRELFFASQEKA ESLKEDDKAGVESYVQDIKELFSVSAGLKAATTWACADIIDKA RQACGGHGYSAYNGFGQAFQDWVVQCTWEGDNTVLTLSAG RALIQSALVYRKEGKLGNATKYLSRSKELANAKRNGRSLEDP KLLVEAWEAVSAGAINAATDAYEELSKQGVSVDECFEQVSQE RFQAARIHTRRALIEAFYSRIATADEKVKPHLIPLANLFALWSIE EDSALFLAEGYFEPEDIIEVTSLVNKYCGIVRKNVIGYTDAFNL SDYFINAAIGRYDGDVYKNYFEKVKQQYPPEGGKPHYYEDV MKPFLHRERIPDVPMEPEDIQ
SEQ ID NO: 3782	POX5 >gi|50549457|ref|XP_502 199.1 | YALI0C23859 p [Yarrowia lipolytica]	MNNNPTNVILGGKEYDTFTEPPAQMELERAKTQFKVRDVTNF LTGSEQETLLTERIMREIERDPVLNVAGDYDADLPTKRRQAVE RIGALARYLPKDSEKEAILRGQLHGIVDMGTRTRIAVHYGLFM GAIRGSGTKEQYDYWVAKGAATLHKFYGCFAMTELGHGSNV AGLETTATLDKDTDEFIINTPNSGATKWWIGGAAHSATHTACL ARLIVDGKDYGVKIFIVQLRDLNSHSLLNGIAIGDIGKKMGRD AIDNGWIQFTDVRIPRQNMLMRYDRVSRDGEVTTSELAQLTY GALLSGRVTMIAESHLLSARFLTIALRYACIRRQFGAVPDKPET KLIDYPYHQRRLLPLLAYTYAMKMGADEAQQQYNSSFGALL KLNPVKDAEKFAVATADLKALFASSAGMKAFTTWAAAKIIDE CRQACGGHGYSGYNGFGQAYADWVVQCTWEGDNNVLCLSM GRSLIQSCIAMRKKKGH VGKS VEYLQRRDELQN ARVDNKPLT DPAVLITAWEKVACEAINRATDSFIKLTQEGLSPDQAFEELSQQ RFECARIHTRKHLITSFYARISKAKARVKPHLTVLANLFAVWSI EEDSGLFLREGCFEPAEMDEITALVDELCCEAREQVIGFTDAFN LSDFFINAPIGRFDGDAYKHYMDEVKAANNPRNTHAPYYETK LRPFLFRPDEDEEICDLDE
SEQ ID	POX6 >gi|50552444|r	MLSQQSLNTFTEPPVEMARERNQTSFNPRLLTYFLDGGEKNTL LMDRLMQEYERDPVFRNEGDYDITDVAQSRELAFKRIAKLIEY

SEQ ID NO:	Description	Sequence
NO: 3783	eflXP_503632.1 | YALI0E06567p [Yarrowia lipolytica]	VHTDDEETYLYRCMLLGQIDMGAFARYAIHHGVWGGAIRGA GTPEQYEFWVKKGSLSVKKFYGSFSMTELGHGSNLVGLETTA TLDKNADEFVINTPNVAATKWWIGGAADTATHTAVFARLIVD GEDHGVKTFVVQLRDVETHNLMPGIAIGDCGKKMGRQGTDN GWIQFTHVRIPRQNMLMRYCHVDSDGNVTEPMMAQMAYGA LLAGRVGMAMDSYFTSRKFLTIALRYATIRRAFAAGGGQETK LIDYPYHQRRLLPLMA QTYAIKCTADKVRDQFVKVTDMLLNLDVSDQEAVPKAIAEAK ELFSVSAGVKATTTWACAHTIDQCRQACGGHGYSAYNGFGR AYSDWVIQCTWEGDNNILCLSAGRALVQSNRAVRAGKPIGGP TAYLAAPAGSPKLAGRNLYDPKVMIGAWETVSRALINRTTDE FEVLAKKGLSTAQAYEELSQQRFLCTRIHTRLYMVKNFYERIA EEGTEFTKEPLTRLANLYAFWSVEEEAGIFLREGYITPQELKYIS AEIRKQLLEVRKDVIGYTDAFNVPDFFLNSAIGRADGDVYKNY FKVVNTQNPPQDPRPPYYESVIRPFLFRKDEDEEICSLEDE
SEQ ID NO: 3784	pAA298 (Nucleic Acid Seq.)	gatctggaatccctcggcgtcggtcttgggggtgggggcattctttcttggtcttgggaacgccaac gctttgttgtttgggttcttgaacacggactgctcgaaaaagtaccagtatgatgccttacctctgctgg ctttccacaagtatgggagggcattgacagcgactgtcttggctaacagcacgtcgtcggcaattaa atatttggcttccaataactgactaccaaggatggcagcagcggctatttctaatcctgacatgtttctc gtacgtagtagtgaatgaagggaaggtggaataatatcaagggcgaattctgcagatatccatcac actggcggccgctcgagcatgcatctagagggcccaattcgccctatagtgagtcgtattacaattc actggccgtcgttttacaacgtcgtgactgggaaaaccctggcgttacccaacttaatcgccttgcag cacatccccctttcgccagctggcgtaatagcgaagaggcccgcaccgatcgcccttcccaacagt tgcgcagcctatacgtacggcagtttaaggtttacacctataaaagagagagccgttatcgtctgtttg tggatgtacagagtgatattattgacacgccggggcgacggatggtgatccccctggccagtgcac gtctgctgtcagataaagtctcccgtgaactttacccggtggtgcatatcggggatgaaagctggcg catgatgaccaccgatatggccagtgtgccggtctccgttatcggggaagaagtggctgatctcag ccaccgcgaaaatgacatcaaaaacgccattaacctgatgttctggggaatataaatgtcaggcatg agattatcaaaaaggatcttcacctagatccttttcacgtagaaagccagtccgcagaaacggtgct gaccccggatgaatgtcagctactgggctatctggacaagggaaaacgcaagcgcaaagagaaa

SEQ ID NO:	Description	Sequence
		gcaggtagcttgcagtgggcttacatggcgatagctagactgggcggttttatggacagcaagcga accggaattgccagctggggcgccctctggtaaggttgggaagccctgcaaagtaaactggatgg ctttctcgccgccaaggatctgatggcgcaggggatcaagctctgatcaagagacaggatgagga tcgtttcgcatgattgaacaagatggattgcacgcaggttctccggccgcttgggtggagaggctatt cggctatgactgggcacaacagacaatcggctgctctgatgccgccgtgttccggctgtcagcgca ggggcgcccggttctttttgtcaagaccgacctgtccggtgccctgaatgaactgcaagacgaggc agcgcggctatcgtggctggccacgacgggcgttccttgcgcagctgtgctcgacgttgtcactga agcgggaagggactggctgctattgggcgaagtgccggggcaggatctcctgtcatctcaccttg ctcctgccgagaaagtatccatcatggctgatgcaatgcggcggctgcatacgcttgatccggctac ctgcccattcgaccaccaagcgaaacatcgcatcgagcgagcacgtactcggatggaagccggt cttgtcgatcaggatgatctggacgaagagcatcaggggctcgcgccagccgaactgttcgccag gctcaaggcgagcatgcccgacggcgaggatctcgtcgtgacccatggcgatgcctgcttgccga atatcatggtggaaaatggccgcttttctggattcatcgactgtggccggctgggtgtggcggaccg ctatcaggacatagcgttggctacccgtgatattgctgaagagcttggcggcgaatgggctgaccg cttcctcgtgctttacggtatcgccgctcccgattcgcagcgcatcgccttctatcgccttcttgacga gttcttctgaattattaacgcttacaatttcctgatgcggtattttctccttacgcatctgtgcggtatttca caccgcatacaggtggcacttttcggggaaatgtgcgcggaacccctatttgtttatttttctaaatataca ttcaaatatgtatccgctcatgagacaataaccctgataaatgcttcaataatagcacgtgaggaggg ccaccatggccaagttgaccagtgccgttccggtgctcaccgcgcgcgacgtcgccggagcggt cgagttctggaccgaccggctcgggttctcccgggacttcgtggaggacgacttcgccggtgtggt ccgggacgacgtgaccctgttcatcagcgcggtccaggaccaggtggtgccggacaacaccctg gcctgggtgtgggtgcgcggcctggacgagctgtacgccgagtggtcggaggtcgtgtccacga acttccgggacgcctccgggccggccatgaccgagatcggcgagcagccgtgggggcgggag ttcgccctgcgcgacccggccggcaactgcgtgcacttcgtggccgaggagcaggactgacacg tgctaaaacttcatttttaatttaaaaggatctaggtgaagatccttttgataatctcatgaccaaaatcc cttaacgtgagttttcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagat ccttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgc cggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagataccaaatac tgtccttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcg ctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactca

SEQ ID NO:	Description	Sequence
		agacgatagttaccggataaggcgcagcggtcgggctgaacgggggttcgtgcacacagccca gcttggagcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccac gcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggagagc gcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctg acttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgc ggccttttacggttcctgggcttttgctggccttttgctcacatgttctttcctgcgttatcccctgattct gtggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcg cagcgagtcagtgagcgaggaagcggaagagcgcccaatacgcaaaccgcctctccccgcgcg ttggccgattcattaatgcagctggcacgacaggtttcccgactggaaagcgggcagtgagcgca acgcaattaatgtgagttagctcactcattaggcaccccaggctttacactttatgcttccggctcgtat gttgtgtggaattgtgagcggataacaatttcacacaggaaacagctatgaccatgattacgccaag ctatttaggtgacactatagaatactcaagctatgcatcaagcttggtaccgagctcggatccactagt aacggccgccagtgtgctggaattcgcccttccgttaaacaaaaatcagtctgtaaaaaaggttcta aataaatattctgtctagtgtacacattctcccaaaatagtgaaatccagctctacaatttggctttaccg gtacaaatcaaagaccaatcgtcttcagtcaactcctggtacttgtcgccattcaaccagtagatcaa atccttgccgtcttcacccttggcaacttttggttcttgaattggttcttaggaaccttgtggttgtgcga agcctcaatggtgccaatcttgatgaacgcaggttgagcatacacaggcaaagacttggtcacgtg agagtgaatcaatttcaagatttcttcatggctcaactcgtccttggcttcacagacggcaaaacagg ctctaccttcgtggtttggcaccttgacaccgacaacgacggactgcttcaaggccttggagcccat caattcgttctcgacctcggtggcggagacgttttcggacttccaacggaaagtgtcacctaatctgt cgacaaagtacaacaatttgtcctcgtccatcttcaacaagtcaccggatctgtaccacgcgtcacct tttttgaaaacattggtgaggattttgctgttggtggcggacttgttaccataataaccctggaaggattt ctgcacgtcgttagggttcaagattctcatcaacaactcacctggctcgttgtaagcggcctcggtac agaacccggtcttggggtccttgtagatttcactctcgtcttctgggtccatcttggccaatttctgctg ggtagacaataacaagctgatgagggacccgtacttacgacaggcgccgacaccgtactcaccgt actgcaagttggtggtggcgataggggactcggtggcggcgtagaactcaccgataccttcaatgt ggaatctgcgcttgaactcagaccatatatctggacgcaacccgttaccgtaggcaattctgacattg tgtctgtcttggtctggatgaggcttggagttcaacaagtaacgacagacctcaccgacgtattgcac gtgggtggcaccacataatctggcctgggtccagaacgaagtagcggagaatttctgggacacgg agacacagccaccgacaatcaaagtaggacacaaccccaacatggccgcggtggagtggtaca

SEQ ID NO:	Description	Sequence
		agggcatggcggtcaagacgttcgatttcgagtcaatcttcatgatgtggccaaagaaaaccgagg ccatgaaggcttttctccaggacatgataccggcttttggcaaaccggtggtacccgaggtgtaaat caatgcacaagcggaggagtcagtatcggttggtcttctggtcttgtcctcggctctgtgttttggagt cgacttgagtctcaatctgtcaaacaaggcaaactcgtcaatgtagtttatttgcacatgtggcaattc ctctctgatctgagcctcggtatctctgattggggaatcacagtccgggtcaacgaaaacttgcgaa gcgttgacaatcttaagacagtggatcaatggcttgtccttggtgttgaagttcaagaacgcaggcaa ggcaccaatgttccacaatgccaaccacaagacaatgaaaagcggcttgttcatacaagaaacacc gatggtgtcgttggcagtgacgccgtactcgttcttcaagatgtatgagtacttcaaaaccatgtcgta caattccttgtaggtgtattcttctaggtcaaactgatcgtcgtagatccactagtaacggccgccagt gtgctggaattcgcccttgggctaacgaaaaggaaaccgctgacgttaaaggtatctacggttgtttc ggtatgacccccggggatctgacgggtacaacgagaattgtattgaattgatcaagaacatgatctt ggtgttacagaacatcaagttcttggaccagactgagaatgcacagatatacaaggcgtcatgtgat aaaatggatgagatttatccacaattgaagaaagagtttatggaaagtggtcaaccagaagctaaac aggaagaagcaaacgaagaggtgaaacaagaagaagaaggtaaataagtattttgtattatataac aaacaaagtaaggaatacagatttatacaataaattgccatactagtcacgtgagatatctcatccatt ccccaactcccaagaaaaaaaaaaagtgaaaaaaaaaatcaaacccaaagatcaacctccccatc atcatcgtcatcaaacccccagctcaattcgcaatggttagcacaaaaacatacacagaaagggca tcagcacacccctccnaggttgcccaacgtttattccgcttaatggagtccaaaaagaccaacctct gcgcctcgatcgacgtgaccacaaccgccgagttcctttcgctcatcgacaagctcggtccccaca tctgtctcgtgaagacgcacatcgatntcatctcagacttcagctacgagggcacgattgagccgtt gcttgtgcttgcagagcgccacgggttcttgatattcgaggacaggaagtttgctgatatcggaaac accgtgatgttgcagtacacctcgggggtataccggatcgcggcgtggagtgacatcacgaacgc gcacggagtgactgggaagggcgtcgttgaagggttgaaacgcggtgcggaggggtagaaaa ggaaaggggcgtgttgatgtnggcggagttgtcgagtaaaggctcgttggcgcatggtgaatatac ccgtgagacgatcgagattgcgaagagtgatcgggagttcgtgattgggttcatcgcgcagcggg acatgggggtagagaagaagggtttgattggatcatcatgacgcctggtgtggggttggatgata aaggcgatgcgttgggccagcagtataggactgttgatgaggtggttctgactggtaccgatgtgat tattgtcgggagagggttgtttggaaaaggaagagaccctgaggtggagggaaagagatacagg gatgctggatggaaggcatacttgaagagaactggtcagttagaataaatattgtaataaataggtct atatacatacactaagcttctaggacgtcattgtagtcttcgaagttgtctgctagtttagttctcatgatt

SEQ ID NO:	Description	Sequence
		tcgaaaaccaataacgcaatggatgtagcagggatggtggttagtgcgttcctgacaaacccagag tacgccgcctcaaaccacgtcacattcgcccttgcttcatccgcatcacttgcttgaaggtatccacg tacgagttgtaatacaccttgaagaacggcttcgtctacggtcgacgacgggtacaacgagaattgt attgaattgatcaagaacatgatcttggtgttacagaacatcaagttcttggaccagactgagaatgc acagatatacaaggcgtcatgtgataaaatggatgagatttatccacaattgaagaaagagtttatgg aaagtggtcaaccagaagctaaacaggaagaagcaaacgaagaggtgaaacaagaagaagaa ggtaaataagtattttgtattatataacaaacaaagtaaggaatacagatttatacaataaattgccata ctagtcacgtgagatatctcatccattccccaactcccaagaaaaaaaaaaagtgaaaaaaaaaatc aaacccaaagatcaacctccccatcatcatcgtcatcaaacccccagctcaattcgcagagctcggt acccggg
SEQ ID NO: 3785	Candida strain ATCC20336 Thioesterase PTE1	MRGMEKPHSLFRRMSTAPFAIIQPPISILSATLITSEFFFVYSLYN FFHFIIDLFYYYYTHIYPHPvAMIENISGNGNYPQNHEVDLEKEF GVEKIGINLYRDKSPIPKPDRRSRGAYGGYLAGQALLVAMKST PPEYRPHSFHSYFIKAVNDKETLEWRVEETSNGRNYANRSLQA FQAGNLVYTANVSLTKKNSAKKAEEATGVKPFEFQGKPHEWF EKHKRDDLPLATPSSSLLIYHKFFPEVVSLEASKEEESKPAADR ELSWYFKWGINNEEGHHQPLVNLNSDYQYVGMAALTDAVYL NRLLRILRVEDADHTQLVHYFSVSLDHTMYFHDDDFDVTKW MGFTFKVTRFSHNRALCQGEVYNDKGVHVCTIVQEGLMMLN GLEEGAKL*
SEQ ID NO: 3786	Candida strain ATCC20336 Thioesterase PTE2	MICVFFPTSTFTTAHKFVSNLQSFFLSQQPHTTSYTMPTFNYKD GETIDVQKEFGVVETAPNKYVGVKPLVKPMPHVKGVFGGNL AGQALLVAMKSVGPDFSPHSLHSYFIRAGSDQTPVEWTVQAIS DGNSFCNRFIKGVQNGQVIYIANVSLTKRNSAADAMKKYEEY HAQIRQKGKDGDADEEDEDDDDEDDNAPAKPFGFQTPSHKWI KDRDLDKLPVSDMESNLLLYYKLPPEFVSLKSSTEEESLPVSER RMGALAKWGIENEQGFNQPLTNLDKSFQYVGLANITDGLYLG TLNRILRIDDLTLDERATNYFSVSLDHVIYFHDDDFDVTKWMG FTFRCSRYSHNRVIFEGEIYSDKGVQVASIIQEGLVRFKDGYLK NAKL

Example 54: Additional Nucleotide and Amino Acid Sequences

SEQ ID NO: 3787	Peroxisomal 2,4-dienoyl-CoA reductase >gi|1102790511 sp|P32573.4| SPS19_YEAST	MNTANTLDGKFVTEGSWRPDLFKGKVAFVTGGAGTICRVQT EALVLLGCKAAIVGRDQERTEQAAKGISQLAKDKDAVLAIAN VDVRNFEQVENAVKKTVEKFGKIDFVIAGAAGNFVCDFANLS PNAFKSVVDIDLLGSFNTAKACLKELKKSKGSILFVSATFHYY GVPFQGHVGAAKAGIDALAKNLAVELGPLGIRSNCIAPGAIDN TEGLKRLAGKKYKEKALAKIPLQRLGSTRDIAESTVYIFSPAA SYVTGTVLVVDGGMWHLGTYFGHELYPEALIKSMTSKL
SEQ ID NO: 3788	CvDCR1_NODE5705	atgccaaacactttagatcacaactaccttaagaaaagcgtctggaagtcagatatctttgccggca aagtcatcttcatcactggtggtgccggtaccatctgtcgtgtgcaagccgaagccatggtcttgttg ggtgctaatgctgccatcattggtagaaacgtcgaaaagactgaagaagccgctaaggagattgc ttccttgagaccaggagccaaggtcattggtattggtgccgttgacgtgcgtaagatccaaactatc aaggatgccgttgacagaactgttgccgagttgggcagaatcgactatgtcattgctggtgccgcc ggtaacttcctttgtgacttcaaccacttgtctgccaacgccttcaagtccgtgattgacattgatttgtt gggttctttcaacaccgtcaaggttacctttgaccagttgagaaagaacaaaggtgccgtgttgtttg tcagtgccactttgcactacaccggtgttcctatgcaatcgcacgttggtgctgccaagtctggtgtt gatgccttgagcaacgcccttgctgttgaattgggtccattgggtatcagattcaactgtattgcccc aggtgccattgctggtactgaaggtatgctgagattggccccacctaccgacacgccattggaga cgaagattccattgcagagacaaggtactactgaggacattgctgatgccactgtgtttttgttctcg ccagcttcctcctacattactggtgatgttttggttgttgatggtgctatgtggcaaactggcggaggt attcttaacgacttctacccagacatcattattcaccagaacgccgacccagaaggtaagttgtaa
SEQ ID NO: 3789	CvDCR1_NODE5705	MPNTLDHNYLKKSVWKSDIFAGKVIFITGGAGTICRVQAEAM VLLGANAAIIGRNVEKTEEAAKEIASLRPGAKVIGIGAVDVRK IQTIKDAVDRTVAELGRIDYVIAGAAGNFLCDFNHLSANAFKS VIDIDLLGSFNTVKVTFDQLRKNKGAVLFVSATLHYTGVPMQ SHVGAAKSGVDALSNALAVELGPLGIRFNCIAPGAIAGTEGMS RLAPPTDTPLETKIPLQRQGTTEDIADATVFLFSPASSYITGDVL WDGAMWQTGGGILNDFYPDIIIHQNADPEGKL
SEQ ID	CvDCR2_NODE6821	atgccaaacactttaactgaagcatacaaacaagaaagttcttggaagccagatttgttcaagggca aagtcgtcttcatcaccggtggtgcaggttcgatctgccgtgtccaggccgaagccatggtcttgtt

NO: 3790		gggtgccaacgctgccatcattggtagaaacgtcgaaaagaccgaatccgccgccaaagaaatc gccagcttgagatcaggagccaaggttcttggtattggtggcattgatgtgcgtaaagttgacagct tgaagagcgccgtcgataagaccgttgccgaattgggcagaatcgactttgtcattgctggtgccg ccggtaacttcctttgtgacttcaaccacttgtcttccaacgccttcaaatccatcattgacattgactt gttgggttctttcaacaccgtcaaggtcacttttgaccaattgagaaagaacaagggtgccatcttgt ttgtcagtgccaccttgcactactacggtgttcctttccaaattggtgttggtgctgccaaggctggtg tcgacgccttgagcaatgcccttgccgttgagttgggtccattgggcatcagatccaactgtattgc cccaggtccaatcgacggtaccgaaggtgttgaaagattggtcagagcttccaaggccgaagcc gccaagaaggttccattgcaaagattgggtactacacaagacattgctgacggtaccgtctacttgt tctccccagccgcctcctttgtcactggtgacgttttggttgttgatggtgctagctggcagatctcct ctggtgttggtgccaccaactacccagtctccatcctcaatgcgatcgacgcaccgaagggtggta agttgtaa
SEQ ID NO: 3791	CvDCR2_NODE6821	MPNTLTEAYKQESSWKPDLFKGKVVFITGGAGSICRVQAEAM VLLGANAAIIGRNVEKTESAAKEIASLRSGAKVLGIGGIDVRK VDSLKSAVDKTVAELGRIDFVIAGAAGNFLCDFNHLSSNAFKS IIDIDLLGSFNTVKVTFDQLRKNKGAILFVSATLHYYGVPFQIG VGAAKAGVDALSNALAVELGPLGIRSNCIAPGPIDGTEGVERL VRASKAEAAKKVPLQRLGTTQDIADGTVYLFSPAASFVTGDV LVVDGASWQISSGVGATNYPVSILNAIDAPKGGKL
SEQ ID NO: 3792	DCR; Candida_tropicalis_XP_002545283	MPNTLDQSYVKRSVWKSDIYAGKVVFITGGAGSICRVQAEA MVLLGANAAIIGRNVEKTESAAEEIASLRPGAKVIGIGGVDVR KIQTIKEAVDRTVAELGRIDYVVAGAAGNFLCDFNHLSSNAF KSVIDIDLLGSFNTVKVTFDELRKNKGSVIFVSATLHYTGVPM QAHVGAAKAGVDALSNALAVELGPLGIRFNCIAPGAISGTEG MSRLAPPSDTPLEKKIPLQRQGTTEDIADATVFLFSPAASYVTG DVLVVDGAMWQTGGGIINDFYPDIIVHQNSDPDGKL
SEQ ID NO: 3793	DCR; Candida_tropicalis_XP_002550344	MPNTLTQSYVQESSWKPDLFKGKVVFITGGAGSICRVQAEAM VLLGANAAIIGRNVEKTESAAKEIASLRSGAKVLGVGGIDVRK IDSLKSAVDKTVAELGRIDYVIAGAAGNFLCDFNHLSSNAFKS IVDIDLLGSFNTVKATFDQLRKNKGAILFVSATLHYYGVPYQI GVGAAKAGVDALSNAFAVELGPLGIRSNCIAPGPIDGTEGLSR LSRASKDQTSKKVPLQRLGTTQDIADGTVYLFSPAASFVTGD VLVVDGASWQISSGVGATDYPVSITKANSDPVGKL

SEQ ID NO: 3794	DCR; Yarrowia_lipolytica_XP_501382	MPNTLSPDWLKKSVFKPDIFKGKVVFVTGGAGTICRVQTEAM ILLGADAAIIGRRPEVTQKAADEMAQLRPGAKVIGVGNTDVR EVKSLVAAAERTVKELGRIDFVICGAAGNFLSDFNHLSSNAFK SVISIDLLGSYNTVKACFPELRKNKGSILFVSATLHYYGIPFQT HVSAAKAGIDALSNALAVELGPLGIRSNCIAPGAIAGTEGIDRL LPADMREKALSLIPAQRWGQTEDIANGTVYYVFSEAASYVSGT TIVVDGAGWHTSNGAANNAVYPQMLIAQGNDTPKL
SEQ ID NO: 3795	DCR; Yarrowia_lipolytica_XP_503937	MPNTLTPEWLKTCVYNPGLFDGKVVFVTGGAGTICRVQTEAL ILLGANAAVIGRRPEVTQKAAEEMQQLRPGAKVIGIGNCDVR EVKSLVAAAEKAVQELGRIDYVIAGAAGNFLADFNHLSANAF KSVISIDLLGSYNTVKACFPELRKNKGKVLFVSATLHYRGVSL QSHVSAAKAGIDALSQALAVELGPLGIAVNCLAPGPIDGTEGL GRLLPSDARKRSLQLVPVQRFGTTEDIANGTVFLFSDAASYIS GTTLVIDGAAWHTSARTTYPETVIVQGNKPPKL
SEQ ID NO: 3796	DCR; Yarrowia_lipolytica_XP_502296	MLPADFIESSPYKSNIFVGKVVFVTGGAGTICKDQTEALVRLG ANGAIVGRKKEVTEKAAKELEALRPGARVLGLGEIDVRDIQS LKRAVDTTISELGRIDYVIAGAAGNFVTDINHMSANAFKTVID IDLLGSFNTAKATFEALRASAGSLVFISATAHYHGSPFTAHVG AAKAGIDALMQALAVELGPLGIRVNCIAPGFIAATEGMDRLLP PDMQKTYTRLTPLQRFGTTLDIANATVWLFSEAASYVSGTVI WDGAGWHSAQQTSTLHYYPDMLKHMGEDKSKI
SEQ ID NO: 3797	ScECIl	atgtcgcaagaaattaggcaaaatgagaaaatcagttatcgtattgaaggaccattcttcattattcac ttaatgaaccctgacaatttgaatgcactagaaggtgaagactatatttatttaggagagttactagaa ctagcggacagaaatcgtgatgtatattttacaattatacaaagcagtggtagatttttttccagtggtg ctgatttcaagggtattgcaaaagcccaaggggatgataccaataaatatccttcggaaacaagca agtgggtgtcaaattttgtcgctagaaatgtttatgtcactgatgccttcatcaagcattccaaagtttta atttgctgtttgaatggaccagcaatagggttgagcgcggcactggtagcgttatgtgacattgtgta cagtataaatgacaaggtttatttgctatacccctttgctaacttaggactaattaccgaaggtggtac aacggtctctttgccattgaagtttggcacaaatacgacgtatgaatgcctcatgttcaacaaaccatt caagtacgatataatgtgcgagaacggatttataagcaagaattttaacatgccatcttcaaacgctg aagcgttcaatgcaaaggtcttagaagaattgagggagaaagtgaaagggctatacctgcccagt tgcttagggatgaaaaaattgctgaaatcgaaccacatcgatgcattcaataaggctaactcagtgg aagtaaatgaatctctcaagtattgggtagatggagagcccttaaaaagatttaggcagctgggctc

		gaaacaaaggaagcatcgtttatga
SEQ ID NO: 3798	ScEcilp	MSQEIRQNEKISYRIEGPFFIIHLMNPDNLNALEGEDYIYLGEL LELADRNPvDVYFTIIQ SSGRFF SSGADFKGI AKAQGDDTNKYP SETSKWVSNFVARNVYVTDAFIKHSKVLICCLNGPAIGLSAAL VALCDIVYSINDKVYLLYPFANLGLITEGGTTVSLPLKFGTNTT YECLMFNKPFKYDIMCENGFISKNFNMPSSNAEAFNAKVLEE LREKVKGLYLPSCLGMKKLLKSNHIDAFNKANSVEVNESLKY WVDGEPLKRFRQLGSKQRKHRL
SEQ ID NO: 3799	CvECIl	atgtccgacgaggaatcagatatcttatacgaagtcagagacagaaccgccatcatcaccttgaac atccccaagagattgaacgcattgaacggcgctcaatacttgaagttgggcaagttcttggagaga gccaacaacgaagaggacaccgtcttgaccttgatccaggccctgggcagattcttctccgccgg tgccaatttcgccgacaacgacatggccaaggtcgaaatgtccaagttgttcagtcacgagtactg gttggatagattcgtcgccagaaacatctggttgaccaacttgttcaatgaccacaagaagatcttg gccgctgctgtcaacggtccagttatcggtttgagtactggtttgttgttggtcgacttggtctacg tccacgacttgaacaagttctatctcttggccccatttgccaacttgggtttggttgccgaaggtgctt cctcggccactttgttcaacagattgggctggtcaaaggcttctgaagctttgttgttggccaagcca atcagcggccaggactgttacaacgctggtttcatcaacaagcactacgacggtaagttttcctcca ctgaggagttcaacgaacacgtctacaaggagttgaccgatgcctttgaaaacttgcacgaagatt ccatcttgcagaacaagcaattgttgaagttgtcgagagaccaggctatcaatgttgccaactcgca agaggtcatgagaggtttgggcaagtggttggaaggtgttccacaaaccagatttgctcaattggc ccagaaggaggtcaagcacaaactttag
SEQ ID NO: 3800	CvEcilp	MSDEESDILYEVRDRTAIITLNIPKRLNALNGAQYLKLGKFLE RANNEEDTVLTLIQASGRFFSAGANFADNDMAKVEMSKLFSH EYWLDRFVARNIWLTNLFNDHKKILAAAVNGPVIGLSTGLLL LVDLVYVHDLNKFYLLAPFANLGLVAEGASSATLFNRLGWS KASEALLLAKPISGQDCYNAGFINKHYDGKFSSTEEFNEHVYK ELTDAFENLHEDSILQNKQLLKLSRDQAINVANSQEVMRGLG KWLEGVPQTRFAQLAQKEVKHKL
SEQ ID	CvECI2	atgtccgacgaccttatcacctacgaagtcaaagaccgagctgccgtgatcaccttgaacaacccc aagaagctcaacgccttatcgatcccgcagtacgacaccatctgcaagctcttagaacgagccaa

NO: 3801		cgccgaagaagacaccgtcatcaccttgctccagtccacgggccgggtgttctctgccggggcc aacgccgactccatcgtggggcaggatgccgagctcgagacctggttgaacatgtcggtggcca agcagacgttcttggtgcagacgttcctcgcacacaagaagatccttgccgtcgccttgaacggcc ccgtgattggcttatcggcggcgttcgtggcgctctgcgacttggtctacgtgcacaacgccgcaa agacgttcttcttgaccccgttcgccaacatcgggatccttgccgagggcggcacctcagccacgt tgcccatgcgcgtggggtggtccagggccgcggaagcgttgttgttgtcaaagaggatttcggga gatgacttgcagagagcggggttcttcaataaggactacaaggggcagttcaagtccgcggagg agtttaacgaggtcgtcttgaaggagttgcttgacgccacggaaaacttgcatgaggactcgatcat ccagaacaaggagttgttgaaggctattttcaagccaaagatcagtgaggtcaactcgcaggagg tgtcaagaggtgtgtacaagtggacctctggggtgccaatggatagatttaaaaaattgcttaatggt gagttgaaacataaattatag
SEQ ID NO: 3802	CvEci2p	MSDDLITYEVKDRAAVITLNNPKKLNALSIPQYDTICKLLERA NAEEDTVITLLQSTGRVFSAGANADSIVGQDAELETWLNMSV AKQTFLVQTFLAHKKILAVALNGPVIGLSAAFVALCDLVYVH NAAKTFFLTPFANIGILAEGGTSATLPMRVGWSRAAEALLLSK RISGDDLQRAGFFNKDYKGQFKSAEEFNEVVLKELLDATENL HEDSIIQNKELLKAIFKPKISEVNSQEVSRGVYKWTSGVPMDR FKKLLNGELKHKL
SEQ ID NO: 3803	CvEcilp; Candida_tropicalis_XP_002548748	MSEEETDILYEVKDRTAIITLNIPDKLNALNGAQYLKLGKFLE RANNEEDT VLTLIQ ASGRFF SAGANF ADNDMTKVE VEKLF SH EYWLERFVARNVWLTNLFNDHKKILAAAVNGPVIGLSTGLL MLVDLIYVHDLNKFYLLAPFANLGLVAEGASSATLFARLGWS KASEALLLAKPISGADCYNAGLINKHYDGKFSSTEQFNDQVY KELNDAFENLHEDSILQNKQLLKLSRDQFINHANSQEVMRGL GKWLEGVPQTRFVQLAQKEVKHKL
SEQ ID NO: 3804	CvEcilp; Yarrowia_lipolytica_XP_500719.1	MRSLYINVPGLFPSTSLARETVHHRTEMLRTIRSSSRLGVRAM STAATRRAAQIGFHTRVPT VVTKAPTLRMQTTPFS SSAPAQTF GDKKYEHILTSTPVPKVALVTLNRPKALNALCTPLIKELNEAL QAADADPTIGAIVLTGSEKSFAAGADIKEMKDKTVTSVLNEN FIEEWGNMANIKKPIIAAVNGFALGGGCELAMMADIIYAGAK AKFGQPEIKLGVIPGAGGTQRLTRAIGLYRANHYILTGEMFTA QQAADWGLAAKVYEPAQLVDESVKAAAQIASYGQLAVQAA KASVHQSAEVGLRAGLEFERVRFHGLFGTHDQKEGMAAFAE

		KREPNFKNE
SEQ ID NO: 3805	CvEci2p; Candida_tropicalis_XP_002548746	MSDELITYEVKDRAVLITLNNPKKLNSLTIHQYDTICKYLERA NNEEDTIITLIQ STGRVF SAGANADSIVGQD AELET WMNLSI A KQTFLVQTFLAHKKILAVALNGPVIGLSASFVTLCDLIYVNDL SKTFLLTPFANIGILAEGGTSATLPMRLGWSKASEALLLSKRIS GDDLQNAGLINKHYNGKFKTAEEFNVTVLNELLDATETLHQ DSILQNKQLLQAMFKPRISEVNSQEVSRGVYKWTSGEPMSRF KKLLTGELKHKL
SEQ ID NO: 3806	CvEci2p; Yarrowia_lipolytica_XP_500719.1	MRSLYINVPGLFPSTSLARETVHHRTEMLRTIRSSSRLGVRAM STAATRRAAQIGFHTRVPT VVTKAPTLRMQTTPFS SSAPAQTF GDKKYEHILTSTPVPKVALVTLNRPKALNALCTPLIKELNEAL QAADADPTIGAIVLTGSEKSFAAGADIKEMKDKTVTSVLNEN FIEEWGNMANIKKPIIAAVNGFALGGGCELAMMADIIYAGAK AKFGQPEIKLGVIPGAGGTQRLTRAIGLYRANHYILTGEMFTA QQAADWGLAAKVYEPAQLVDESVKAAAQIASYGQLAVQAA KASVHQSAEVGLRAGLEFERVRFHGLFGTHDQKEGMAAFAE KREPNFKNE
SEQ ID NO: 3807	pAA335	ggtttgattggatcatcatgacgcctggtgtggggttggatgataaaggcgatgcgttgggccagc agtataggactgttgatgaggtggttctgactggtaccgatgtgattattgtcgggagagggttgttt ggaaaaggaagagaccctgaggtggagggaaagagatacagggatgctggatggaaggcata cttgaagagaactggtcagttagaataaatattgtaataaataggtctatatacatacactaagcttct aggacgtcattgtagtcttcgaagttgtctgctagtttagttctcatgatttcgaaaaccaataacgca atggatgtagcagggatggtggttagtgcgttcctgacaaacccagagtacgccgcctcaaacca cgtcacattcgccctttgcttcatccgcatcacttgcttgaaggtatccacgtacgagttgtaatacac cttgaagaacggcttcgtctacgcgcgagacgaaagggcctcgtgatacgcctatttttataggtta atgtcatgataataatggtttcttagacgtcaggtggcacttttcggggaaatgtgcgcggaacccct atttgtttatttttctaaatacattcaaatatgtatccgctcatgagacaataaccctgataaatgcttcaa taatattgaaaaaggaagagtatgagtattcaacatttccgtgtcgcccttattcccttttttgcggcatt ttgccttcctgtttttgctcacccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtg cacgagtgggttacatcgaactggatctcaacagcggtaagatccttgagagttttcgccccgaag aacgttttccaatgatgagcacttttaaagttctgctatgtggcgcggtattatcccgtattgacgccg ggcaagagcaactcggtcgccgcatacactattctcagaatgacttggttgagtactcaccagtca

		cagaaaagcatcttacggatggcatgacagtaagagaattatgcagtgctgccataaccatgagtg ataacactgcggccaacttacttctgacaacgatcggaggaccgaaggagctaaccgcttttttgc acaacatgggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccatacca aacgacgagcgtgacaccacgatgcctgtagcaatggcaacaacgttgcgcaaactattaactgg cgaactacttactctagcttcccggcaacaattaatagactggatggaggcggataaagttgcagg accacttctgcgctcggcccttccggctggctggtttattgctgataaatctggagccggtgagcgt gggtctcgcggtatcattgcagcactggggccagatggtaagccctcccgtatcgtagttatctaca cgacggggagtcaggcaactatggatgaacgaaatagacagatcgctgagataggtgcctcact gattaagcattggtaactgtcagaccaagtttactcatatatactttagattgatttaaaacttcatttttta atttaaaaggatctaggtgaagatccttttgataatctcatgaccaaaatcccttaacgtgagttttcg ttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcctttttttctgcgcgta atctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagcta ccaactctttttccgaaggtaactggcttcagcagagcgcagataccaaatactgttcttctagtgta gccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctg ttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttac cggataaggcgcagcggtcgggctgaacgggggttcgtgcacacagcccagcttggagcga acgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaag ggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgaggga gcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtc gatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggccttttac ggttcctggccttttgctggccttttgctcacatgttctttcctgcgttatcccctgattctgtggataacc gtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtc agtgagcgaggaagcgagttcggcatggcagatcatcatgcctgcaggaagatgaagcgtatga gtattatgagtactgtcggacgttggaaggtggcagagttaagcccgagaaagcaaggaaggag tgggagatgatgagtgatgcggccaaagaggatgtgaaggctgcgtatctgtttttgatagctggt ggtagccgaatagaggaaggcaagcttgttcatattggatgatgatggtagatggtggctgccaaa gtggttgtaaatagaaaaaagtgggtttgggtctgttgatagttagtggtggcggctgtctgtgatta cgtcagcaagtagcacctcggcagttaaaacagcagcaacagaaaaaaaatgtgtgaaagtttga ttcccccacagtctaccacacccagagttccatttatccataatatcacaagcaatagaaaaataaaa aattatcaacaaatcacaacgaaaagattctgcaaaattattttcacttcttcttttgacttcctcttcttct tgttaggttctttccatattttcccctttaaacccatacacaacgcagccagaagagcccgggtctaga tgtgtgctcttccgctccaggcttgttatgactctagagagaagtgtgtgtgtgtgtgtgcgtttgtttta ctatacattcaacatgttctttttctttttttgatatttattccaactataattatacacagattcgtatatacttt actttaccctctttcgtagttttttaatttgatgatttttgagtttcatatccaaggtcaaaacccgaccga

		attcatatgcgacgggtacaacgagaattgtattgaattgatcaagaacatgatcttggtgttacaga acatcaagttcttggaccagactgagaatgcacagatatacaaggcgtcatgtgataaaatggatg agatttatccacaattgaagaaagagtttatggaaagtggtcaaccagaagctaaacaggaagaa gcaaacgaagaggtgaaacaagaagaagaaggtaaataagtattttgtattatataacaaacaaag taaggaatacagatttatacaataaattgccatactagtcacgtgagatatctcatccattccccaact cccaagaaaaaaaaaaagtgaaaaaaaaaatcaaacccaaagatcaacctccccatcatcatcgt catcaaacccccagctcaattcgcaatggttagcacaaaaacatacacagaaagggcatcagcac acccctccaaggttgcccaacgtttattccgcttaatggagtccaaaaagaccaacctctgcgcctc gatcgacgtgaccacaaccgccgagttcctttcgctcatcgacaagctcggtccccacatctgtctc gtgaagacgcacatcgatatcatctcagacttcagctacgagggcacgattgagccgttgcttgtg cttgcagagcgccacgggttcttgatattcgaggacaggaagtttgctgatatcggaaacaccgtg atgttgcagtacacctcgggggtataccggatcgcggcgtggagtgacatcacgaacgcgcacg gagtgactgggaagggcgtcgttgaagggttgaaacgcggtgcggaggggtagaaaaggaa aggggcgtgttgatgttggcggagttgtcgagtaaaggctcgttggcgcatggtgaatatacccgt gagacgatcgagattgcgaagagtgatcgggagttcgtgattgggttcatcgcgcagcgggacat ggggggtagagaagaag
SEQ ID NO: 3808	pAA918 sequence	ggccgcgagcactaggttttgataatttggttcttacagtttatgtattttgattcttccttttttagatactt ttttttatattttattattccttattgatgtaacgacagtcccactataattaacttaaactttgctgtaaatca gatgacaagtgtttccctgtttgcaggggagctctactagtttccttaattcatcttggttatgttgttctt gcaccgcaaggataagctgggcccgaatccaggccacgcataccaactatcgctgcccggaaa aaatacaaactagtaaactgcaaaaaaccatatataaacaataagaattttctctgtatcattcttcga aaaacctcaccttctctctctctctctctctcaccttccttttcccctatcacatcttcatcttcacatcac caactcaaaaccccctaagcagccaaagtaaagaacagagaatgattccacaacatcacactatgtt cttcgccgttaattaagaattcactggccgtcgttttacaacgtcgtgactgggaaaaccctggcgtt acccaacttaatcgccttgcagcacatccccctttcgccagctggcgtaatagcgaagaggcccg caccgatcgcccttcccaacagttgcgcagcctgaatggcgaatggcgcctgatgcggtattttct ccttacgcatctgtgcggtatttcacaccgcatatggtgcactctcagtacaatctgctctgatgccg catagttaagccagccccgacacccgccaacacccgctgacgcgccctgacgggcttgtctgct cccggcatccgcttacagacaagctgtgaccgtctccgggagctgcatgtgtcagaggttttcacc gtcatcaccgaaacgcgcgagacgaaagggcctcgtgatacgcctatttttataggttaatgtcatg ataataatggtttcttagacgtcaggtggcacttttcggggaaatgtgcgcggaacccctatttgttta ttttttctaaatacattcaaatatgtatccgctcatgagacaataaccctgataaatgcttcaataatattg aaaaaggaagagtatgagtattcaacatttccgtgtcgcccttattcccttttttgcggcattttgccttc ctgtttttgctcacccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagt

		gggttacatcgaactggatctcaacagcggtaagatccttgagagttttcgccccgaagaacgtttt ccaatgatgagcacttttaaagttctgctatgtggcgcggtattatcccgtattgacgccgggcaag agcaactcggtcgccgcatacactattctcagaatgacttggttgagtactcaccagtcacagaaaa gcatcttacggatggcatgacagtaagagaattatgcagtgctgccataaccatgagtgataacact gcggccaacttacttctgacaacgatcggaggaccgaaggagctaaccgcttttttgcacaacatg ggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacga gcgtgacaccacgatgcctgtagcaatggcaacaacgttgcgcaaactattaactggcgaactact tactctagcttcccggcaacaattaatagactggatggaggcggataaagttgcaggaccacttct gcgctcggcccttccggctggctggtttattgctgataaatctggagccggtgagcgtgggtctcg cggtatcattgcagcactggggccagatggtaagccctcccgtatcgtagttatctacacgacggg gagtcaggcaactatggatgaacgaaatagacagatcgctgagataggtgcctcactgattaagc attggtaactgtcagaccaagtttactcatatatactttagattgatttaaaacttcatttttaatttaaaag gatctaggtgaagatccttttgataatctcatgaccaaaatcccttaacgtgagttttcgttccactga gcgtcagaccccgtagaaaagatcaaaggatcttcttgagatccttttttttctgcgcgtaatctgctgc ttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactcttt ttccgaaggtaactggcttcagcagagcgcagataccaaatactgtccttctagtgtagccgtagtt aggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtg gctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataag gcgcagcggtcgggctgaacgggggggttcgtgcacacagcccagcttggagcgaacgacctac accgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaagggagaaag gcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttccag ggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgt gatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggccttttacggttcctg gccttttgctggccttttgctcacatgttctttcctgcgttatcccctgattctgtggataaccgtattacc gcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtcagtgagc gaggaagcggaagagcgcccaatacgcaaaccgcctctccccgcgcgttggccgattcattaat gcagctggcacgacaggtttcccgactggaaagcgggcagtgagcgcaacgcaattaatgtgag ttagctcactcattaggcaccccaggctttacactttatgcttccggctcgtatgttgtgtggaattgtg agcggataacaatttcacacaggaaacagctatgaccatgattacgccaagcttttaattaaagata atcacaggggtagagaccttggttatgggctgattggaggaagaaatggggtgatttggcacttga cagcgcgagagtggttaacacctggtttccctcatttgggttctgacatttgataagttgaaagaaca atgcagaattcacatggctaatttggcctcggttccacaacgcactcagcattaaaaaaaaaatacg caatggcagctcggtcgacgcagcagaagcgccgacgtaccgtcgcgttgccccgcccatgcct cgccgacccctccaccgccatcgtttgcccattgtttgtggtagtgcgccgtgacacaaaaacttgt

		cctgtcacatgctgaagttacaccaacataactactatgggattacgtaatcaaaaatttcacagtttt aacaaaaaaaaaatcatacaatcaacattgggacatcttgccctcccccacaaaacttgcttctgca tcaatcatatataaacatcatgaaataagcctaaactcacttctttttttttcatccttcctacttcttctttc atagtaactacttttttttttattaccacacttattcattcataccacgctatcggccggccgcatgcggat ccgagctcgcggccgccagtgtgatggatatctgcagcggccgctaaatattgtaataaataggtc tatatacatacactaagcttctaggacgtcattgtagtcttcgaagttgtctgctagtttagttctcatga tttcgaaaaccaataacgcaatggatgtagcagggatggtggttagtgcgttcctgacaaacccag agtacgccgcctcaaaccacgtcacattcgccctttgcttcatccgcatcacttgcttgaaggtatcc acgtacgagttgtaatacaccttgaagaacggcttcgtctcgacgggtacaacgagaattgtattga attgatcaagaacatgatcttggtgttacagaacatcaagttcttggaccagactgagaatgcacag atatacaaggcgtcatgtgataaaatggatgagatttatccacaattgaagaaagagtttatggaaa gtggtcaaccagaagctaaacaggaagaagcaaacgaagaggtgaaacaagaagaagaaggt aaataagtattttgtattatataacaaacaaagtaaggaatacagatttatacaataaattgccatacaa gtcacgtgagatatctcatccattccccaactcccaagaaaaaaaaaaagtgaaaaaaaaaatcaa acccaaagatcaacctccccatcatcatcgtcatcaaacccccagctcaattcgcaatggttagcac aaaaacatacacagaaagggcatcagcacacccctccaaggttgcccaacgtttattccgcttaat ggagtccaaaaagaccaacctctgcgcctcgatcgacgtgaccacaaccgccgagttcctttcgc tcatcgacaagctcggtccccacatctgtctcgtgaagacgcacatcgatatcatctcagacttcag ctacgagggcacgattgagccgttgcttgtgcttgcagagcgccacgggttcttgatattcgagga caggaagtttgctgatatcggaaacaccgtgatgttgcagtacacctcgggggtataccggatcgc ggcgtggagtgacatcacgaacgcgcacggagtgactgggaagggcgtcgttgaagggttgaa acgcggtgcggaggggtagaaaaggaaaggggcgtgttgatgttggcggagttgtcgagtaa aggctcgttggcgcatggtgaatatacccgtgagacgatcgagattgcgaagagtgatcgggagt tcgtgattgggttcatcgcgcagcgggacatgggggtagagaagaagggtttgattggatcatc atgacgcctggtgtggggttggatgataaaggcgatgcgttgggccagcagtataggactgttgat gaggtggttctgactggtaccgatgtgattattgtcgggagagggttgtttggaaaaggaagagac cctgaggtggagggaaagagatacagggatgctggatggaaggcatacttgaagagaactggtc agttagaataaatattgtaataaataggtctatatacatacactaagcttctaggacgtcattgtagtctt cgaagttgtctgctagtttagttctcatgatttcgaaaaccaataacgcaatggatgtagcagggatg gtggttagtgcgttcctgacaaacccagagtacgccgcctcaaaccacgtcacattcgcccttgct tcatccgcatcacttgcttgaaggtatccacgtacgagttgtaatacaccttgaagaacggcttcgtc tgc
SEQ ID	Rat Dcilp	MATAMTVSSKLLGLLMQQLRGTRQLYFNVSLRSLSSSAQEAS KRIPEEVSDHNYESIQVTSAQKHVLHVQLNRPEKRNAMNRAF

NO: 3809	>gil12018256lref]NP_0721 16.1 1 delta(3,5)-Delta(2,4)-dienoyl-CoA isomerase, mitochondrial precursor [Rattus norvegicus]	WRELVECFQKISKDSDCRAVVVSGAGKMFTSGIDLMDMASDI LQPPGDDVARIAWYLRDLISRYQKTFTVIEKCPKPVITAIHGGC IGGGVDLISACDIRYCTQDAFFQVKEVDVGLAADVGTLQRLP KVIGNRSLVNELTFTARKMMADEALDSGLVSRVFPDKDVML NAAFALAADISSKSPVAVQGSKINLIYSRDHSVDESLDYMAT WNMSMLQTQDIIKSVQAAMEKKDSKSITFSKL
SEQ ID NO: 3810	Mouse Dcilp >gil7949037lref]NP_058052.1l delta(3,5)-Delta(2,4)-dienoyl-CoA isomerase, mitochondrial precursor [Mus musculus]	MATAMTVSSKLRGLLMQQLRGTSQLYFNISLRSLSSSAQEAS KRAPEEVSDHNYESIQVTSAQKHVLHVQLNRPEKRNAMNRA FWRELVECFQKISKDSDCRAVVVSGAGKMFTSGIDLMDMAS ELMQPSGDDAARIAWYLRDLISKYQKTFTVIEKCPKPVIAAIH GGCIGGGVDLVSACDIRYCTQDAFFQIKEVDMGLAADVGTLQ RLPKVIGNQSLVNELTFSARKMMADEALDSGLVSRVFQDKD AMLNAAFALAADISSKSPVAVQGSKINLIYSRDHSVDESLDY MATWNMSMLQTQDIIKSVQAAMEKRDTKSITFSKL
SEQ ID NO: 381 1	CtDcilp (putative) >gil255732940lref]XP_002551 **393.11** hypothetical protein CTRG_05691 [Candida tropicalis MYA-3404]	MTFNPESYKYEFYTVTQIEDGFAHVQYTNPKTLNAFTEQNWK DYGEIFRRLDKEEDIQLILFSSGVEKSFSSGLNLKTAANLFKSE TNSTEEAIKELHEHITGFQYDIGTPARISTPTIGILNGLNLGLAL DIAAAFTIRIAVEGAQFSIAEVNIGITADIGSLQRLPAIINNKSLL FQHALMGDKWGTDEAMKLGFVSTVVPTIEEGIEYAKTLGEKI CSVPAWAIKGTKKHIQDIINGTTVEQGLKDVANWNAKNITLS RGKL
SEQ	CvDcilp	MTFDPSTYKYEFFTVSEIQEGFAHVQYTNPKTLNAFSEQNWK

ID NO: 3812	(putative)	DYGSIFARLDAEEDIQVILFSSGVPRSFSSGLNLKTAANLFSSE QQSTAEAIKDLHVHIRDFQSDIGTPARISTPTIAILNGLNLGLAL DMAAAYTIRIAVEGAQFSIAEVNIGITADIGSLQRLPALINNKS LLFQHALLGDKWGPEEAEKLGFVSTVVPSVEAGIEYAKALGA KICEAPAWAIKGTKKHIQDVINGTSVDQGLADVANWNARNIT LSKGKL
SEQ ID NO: 3813	ScDcilp (YOR180c) >gi|398365593| ref]NP_0 14823 .3| Dcilp [Saccharomyce s cerevisiae S288c]	MSSRVCYHINGPFFIIKLIDPKHLNSLTFEDFVYIALLLHKANDI DSVLFTVLQSSGKYFSSGGKFSAVNKLNDGDVTSEVEKVSKL VSAISSPNIFVANAFAIHKKVLVCCLNGPAIGLSASLVALCDIV YSQNDSVFLLFPFSNLGFVAEVGTSVTLTQKLGINSANEHMIF STPVLFKELIGTIITKNYQLTNTETFNEKVLQDIKQNLEGLYPK SVLGMKELLHSEMKQKLIKAQAMETNGTLPFWASGEPFKRF KQLQEGNRRHKL

Example 55: Examples of certain non-limiting embodiments

Listed hereafter are non-limiting examples of certain embodiments of the technology.

A1. A genetically modified yeast, comprising an active, modified endogenous acyl-CoA oxidase polypeptide or an active, modified endogenous acyl-CoA dehydrogenase polypeptide, which yeast is capable of producing a diacid from a feedstock comprising one or more components from a vegetable oil.

A2. The genetically modified yeast of embodiment A1, wherein the yeast is a genetically modified Candida spp. yeast.

A2.1. The genetically modified yeast of embodiment A2, wherein the Candida spp. yeast is chosen from C. tropicalis and C. viswanathii.

A2.2. The genetically modified yeast of embodiment A1, wherein the Candida spp. yeast is a genetically modified ATCC20336 yeast.

A2.3. The genetically modified yeast of any one of embodiments A2 to A2.2, wherein the endogenous acyl-CoA oxidase polypeptide is a POX4 polypeptide.

A2.4. The genetically modified yeast of embodiment A2.3, wherein the POX4 polypeptide comprises a modified amino acid sequence of SEQ ID NO: 30.

A2.5. The genetically modified yeast of embodiment A2.3 or A2.4, wherein the POX4 polypeptide comprises an amino acid modification at one or more amino acid positions chosen from 88, 90, 96, 98, 99, 100, 102, 103, 302, 309, 310, 473, 474, 475, 476, 477, 478, 479, 480, 481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 501, 502, 503, 504 and 505.

A2.6. The genetically modified yeast of any one of embodiments A2 to A2.2, wherein the endogenous acyl-CoA oxidase polypeptide is a POX5 polypeptide.

A2.7. The genetically modified yeast of embodiment A2.6, wherein the POX5 polypeptide comprises a modified amino acid sequence of SEQ ID NO: 32.

A2.8. The genetically modified yeast of embodiment A2.6 or A2.7, wherein the POX5 polypeptide comprises an amino acid modification at one or more amino acid positions chosen from 81, 82, 83, 84, 85, 86, 88, 93, 94, 95, 96, 98, 102, 284, 287, 290, 291, 292, 294, 295, 428, 429, 436, 453, 454, 455, 456, 457, 458, 459, 460, 461, 462 and 463.

A2.9. The genetically modified yeast of any one of embodiments A2 to A2.2, wherein the acyl-CoA dehydrogenase polypeptide is chosen from ACAD, VLCAD, LCAD, MCAD and SCAD polypeptides.

A2. 10. The genetically modified yeast of embodiment A2.9, wherein the acyl-CoA dehydrogenase polypeptide comprises a modified amino acid sequence of SEQ ID NO: 3685.

A2. 11. The genetically modified yeast of embodiment A2.9 or A2. 10, wherein the the acyl-CoA dehydrogenase polypeptide comprises an amino acid modification at VLCAD position 461.

A2. 12. The genetically modified yeast of any one of embodiments 2.5, 2.8 and 2.1 1, wherein at least one of the amino acid modifications is an amino acid substitution.

A2. 13. The genetically modified yeast of embodiment A2. 12, wherein at least one of the one or more amino acid substitutions is conservative.

A2. 14. The genetically modified yeast of embodiment A2. 12, wherein at least one of the one or more amino acid substitutions is not conservative.

A3. The genetically modified yeast of embodiment Al, wherein the yeast is a genetically modified Yarrowia spp. yeast.

A3.1. The genetically modified yeast of embodiment A3.1, wherein the Yarrowia spp. yeast is Y. lipolytica.

A3.2. The genetically modified yeast of embodiment A3 or A3.1, wherein the endogenous acyl-CoA oxidase polypeptide is chosen from a POX1 polypeptide, POX2 polypeptide, POX3 polypeptide, POX4 polypeptide, POX5 polypeptide or POX6 polypeptide.

A3.3. The genetically modified yeast of embodiment 3.2, wherein the endogenous acyl-CoA oxidase polypeptide is chosen from SEQ ID NOs: 3778 to 3783.

A4. The genetically modified yeast of embodiment Al, wherein the yeast is a genetically modified Pichia spp. yeast.

A4. 1. The genetically modified yeast of embodiment A4. 1, wherein the Pichia spp. yeast is chosen from P. pastoris, P. membranifaciens, P. kluyveri, P. guilliermondii, P. heedii and P. subpelliculosa.

A5. The genetically modified yeast of embodiment Al, wherein the yeast is a genetically modified Saccharomyces spp. yeast.

A5.1. The genetically modified yeast of embodiment A5.1, wherein the Saccharomyces spp. yeast is chosen from S. cerevisiae, S. bayanus, S. pastorianus and S. carlsbergensis.

A6. The genetically modified yeast of embodiment Al, wherein the yeast is a genetically modified Kluyveromyces spp. yeast.

A6. 1. The genetically modified yeast of embodiment A6. 1, wherein the Kluyveromyces spp. yeast is chosen from K. lactis and K. marxianus.

A7. The genetically modified yeast of any one of embodiments A1 to A6.1, wherein the modified endogenous acyl-CoA oxidase polypeptide comprises an amino acid modification in the N-terminal loop.

A8. The genetically modified yeast of any one of embodiments A1 to A7, wherein the modified endogenous acyl-CoA oxidase polypeptide comprises an amino acid modification in the D alpha helix.

A9. The genetically modified yeast of any one of embodiments A1 to A8, wherein the modified endogenous acyl-CoA oxidase polypeptide comprises an amino acid modification in the loop between the D alpha helix and the E' alpha helix.

A10. The genetically modified yeast of any one of embodiments A1 to A9, wherein the modified endogenous acyl-CoA oxidase polypeptide comprises an amino acid modification to an amino acid in effective contact with carbons 6 to 9 in a feedstock component.

A1 1. The genetically modified yeast of any one of embodiments A1 to A10, wherein the modified endogenous acyl-CoA oxidase polypeptide comprises an amino acid modification to an amino acid in effective contact with carbons 10 to 12 in a feedstock component.

A12. The genetically modified yeast of any one of embodiments A1 to A1 1, wherein the modified endogenous acyl-CoA oxidase polypeptide comprises an amino acid modification in the L alpha helix.

A13. The genetically modified yeast of any one of embodiments A1 to A12, wherein the modified endogenous acyl-CoA oxidase polypeptide comprises an amino acid modification in the loop C-terminal to the L alpha helix.

A14. The genetically modified yeast of any one of embodiments A1 to A13, wherein the modified endogenous acyl-CoA oxidase polypeptide comprises an amino acid modification in the loop between the L alpha helix and the M alpha helix.

A15. The genetically modified yeast of any one of embodiments A7 to A14, wherein the amino acid modification comprises an amino acid substitution.

A16. The genetically modified yeast of embodiment A15, wherein the amino acid substitution is conservative.

A17. The genetically modified yeast of embodiment A15, wherein the amino acid substitution is not conservative.

A17.1. A genetically modified yeast comprising (i) an active, modified endogenous acyl-CoA oxidase polypeptide or an active, modified endogenous acyl-CoA dehydrogenase polypeptide and (ii) a genetic modification that alters an enoyl CoA isomerase activity, wherein the yeast is capable of producing a diacid from a feedstock comprising one or more components from a vegetable oil.

A18. The genetically modified yeast of any one of embodiments A1 to A17.1, which comprises a genetic modification that reduces an enoyl CoA isomerase activity.

A18.1 The genetically modified yeast of embodiment A18, which comprises a disruption, deletion or knockout of (i) a polynucleotide that encodes an enoyl CoA isomerase polypeptide or (ii) a promoter operably linked to a polynucleotide that encodes an enoyl CoA isomerase polypeptide, whereby the enoyl CoA isomerase activity is reduced or removed.

A19. The genetically modified yeast of embodiment A18.1, wherein the genetic modification disrupts a polynucleotide that encodes an enoyl CoA isomerase polypeptide.

A19. 1. The genetically modified yeast of any one of embodiments A1 to A19, which comprises a genetic modification that increases an enoyl CoA isomerase activity.

A19.2. The genetically modified yeast of embodiment A19.1, which comprises (i) an increased copy number of a polynucleotide that encodes an enoyl CoA isomerase polypeptide or (ii) a promoter inserted and operably linked to a polynucleotide that encodes an enoyl CoA isomerase polypeptide.

A20. The genetically modified yeast of any one of embodiments A18.1, A18.2 and A19.2, wherein the enoyl CoA isomerase polypeptide is a polypeptide native to the yeast.

A20.1. The genetically modified yeast of any one of embodiments A17.1 to A20, wherein the activity of one or more or all enoyl CoA isomerase polypeptide types in the yeast is altered.

A20.2. The genetically modified yeast of embodiment A20. 1, wherein the yeast comprises two enoyl CoA isomerase polypeptide types, and one or both of the polypeptide types are altered.
A21. The genetically modified yeast of any one of embodiments A17.1 to A20, wherein the enoyl CoA isomerase activity is provided by a polypeptide present in aCandida spp. yeast.
A22. The genetically modified yeast of any one of embodiments A17.1 to A21, wherein the enoyl CoA isomerase activity is provided by a polypeptide comprising the amino acid sequence of SEQ ID NO: 3675, 3677, 3800, 3802, 3803 or 3805.

A22. 1 The genetically modified yeast of any one of embodiments A17.1 to A21, wherein the yeast is a Yarrowia yeast.

A22.2. The genetically modified yeast of embodiment A22. 1, wherein the yeast is a Y. lipolytica yeast.

A22.3. The genetically modified yeast of any one of embodiments A22.2, wherein the enoyl CoA isomerase activity is provided by a polypeptide present in a Yarrowia yeast or a Y. lipolytica yeast.

A22.4. The genetically modified yeast of embodiment A22.2, wherein the enoyl CoA isomerase activity is provided by a polypeptide comprising the amino acid sequence of SEQ ID NO: 3804 or 3806.

A23. The genetically modified yeast of any one of embodiments A1 to A22.4, which comprises a genetic modification that reduces the cytoplasmic activity of an acyl-CoA synthetase (ACS) polypeptide.

A24. The genetically modified yeast of any one of embodiments A1 to A23, which comprises a genetic modification that reduces the peroxisomal activity of an acyl-CoA synthetase (ACS) polypeptide.

A25. The genetically modified yeast of embodiment A23 or A24, wherein the genetic modification disrupts a polynucleotide that encodes the acyl-CoA synthetase (ACS) polypeptide.

A26. The genetically modified yeast of embodiment A25, wherein the genetic modification disrupts an ACS1 polypeptide or ACS2 polypeptide.

A27. The genetically modified yeast of any one of embodiments A23 to A26, wherein the genetic modification disrupts a polynucleotide that encodes a long-chain acyl-CoA synthetase polypeptide.

A28. The genetically modified yeast of embodiment A27, wherein the genetic modification disrupts a FAT1 polypeptide.

A29. The genetically modified yeast of any one of embodiments A23 to A28, wherein the acyl-CoA synthetase (ACS) polypeptide is a polypeptide native to the yeast.

A30. The genetically modified yeast of embodiment A29, wherein the yeast is a Candida spp. yeast.

A31. The genetically modified yeast of embodiment A30, wherein the acyl-CoA synthetase (ACS) polypeptide comprises an amino acid sequence chosen from SEQ ID NOs: 80, 82, 84, 158 and 159.

A32. The genetically modified yeast of embodiment A30, wherein the FAT1 polypeptide comprises the amino acid sequence of SEQ ID NO: 90.

A33. The genetically modified yeast of any one of embodiments A1 to A32, which comprises a genetic modification that reduces the activity of a PXA polypeptide.

A34. The genetically modified yeast of embodiment A33, wherein the genetic modification disrupts a polynucleotide that encodes the PXA polypeptide.

A35. The genetically modified yeast of embodiment A33 or A34, wherein the PXA polypeptide is a PXA1 polypeptide or a PXA2 polypeptide, or a PXA1 polypeptide and a PXA2 polypeptide.

A36. The genetically modified yeast of any one of embodiments A33 to A35, wherein the PXA polypeptide is native to the yeast.

A37. The genetically modified yeast of embodiment A36, wherein the yeast is a Candida spp. yeast.

A38. The genetically modified yeast of embodiment A37, wherein the PXA1 polypeptide comprises the amino acid sequence of SEQ ID NO: 92.

A39. The genetically modified yeast of embodiment A37, wherein the PXA2 polypeptide comprises the amino acid sequence of SEQ ID NO: 94.

A40. The genetically modified yeast of any one of embodiments A1 to A39, comprising an active, modified endogenous acyl-CoA oxidase polypeptide and no active, modified endogenous acyl-CoA dehydrogenase polypeptide.

A41. The genetically modified yeast of any one of embodiments A1 to A39, comprising no active, modified endogenous acyl-CoA oxidase polypeptide and an active, modified endogenous acyl-CoA dehydrogenase polypeptide.

A42. A method for producing a diacid, comprising:

(1) contacting a genetically modified yeast of any of embodiments A1 to A41 with a feedstock comprising one or more components from a feedstock containing at least one unsaturated vegetable fatty acid capable of being converted by the yeast to a diacid;

(2) culturing the yeast under conditions in which the diacid is produced from the feedstock; and

(3) hydrogenating the diacid to remove any unsaturation in the carbon chain.

A43. The method of A42, wherein the unsaturated vegeteable fatty acid is linoleic acid or linolenic acid.

A44. The method of A43, wherein the diacid is dodecanedioic acid or sebacic acid.

B1. A genetically modified yeast, comprising a heterologous acyl-CoA oxidase polypeptide or a heterologous acyl-CoA dehydrogenase polypeptide, which yeast is capable of producing a diacid from a feedstock comprising one or more components from a vegetable oil.

B2. The genetically modified yeast of embodiment B1, wherein the heterologous acyl-CoA oxidase polypeptide is a native polypeptide.

B3. The genetically modified yeast of embodiment B1, wherein the heterologous acyl-CoA oxidase polypeptide is an active, modified polypeptide.

B4. The genetically modified yeast of embodiment B1, wherein the heterologous acyl-CoA dehydrogenase polypeptide is a native polypeptide.

B5. The genetically modified yeast of embodiment B1, wherein the heterologous acyl-CoA dehydrogenase polypeptide is an active, modified polypeptide.

B6. The genetically modified yeast of embodiment B1 or B2, wherein the heterologous acyl-CoA oxidase polypeptide is chosen from a polypeptide having an amino acid sequence set forth in SEQ ID NO: 51 to SEQ ID NO: 3673.

B7. The genetically modified yeast of embodiment B1 or B4, wherein the heterologous acyl-CoA dehydrogenase polypeptide is chosen from SEQ ID NOs: 3679 to 3683, 3686, 3689, 3691, 3693, 3695, 3697, 3699, 3701 and 3703.

B8. The genetically modified yeast of any one of embodiments B1 to B7, which is chosen from a Candida spp. yeast, Yarrowia spp. yeast, Pichia spp. yeast, Saccharomyces spp. yeast and Kluyveromyces spp. yeast.

B9. The genetically modified yeast of embodiment B8, which is chosen from C. tropicalis, C. viswanathii, Y. lipolytica, P. pastoris, P. membranifaciens, P. kluyveri, P. guilliermondii, P. heedii, P. subpelliculosa, S. cerevisiae, S. bayanus, S. pastorianus, S. carlsbergensis, K. lactis and K. marxianus.

BIO. The genetically modified yeast of any one of embodiments B1 to B9, which comprises a genetic modification that reduces the activity of an enoyl CoA isomerase polypeptide.

B11. The genetically modified yeast of embodiment BIO, wherein the genetic modification disrupts a polynucleotide that encodes the enoyl CoA isomerase polypeptide.

B12. The genetically modified yeast of embodiment BIO or B11, wherein the enoyl CoA isomerase polypeptide is a polypeptide native to the yeast.

B13. The genetically modified yeast of embodiment B12, wherein the yeast is a Candida spp. yeast.

B14. The genetically modified yeast of embodiment B13, wherein the enoyl CoA isomerase polypeptide comprises the amino acid sequence of SEQ ID NO: 3675 or 3677.

B15. The genetically modified yeast of any one of embodiments B1 to B14, which comprises a genetic modification that reduces the cytoplasmic activity of an acyl-CoA synthetase (ACS) polypeptide.

B16. The genetically modified yeast of any one of embodiments B1 to B15, which comprises a genetic modification that reduces the peroxisomal activity of an acyl-CoA synthetase (ACS) polypeptide.

B17. The genetically modified yeast of embodiment B15 or B16, wherein the genetic modification disrupts a polynucleotide that encodes the acyl-CoA synthetase (ACS) polypeptide.

B18. The genetically modified yeast of embodiment B17, wherein the genetic modification disrupts an ACS1 polypeptide or ACS2 polypeptide.

B19. The genetically modified yeast of any one of embodiments B15 to B18, wherein the genetic modification disrupts a polynucleotide that encodes a long-chain acyl-CoA synthetase polypeptide.

B20. The genetically modified yeast of embodiment B19, wherein the genetic modification disrupts a FAT1 polypeptide.

B21. The genetically modified yeast of any one of embodiments B15 to B20, wherein the acyl-CoA synthetase (ACS) polypeptide is a polypeptide native to the yeast.

B22. The genetically modified yeast of embodiment B21, wherein the yeast is a Candida spp. yeast.

B23. The genetically modified yeast of embodiment B22, wherein the acyl-CoA synthetase (ACS) polypeptide comprises an amino acid sequence chosen from SEQ ID NOs: 80, 82, 84, 158 and 159.

B24. The genetically modified yeast of embodiment B21, wherein the FAT1 polypeptide comprises the amino acid sequence of SEQ ID NO: 90.

B25. The genetically modified yeast of any one of embodiments B1 to B24, which comprises a genetic modification that reduces the activity of a PXA polypeptide.

B26. The genetically modified yeast of embodiment B25, wherein the genetic modification disrupts a polynucleotide that encodes the PXA polypeptide.

B27. The genetically modified yeast of embodiment B25 or B26, wherein the PXA polypeptide is a PXA1 polypeptide or a PXA2 polypeptide, or a PXA1 polypeptide and a PXA2 polypeptide.

B28. The genetically modified yeast of any one of embodiments B25 to B27, wherein the PXA polypeptide is native to the yeast.

B29. The genetically modified yeast of embodiment B28, wherein the yeast is a Candida spp. yeast.

B30. The genetically modified yeast of embodiment B29, wherein the PXA1 polypeptide comprises the amino acid sequence of SEQ ID NO: 92.

B31. The genetically modified yeast of embodiment B29, wherein the PXA2 polypeptide comprises the amino acid sequence of SEQ ID NO: 94.

B32. The genetically modified yeast of any one of embodiments B1 to B31, comprising an active, modified endogenous acyl-CoA oxidase polypeptide and no active, modified endogenous acyl-CoA dehydrogenase polypeptide.

B33. The genetically modified yeast of any one of embodiments B1 to B31, comprising no active, modified endogenous acyl-CoA oxidase polypeptide and an active, modified endogenous acyl-CoA dehydrogenase polypeptide.

CI. The genetically modified yeast of any one of embodiments A1 to A41 and B1 to B33, comprising one or more genetic modifications that reduce the activity of one or more native endogenous acyl-CoA oxidase polypeptides.

C2. The genetically modified yeast of embodiment CI, comprising genetic modifications that reduce the activity of all native endogenous acyl-CoA oxidase polypeptides.

C3. The genetically modified yeast of embodiment C1 or C2, wherein the genetic modifications partially block beta oxidation activity.

C4. The genetically modified yeast of any one of embodiments A1 to A41, B1 to B33, and C1 to C3, wherein the diacid is a C4 to C24 diacid.

C5. The genetically modified yeast of embodiment C4, wherein the diacid is a CIO, C12, CI4, CI6, CI8 or C20 diacid.

C6. The genetically modified yeast of embodiment C5, wherein the diacid is a CIO diacid.

C7. The genetically modified yeast of embodiment C5, wherein the diacid is a C12 diacid.

C8. The genetically modified yeast of embodiment C5, wherein the diacid is a CI8 diacid.

C9. The genetically modified yeast of any one of embodiments C4 to C8, wherein the diacid contains no unsaturation.

CIO. The genetically modified yeast of any one of embodiments C4 to C8, wherein the diacid contains one or more unsaturations.

CI0.1. The genetically modified yeast of any one of embodiments C4 to CIO, wherein the diacid is the predominant diacid in a mixture of diacids.

CI1. The genetically modified yeast of any one of embodiments A1 to A41, B1 to B33, and CI to CI0.1, wherein the feedstock comprises a substantially pure oil.

CI2. The genetically modified yeast of any one of embodiments A1 to A41, B1 to B33, and CI to CIO, wherein the feedstock comprises a plurality of fatty acids.

C13. The genetically modified yeast of embodiment C12, wherein the feedstock comprises a soapstock.

C14. The genetically modified yeast of embodiment C12, wherein the feedstock comprises a fatty acid distillate.

C15. The genetically modified yeast of any one of embodiments A1 to A41, B1 to B33, and CI to CI4, wherein the vegetable oil is from a plant chosen from palm, palm kernel, coconut, soy, safflower, canola, palm, palm kernel or combination thereof.

D1. A method for producing a diacid, comprising:

contacting a genetically modified yeast of any one of embodiments 1 to A41, B1 to B33, and C1 to C15 with a feedstock comprising one or more components from a vegetable oil capable of being converted by the yeast to a diacid; and

culturing the yeast under conditions in which the diacid is produced from the feedstock.

D2. The method of embodiment Dl, wherein the diacid is a C4 to C24 diacid.

D3. The method of embodiment D2, wherein the diacid is a C10, C12, C14, C16, C18 or C20 diacid.

D4. The method of embodiment D3, wherein the diacid is a C10 diacid.

D5. The method of embodiment D3, wherein the diacid is a C12 diacid.

D6. The method of embodiment D3, wherein the diacid is a C18 diacid.

D7. The method of any one of embodiments D1 to D6, wherein the diacid contains no unsaturation.

D8. The method of any one of embodiments D1 to D6, wherein the diacid contains one or more unsaturations.

D8.1. The method of any one of embodiments D2 to D8, wherein the diacid is the predominant diacid in a mixture of diacids.

D9. The method of any one of embodiments D1 to D8.1, wherein the feedstock comprises a substantially pure oil.

D10. The method of any one of embodiments D1 to D8, wherein the feedstock comprises a plurality of fatty acids.

D11. The method of embodiment D10, wherein the feedstock comprises a soapstock.

D12. The method of embodiment D10, wherein the feedstock comprises a fatty acid distillate.

D13. The method of any one of embodiments D1 to D12, wherein the vegetable oil is from a plant chosen from palm, palm kernel, coconut, soy, safflower, canola, palm, palm kernel or combination thereof.

El. An isolated nucleic acid, comprising a polynucleotide that encodes a modified acyl-CoA oxidase polypeptide from a yeast.

E2. The isolated nucleic acid of embodiment El, wherein the modified acyl-CoA oxidase polypeptide comprises an amino acid modification in the N-terminal loop.

E3. The isolated nucleic acid of embodiment El, wherein the modified endogenous acyl-CoA oxidase polypeptide comprises an amino acid modification in the D alpha helix.

E4. The isolated nucleic acid of embodiment El, wherein the modified endogenous acyl-CoA oxidase polypeptide comprises an amino acid modification in the loop between the D alpha helix and the E' alpha helix.

E5. The isolated nucleic acid of embodiment El, wherein the modified endogenous acyl-CoA oxidase polypeptide comprises an amino acid modification to an amino acid in effective contact with carbons 6 to 9 in a feedstock component.

E6. The isolated nucleic acid of embodiment El, wherein the modified endogenous acyl-CoA oxidase polypeptide comprises an amino acid modification to an amino acid in effective contact with carbons 10 to 12 in a feedstock component.

E7. The isolated nucleic acid of embodiment El, wherein the modified endogenous acyl-CoA oxidase polypeptide comprises an amino acid modification in the L alpha helix.

E8. The isolated nucleic acid of embodiment El, wherein the modified endogenous acyl-CoA oxidase polypeptide comprises an amino acid modification in the loop C-terminal to the L alpha helix.

E9. The isolated nucleic acid of embodiment El, wherein the modified endogenous acyl-CoA oxidase polypeptide comprises an amino acid modification in the loop between the L alpha helix and the M alpha helix.

E10. The isolated nucleic acid of any one of embodiments E2 to E9, wherein the amino acid modification comprises an amino acid substitution.

E11. The isolated nucleic acid of embodiment E10, wherein the amino acid substitution is conservative.

E12. The isolated nucleic acid of embodiment E10, wherein the amino acid substitution is not conservative.

E13. The isolated nucleic acid of any one of embodiments E1 to E12, wherein the yeast is chosen from a Candida spp. yeast, Yarrowia spp. yeast, Pichia spp. yeast, Saccharomyces spp. yeast and Kluyveromyces spp. yeast.

E14. The isolated nucleic acid of embodiment E13, wherein the yeast is chosen from C. tropicalis, C. viswanathii, Y. lipolytica, P. pastoris, P. membranifaciens, P. kluyveri, P. guilliermondii, P. heedii, P. subpelliculosa, S. cerevisiae, S. bayanus, S. pastorianus, S. carlsbergensis, K. lactis and K. marxianus.

E15. The isolated nucleic acid of embodiment E13, wherein the yeast is a Candida spp. yeast.

E16. The isolated nucleic acid of embodiment E15, wherein the yeast is chosen from C. tropicalis and C. viswanathii.

E17. The isolated nucleic acid of embodiment E16, wherein the yeast is a genetically modified ATCC20336 yeast.

E18. The isolated nucleic acid of any one of embodiments E15 to E17, wherein the endogenous acyl-CoA oxidase polypeptide is a POX4 polypeptide.

E19. The isolated nucleic acid of embodiment E18, wherein the POX4 polypeptide comprises a modified amino acid sequence of SEQ ID NO: 30.

E20. The isolated nucleic acid of embodiment E18 or E19, wherein the POX4 polypeptide comprises amino acid modifications at one or more amino acid positions chosen from 88, 90, 96, 98, 99, 100, 102, 103, 302, 309, 310, 473, 474, 475, 476, 477, 478, 479, 480, 481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 501, 502, 503, 504 and 505.

E21. The isolated nucleic acid of any one of embodiments E15 to E17, wherein the endogenous acyl-CoA oxidase polypeptide is a POX5 polypeptide.

E22. The isolated nucleic acid of embodiment E21, wherein the POX5 polypeptide comprises a modified amino acid sequence of SEQ ID NO: 32.

E23. The isolated nucleic acid of embodiment E21 or E22, wherein the POX5 polypeptide comprises amino acid modifications at one or more amino acid positions chosen from 81, 82, 83, 84, 85, 86, 88, 93, 94, 95, 96, 98, 102, 284, 287, 290, 291, 292, 294, 295, 436, 453, 454, 455, 456, 457, 458, 459, 460, 461, 462 and 463.

E24. The isolated nucleic acid of embodiment E20 or E23, wherein at least one of the amino acid modifications is an amino acid substitution.

E25. The isolated nucleic acid of embodiment E24, wherein at least one of the amino acid substitutions is conservative.

E26. The isolated nucleic acid of embodiment E24, wherein at least one of the amino acid substitutions is non-conservative.

F1. The isolated nucleic acid of any one of embodiments E1 to E26, which is an expression vector.

F2. A cell comprising a nucleic acid of any one of embodiments E1 to F1.

F3. The cell of embodiment F2, which is a bacterium.

F4. The cell of embodiment F2, which is a yeast.

F5. The cell of embodiment F4, which is a Candida spp. yeast.

F6. The cell of embodiment F5, wherein the Candida spp. yeast is chosen from C. tropicalis and C. viswanathii.

F7. The cell of embodiment F6, wherein the Candida spp. yeast is a genetically modified ATCC20336 yeast.

F8. The cell of embodiment F4, which is chosen from a Yarrowia spp. yeast, Pichia spp. yeast, Saccharomyces spp. yeast and Kluyveromyces spp. yeast.

F9. The cell of embodiment F8, which is chosen from Y. lipolytica, P. pastoris, P. membranifaciens, P. kluyveri, P. guilliermondii, P. heedii, P. subpelliculosa, S. cerevisiae, S. bayanus, S. pastorianus, S. carlsbergensis, K. lactis and K. marxianus.

G1. A genetically modified yeast comprising (i) an active, modified endogenous acyl-coA oxidase polypeptide or an active, modified endogenous acyl-coA dehydrogenase polypeptide; (ii) a genetic modification that alters an enoyl co-A isomerase activity; and (iii) a genetic modification that alters a dieneoyl CoA reductase (DCR) activity, wherein the yeast is capable of producing a diacid from a feedstock comprising one or more components from a vegetable oil.

G2. The genetically modified yeast of embodiment Gl, wherein the feedstock comprises at least one long chain unsaturated fatty acid.

G3. The genetically modified yeast of embodiment G2, wherein the long chain unsaturated fatty acid is linoleic acid or linolenic acid.

G4. The genetically modified yeast of any one of embodiments G1 to G3, which comprises a genetic modification that reduces an enoyl CoA isomerase activity.

G5. The genetically modified yeast of embodiment G4, which comprises a disruption, deletion or knockout of (i) a polynucleotide that encodes an enoyl CoA isomerase polypeptide or (ii) a promoter operably linked to a polynucleotide that encodes an enoyl CoA isomerase polypeptide, whereby the enoyl CoA isomerase activity is reduced or removed.

G6. The genetically modified yeast of embodiment G5, wherein the genetic modification disrupts a polynucleotide that encodes an enoyl CoA isomerase polypeptide.

G7. The genetically modified yeast of any one of embodiments G1 to G6, which comprises a genetic modification that increases an enoyl CoA isomerase activity.

G8. The genetically modified yeast of embodiment G7, which comprises (i) an increased copy number of a polynucleotide that encodes an enoyl CoA isomerase polypeptide or (ii) a promoter inserted and operably linked to a polynucleotide that encodes an enoyl CoA isomerase polypeptide.

G9. The genetically modified yeast of any one of embodiments G5, G6 and G8, wherein the enoyl CoA isomerase polypeptide is a polypeptide native to the yeast.

G10. The genetically modified yeast of any one of embodiments G1 to G9, wherein the activity of one or more or all enoyl CoA isomerase polypeptide types in the yeast is altered.

G11. The genetically modified yeast of embodiment G10, wherein the yeast comprises two enoyl CoA isomerase polypeptide types, and one or both of the polypeptide types are altered.

G12. The genetically modified yeast of any one of embodiments G1 to G11, wherein the enoyl CoA isomerase activity is provided by polypeptide present in a Candida spp. yeast.

G13. The genetically modified yeast of any one of embodiments G1 to G12, wherein the enoyl CoA isomerase activity is provided by a polypeptide comprising the amino acid sequence of SEQ ID NO: 3675, 3677, 3800, 3802, 3803 or 3805.

G14. The genetically modified yeast of any one of embodiments G1 to G11, wherein the enoyl CoA isomerase activity is provided by a polypeptide present in a Yarrowia yeast.

G15. The genetically modified yeast of embodiment G14, wherein the Yarrowia yeast is a Y. lipolytica yeast.

G16. The genetically modified yeast of any one of embodiments G1 to G15, wherein the enoyl CoA isomerase activity is provided by a polypeptide comprising the amino acid sequence of SEQ ID NO: 3804 or 3806.

G17. The genetically modified yeast of any one of embodiments G1 to G16, which comprises a genetic modification that reduces a dieneoyl CoA reductase activity.

G18. The genetically modified yeast of embodiment G17, which comprises a disruption, deletion or knockout of (i) a polynucleotide that encodes a dieneoyl CoA reductase polypeptide or (ii) a promoter operably linked to a polynucleotide that encodes a dieneoyl CoA reductase polypeptide, whereby the dieneoyl CoA reductase activity is reduced or removed.

G19. The genetically modified yeast of embodiment G18, wherein the genetic modification disrupts a polynucleotide that encodes a dieneoyl CoA reductase polypeptide.

G20. The genetically modified yeast of any one of embodiments G1 to G19, which comprises a genetic modification that increases a dieneoyl CoA reductase activity.

G21. The genetically modified yeast of embodiment G20, which comprises (i) an increased copy number of a polynucleotide that encodes a dieneoyl CoA reductase polypeptide or (ii) a promoter inserted and operably linked to a polynucleotide that encodes a dieneoyl CoA reductase polypeptide.

G22. The genetically modified yeast of any one of embodiments G18, G19 and G21, wherein the dieneoyl CoA reductase polypeptide is a polypeptide native to the yeast.

G23. The genetically modified yeast of any one of embodiments G1 to G22, wherein the activity of one or more or all dieneoyl CoA reductase polypeptide types in the yeast is altered.

G24. The genetically modified yeast of embodiment G23, wherein the yeast comprises two dieneoyl CoA reductase polypeptide types, and one or both of the polypeptide types are altered.

G25. The genetically modified yeast of any one of embodiments G1 to G24, wherein the dieneoyl CoA reductase activity is provided by a polypeptide present is a Candida spp. yeast.

G26. The genetically modified yeast of any one of embodiments G1 to G25, wherein the dieneoyl CoA reductase activity is provided by a polypeptide comprising the amino acid sequence of SEQ ID NO: 3789, 3791, 3792 or 3793.

G27. The genetically modified yeast of any one of embodiments G1 to G26, wherein the dieneoyl CoA reductase activity is provided by a polypeptide present in a Yarrowia yeast.

G28. The genetically modified yeast of embodiment G27, wherein the Yarrowia yeast is a Y. lipolytica yeast.

G29. The genetically modified yeast of any one of embodiments G1 to G28, wherein the dieneoyl CoA reductase activity is provided by a polypeptide comprising the amino acid sequence of SEQ ID NO: 3794 or 3795.

G30. The genetically modified yeast of any one of embodiments G1 to G29, wherein the yeast is a genetically modified Candida spp. yeast.

G31. The genetically modified yeast of embodiment G30, wherein the Candida spp. yeast is chosen from C. tropicalis and C. viswanathii.

G32. The genetically modified yeast of embodiment G30, wherein the Candida spp. yeast is a genetically modified ATCC20336 yeast.

G33. The genetically modified yeast of any one of embodiments G30 to G32, wherein the endogenous acyl-CoA oxidase polypeptide is a POX4 polypeptide.

G34. The genetically modified yeast of embodiment G33, wherein the POX4 polypeptide comprises a modified amino acid sequence of SEQ ID NO: 30.

G35. The genetically modified yeast of embodiment G33 or G34, wherein the POX4 polypeptide comprises an amino acid modification at one or more amino acid positions chosen from 88, 90, 96, 98, 99, 100, 102, 103, 302, 309, 310, 473, 474, 475, 476, 477, 478, 479, 480, 481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 501, 502, 503, 504 and 505.

G36. The genetically modified yeast of any one of embodiments G30 to G35, wherein the endogenous acyl-CoA oxidase polypeptide is a POX5 polypeptide.

G37. The genetically modified yeast of embodiment G36, wherein the POX5 polypeptide comprises a modified amino acid sequence of SEQ ID NO: 32.

G38. The genetically modified yeast of embodiment G36 or G37, wherein the POX5 polypeptide comprises an amino acid modification at one or more amino acid positions chosen from 81, 82, 83, 84, 85, 86, 88, 93, 94, 95, 96, 98, 102, 284, 287, 290, 291, 292, 294, 295, 428, 429, 436, 453, 454, 455, 456, 457, 458, 459, 460, 461, 462 and 463.

G39. The genetically modified yeast of any one of embodiments G1 to G38, wherein the acyl-CoA dehydrogenase polypeptide is chosen from ACAD, VLCAD, LCAD, MCAD and SCAD polypeptides.

G40. The genetically modified yeast of embodiment G39, wherein the acyl-CoA dehydrogenase polypeptide comprises a modified amino acid sequence of SEQ ID NO: 3685.

G41 . The genetically modified yeast of embodiment G39 or G40, wherein the the acyl-CoA dehydrogenase polypeptide comprises an amino acid modification at VLCAD position 461.

G42. The genetically modified yeast of any one of embodiments G35, G38 and G41, wherein at least one of the amino acid modifications is an amino acid substitution.

G43. The genetically modified yeast of embodiment G42, wherein at least one of the one or more amino acid substitutions is conservative.

G44. The genetically modified yeast of embodiment G42, wherein at least one of the one or more amino acid substitutions is not conservative.

G45. The genetically modified yeast of any one of embodiments G1 to G29, wherein the yeast is a genetically modified Yarrowia spp. yeast.

G46. The genetically modified yeast of embodiment G42, wherein the Yarrowia spp. yeast is Y. lipolytica.

G47. The genetically modified yeast of embodiment G45 or G46, wherein the endogenous acyl-CoA oxidase polypeptide is chosen from a POX1 polypeptide, POX2 polypeptide, POX3 polypeptide, POX4 polypeptide, POX5 polypeptide or POX6 polypeptide.

G48. The genetically modified yeast of embodiment G47, wherein the endogenous acyl-CoA oxidase polypeptide is chosen from SEQ ID NOs: 3778 to 3783.

G49. The genetically modified yeast of any one of embodiments G1 to G29, wherein the yeast is a genetically modified Pichia spp. yeast.

G50. The genetically modified yeast of embodiment G49, wherein the Pichia spp. yeast is chosen from P. pastoris, P. membranifaciens, P. kluyveri, P. guilliermondii, P. heedii and P. subpelliculosa.

G51. The genetically modified yeast of embodiment G1 to G29, wherein the yeast is a genetically modified Saccharomyces spp. yeast.

G52. The genetically modified yeast of embodiment G51, wherein the Saccharomyces spp. yeast is chosen from S. cerevisiae, S. bayanus, S. pastorianus and S. carlsbergensis.

G53. The genetically modified yeast of embodiment G1 to G29, wherein the yeast is a genetically modified Kluyveromyces spp. yeast.

G54. The genetically modified yeast of embodiment G53, wherein the Kluyveromyces spp. yeast is chosen from K. lactis and K. marxianus.

G55. The genetically modified yeast of any one of embodiments G1 to G54, wherein the modified endogenous acyl-CoA oxidase polypeptide comprises an amino acid modification in the N-terminal loop.

G56. The genetically modified yeast of any one of embodiments G1 to G55, wherein the modified endogenous acyl-CoA oxidase polypeptide comprises an amino acid modification in the D alpha helix.

G57. The genetically modified yeast of any one of embodiments G1 to G56, wherein the modified endogenous acyl-CoA oxidase polypeptide comprises an amino acid modification in the loop between the D alpha helix and the E' alpha helix.

G58. The genetically modified yeast of any one of embodiments G1 to G57, wherein the modified endogenous acyl-CoA oxidase polypeptide comprises an amino acid modification to an amino acid in effective contact with carbons 6 to 9 in a feedstock component.

G59. The genetically modified yeast of any one of embodiments G1 to G58, wherein the modified endogenous acyl-CoA oxidase polypeptide comprises an amino acid modification to an amino acid in effective contact with carbons 10 to 12 in a feedstock component.

G60. The genetically modified yeast of any one of embodiments G1 to G59, wherein the modified endogenous acyl-CoA oxidase polypeptide comprises an amino acid modification in the L alpha helix.

G61 . The genetically modified yeast of any one of embodiments G1 to G60, wherein the modified endogenous acyl-CoA oxidase polypeptide comprises an amino acid modification in the loop C-terminal to the L alpha helix.

G62. The genetically modified yeast of any one of embodiments G1 to G61, wherein the modified endogenous acyl-CoA oxidase polypeptide comprises an amino acid modification in the loop between the L alpha helix and the M alpha helix.

G63. The genetically modified yeast of any one of embodiments G55 to G62, wherein the amino acid modification comprises an amino acid substitution.

G64. The genetically modified yeast of embodiment G63, wherein the amino acid substitution is conservative.

G65. The genetically modified yeast of embodiment G63, wherein the amino acid substitution is not conservative.

G66. A method for producing a diacid, comprising:

(a) contacting a genetically modified yeast of any one of embodiments G1 to G65 with a feedstock comprising a vegetable oil; and

(b) culturing the yeast under conditions in which diacid is produced from the feedstock.

G67. The method of embodiment G66, comprising purifying the diacid from the yeast.

H1. A method for producing a diacid, comprising:

(a) contacting a genetically modified yeast of any of embodiments A1 to A44, B1 to B33, C1 to C15 and G1 to G65 with a feedstock comprising an unsaturated vegetable fatty acid capable of being converted by the yeast to a diacid;

(b) culturing the yeast under conditions in which diacid containing one or more unsaturations is produced from the feedstock; and

(c) hydrogenating and removing one or more of the unsaturations.

H2. The method of embodiment H1, wherein the unsaturated vegeteable fatty acid is linoleic acid or linolenic acid.

H3. The method of embodiment H2, wherein the diacid is dodecanedioic acid or sebacic acid.

* * *

The entirety of each patent, patent application, publication and document referenced herein hereby is incorporated by reference. Citation of the above patents, patent applications, publications and documents is not an admission that any of the foregoing is pertinent prior art,

nor does it constitute any admission as to the contents or date of these publications or documents.

Modifications may be made to the foregoing without departing from the basic aspects of the technology. Although the technology has been described in substantial detail with reference to one or more specific embodiments, those of ordinary skill in the art will recognize that changes may be made to the embodiments specifically disclosed in this application, yet these modifications and improvements are within the scope and spirit of the technology.

The technology illustratively described herein suitably may be practiced in the absence of any element(s) not specifically disclosed herein. Thus, for example, in each instance herein any of the terms "comprising," "consisting essentially of," and "consisting of may be replaced with either of the other two terms. The terms and expressions which have been employed are used as terms of description and not of limitation, and use of such terms and expressions do not exclude any equivalents of the features shown and described or portions thereof, and various modifications are possible within the scope of the technology claimed. The term "a" or "an" can refer to one of or a plurality of the elements it modifies (e.g., "a reagent" can mean one or more reagents) unless it is contextually clear either one of the elements or more than one of the elements is described. The term "about" as used herein refers to a value within 10% of the underlying parameter (i.e., plus or minus 10%>), and use of the term "about" at the beginning of a string of values modifies each of the values (i.e., "about 1, 2 and 3" refers to about 1, about 2 and about 3). For example, a weight of "about 100 grams" can include weights between 90 grams and 110 grams. Further, when a listing of values is described herein (e.g., about 50%>, 60%), 70%), 80%), 85%o or 86%>) the listing includes all intermediate and fractional values thereof (e.g., 54%o, 85.4%o). Thus, it should be understood that although the present technology has been specifically disclosed by representative embodiments and optional features, modification and variation of the concepts herein disclosed may be resorted to by those skilled in the art, and such modifications and variations are considered within the scope of this technology.

Certain embodiments of the technology are set forth in the claim(s) that follow(s).

We claim:

1. A genetically modified Yarrowia lipoytica yeast comprising (i) an active, modified endogenous acyl-coA oxidase polypeptide or an active, modified endogenous acyl-coA dehydrogenase polypeptide; (ii) a genetic modification that alters a native enoyl co-A isomerase activity; and (iii) a genetic modification that alters a native dieneoyl CoA reductase (DCR) activity, wherein the yeast is capable of producing a diacid from a feedstock comprising one or more components from a vegetable oil.

2. A genetically modified yeast comprising an active, modified endogenous acyl-CoA oxidase polypeptide or an active, modified endogenous acyl-CoA dehydrogenase polypeptide, which yeast is capable of producing a diacid from a feedstock comprising one or more components from a vegetable oil.

3. The genetically modified yeast of claim 2, wherein the yeast is a genetically modified Candida spp. yeast.

4. The genetically modified yeast of claim 3, wherein the Candida spp. yeast is selected from the group consisting of: C. tropicalis and C. viswanathii.

5. The genetically modified yeast of claim , wherein the Candida spp. yeast is a genetically modified ATCC20336 yeast.

6. The genetically modified yeast of any one of claim 2-5, wherein the endogenous acyl-CoA oxidase polypeptide is a POX4 polypeptide.

7. The genetically modified yeast of claim 6, wherein the POX4 polypeptide comprises a modified amino acid sequence of SEQ ID NO: 30.

8. The genetically modified yeast of any one of claims 6-7, wherein the POX4 polypeptide comprises an amino acid modification at one or more amino acid positions selected from the group consisting of: 88, 90, 96, 98, 99, 100, 102, 103, 302, 309, 310, 473, 474, 475, 476, 477, 478, 479, 480, 481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 501, 502, 503, 504 and 505.

9. The genetically modified yeast of any one of claims 3-5, wherein the endogenous acyl-CoA oxidase polypeptide is a POX5 polypeptide.

10. The genetically modified yeast of claim 9, wherein the POX5 polypeptide comprises a modified amino acid sequence of SEQ ID NO: 32.

11. The genetically modified yeast of any one of claims 9-10, wherein the POX5 polypeptide comprises an amino acid modification at one or more amino acid positions selected from the group consisting of: 81, 82, 83, 84, 85, 86, 88, 93, 94, 95, 96, 98, 102, 284, 287, 290, 291, 292, 294, 295, 428, 429, 436, 453, 454, 455, 456, 457, 458, 459, 460, 461, 462 and 463.

12. The genetically modified yeast of any one of claims 3-5, wherein the acyl-CoA dehydrogenase polypeptide is selected from the group consisting of: ACAD, VLCAD, LCAD, MCAD and SCAD polypeptides.

13. The genetically modified yeast of claim 12, wherein the acyl-CoA dehydrogenase polypeptide comprises a modified amino acid sequence of SEQ ID NO: 3685.

14. The genetically modified yeast of any one of claims 12-13, wherein the acyl-CoA dehydrogenase polypeptide comprises an amino acid modification at VLCAD position 461.

15. The genetically modified yeast of any one of claims 8, 11 and 14, wherein at least one of the amino acid modifications is an amino acid substitution.

16. The genetically modified yeast of claim 15, wherein at least one of the one or more amino acid substitutions is conservative.

17. The genetically modified yeast of claim 15, wherein at least one of the one or more amino acid substitutions is not conservative.

18. The genetically modified yeast of claim 2, wherein the yeast is a genetically modified Yarrowia spp. yeast.

19. The genetically modified yeast of claim 18, wherein the Yarrowia spp. yeast is Y. lipolytica.

20. The genetically modified yeast of any one of claims 18-19, wherein the endogenous acyl-CoA oxidase polypeptide is selected from the group consisting of: a POX1 polypeptide, POX2 polypeptide, POX3 polypeptide, POX4 polypeptide, POX5 polypeptide or POX6 polypeptide.

21. The genetically modified yeast of claim 20, wherein the endogenous acyl-CoA oxidase polypeptide is selected from the group consisting of: SEQ ID NOs: 3778 to 3783.

22. The genetically modified yeast of claim 2, wherein the yeast is a genetically modified Pichia spp. yeast.

23. The genetically modified yeast of claim 22, wherein the Pichia spp. yeast is selected from the group consisting of: P. pastoris, P. membranifaciens, P. kluyveri, P. guilliermondii, P. heedii and P. subpelliculosa.

24. The genetically modified yeast of claim 2, wherein the yeast is a genetically modified Saccharomyces spp. yeast.

25. The genetically modified yeast of claim 24, wherein the Saccharomyces spp. yeast is selected from the group consisting of: S. cerevisiae, S. bayanus, S. pastorianus and S. carlsbergensis.

26. The genetically modified yeast of claim 2, wherein the yeast is a genetically modified Kluyveromyces spp. yeast.

27. The genetically modified yeast of claim 26, wherein the Kluyveromyces spp. yeast is selected from the group consisting of: K. lactis and K. marxianus.

28. The genetically modified yeast of any one of claims 2-27, wherein the modified endogenous acyl-CoA oxidase polypeptide comprises an amino acid modification in the N-terminal loop.

29. The genetically modified yeast of any one of claims 2-28, wherein the modified endogenous acyl-CoA oxidase polypeptide comprises an amino acid modification in the D alpha helix.

30. The genetically modified yeast of any one of claims 2-29, wherein the modified endogenous acyl-CoA oxidase polypeptide comprises an amino acid modification in the loop between the D alpha helix and the E' alpha helix.

31. The genetically modified yeast of any one of claims 2-30, wherein the modified endogenous acyl-CoA oxidase polypeptide comprises an amino acid modification to an amino acid in effective contact with carbons 6 to 9 in a feedstock component.

32. The genetically modified yeast of any one of claims 2-31, wherein the modified endogenous acyl-CoA oxidase polypeptide comprises an amino acid modification to an amino acid in effective contact with carbons 10 to 12 in a feedstock component.

33. The genetically modified yeast of any one of claims 2-32, wherein the modified endogenous acyl-CoA oxidase polypeptide comprises an amino acid modification in the L alpha helix.

34. The genetically modified yeast of any one of claims 2-33, wherein the modified endogenous acyl-CoA oxidase polypeptide comprises an amino acid modification in the loop C-terminal to the L alpha helix.

35. The genetically modified yeast of any one of claims 2-34, wherein the modified endogenous acyl-CoA oxidase polypeptide comprises an amino acid modification in the loop between the L alpha helix and the M alpha helix.

36. The genetically modified yeast of any one of claims 28-35, wherein the amino acid modification comprises an amino acid substitution.

37. The genetically modified yeast of claim 36, wherein the amino acid substitution is conservative.

38. The genetically modified yeast of claim 36, wherein the amino acid substitution is not conservative.

39. A genetically modified yeast comprising (i) an active, modified endogenous acyl-CoA oxidase polypeptide or an active, modified endogenous acyl-CoA dehydrogenase polypeptide and (ii) a genetic modification that alters an enoyl CoA isomerase activity, wherein the yeast is capable of producing a diacid from a feedstock comprising one or more components from a vegetable oil.

40. The genetically modified yeast of any one of claims 2-39, which comprises a genetic modification that reduces an enoyl CoA isomerase activity.

41. The genetically modified yeast of claim 40, which comprises a disruption, deletion or knockout of (i) a polynucleotide that encodes an enoyl CoA isomerase polypeptide or (ii) a promoter operably linked to a polynucleotide that encodes an enoyl CoA isomerase polypeptide, whereby the enoyl CoA isomerase activity is reduced or removed.

42. The genetically modified yeast of claim 41, wherein the genetic modification disrupts a polynucleotide that encodes an enoyl CoA isomerase polypeptide.

43. The genetically modified yeast of any one of claims 2-42, which comprises a genetic modification that increases an enoyl CoA isomerase activity.

44. The genetically modified yeast of claim 43, which comprises (i) an increased copy number of a polynucleotide that encodes an enoyl CoA isomerase polypeptide or (ii) a promoter inserted and operably linked to a polynucleotide that encodes an enoyl CoA isomerase polypeptide.

45. The genetically modified yeast of any one of claims 40-44, wherein the enoyl CoA isomerase polypeptide is a polypeptide native to the yeast.

46. The genetically modified yeast of any one of claims 39-45, wherein the activity of one or more or all enoyl CoA isomerase polypeptide types in the yeast is altered.

47. The genetically modified yeast of claim 46, wherein the yeast comprises two enoyl CoA isomerase polypeptide types, and one or both of the polypeptide types are altered.

48. The genetically modified yeast of any one of claims 39-45, wherein the enoyl CoA isomerase activity is provided by a polypeptide present in a Candida spp. yeast.

49. The genetically modified yeast of any one of claims 39-48, wherein the enoyl CoA isomerase activity is provided by a polypeptide comprising the amino acid sequence of SEQ ID NO: 3675, 3677, 3800, 3802, 3803 or 3805.

50. The genetically modified yeast of any one of claims 39-48, wherein the yeast is a Yarrowia yeast.

51. The genetically modified yeast of claim 50, wherein the yeast is a Y. lipolytica yeast.

52. The genetically modified yeast of claim 51, wherein the enoyl CoA isomerase activity is provided by a polypeptide present in a Yarrowia yeast or a Y. lipolytica yeast.

53. The genetically modified yeast of claim 51, wherein the enoyl CoA isomerase activity is provided by a polypeptide comprising the amino acid sequence of SEQ ID NO: 3804 or 3806.

54. The genetically modified yeast of any one of claims 2-53, which comprises a genetic modification that reduces the cytoplasmic activity of an acyl-CoA synthetase (ACS) polypeptide.

55. The genetically modified yeast of any one of claims 2-54, which comprises a genetic modification that reduces the peroxisomal activity of an acyl-CoA synthetase (ACS) polypeptide.

56. The genetically modified yeast of claim 54 or 55, wherein the genetic modification disrupts a polynucleotide that encodes the acyl-CoA synthetase (ACS) polypeptide.

57. The genetically modified yeast of claim 56, wherein the genetic modification disrupts an ACS1 polypeptide or ACS2 polypeptide.

58. The genetically modified yeast of any one of claims 54-57, wherein the genetic modification disrupts a polynucleotide that encodes a long-chain acyl-CoA synthetase polypeptide.

59. The genetically modified yeast of claim 58, wherein the genetic modification disrupts a FATl polypeptide.

60. The genetically modified yeast of any one of claims 54-59, wherein the acyl-CoA synthetase (ACS) polypeptide is a polypeptide native to the yeast.

61. The genetically modified yeast of claim 60, wherein the yeast is a Candida spp. yeast.

62. The genetically modified yeast of claim 61, wherein the acyl-CoA synthetase (ACS) polypeptide comprises an amino acid sequence selected from the group consisting of: SEQ ID NOs: 80, 82, 84, 158 and 159.

63. The genetically modified yeast of claim 61, wherein the FATl polypeptide comprises the amino acid sequence of SEQ ID NO: 90.

64. The genetically modified yeast of any one of claims 2-63, which comprises a genetic modification that reduces the activity of a PXA polypeptide.

65. The genetically modified yeast of claim 64, wherein the genetic modification disrupts a polynucleotide that encodes the PXA polypeptide.

66. The genetically modified yeast of claim 64 or 65, wherein the PXA polypeptide is a PXAl polypeptide or a PXA2 polypeptide, or a PXAl polypeptide and a PXA2 polypeptide.

67. The genetically modified yeast of any one of claims 64-66, wherein the PXA polypeptide is native to the yeast.

68. The genetically modified yeast of claim 67, wherein the yeast is a Candida spp. yeast.

69. The genetically modified yeast of claim 68, wherein the PXA1 polypeptide comprises the amino acid sequence of SEQ ID NO: 92.

70. The genetically modified yeast of claim 68, wherein the PXA2 polypeptide comprises the amino acid sequence of SEQ ID NO: 94.

71. The genetically modified yeast of any one of claims 2-70, comprising an active, modified endogenous acyl-CoA oxidase polypeptide and no active, modified endogenous acyl-CoA dehydrogenase polypeptide.

72. The genetically modified yeast of any one of claims 2-70, comprising no active, modified endogenous acyl-CoA oxidase polypeptide and an active, modified endogenous acyl-CoA dehydrogenase polypeptide.

73. A method for producing a diacid, comprising:

(1) contacting a genetically modified yeast of any of claims 1-72 with a feedstock comprising one or more components from a feedstock containing at least one unsaturated vegetable fatty acid capable of being converted by the yeast to a diacid;

(2) culturing the yeast under conditions in which the diacid is produced from the feedstock; and

(3) hydrogenating the diacid to remove any unsaturation in the carbon chain.

74. The method of claim 73, wherein the unsaturated vegeteable fatty acid is linoleic acid or linolenic acid.

75. The method of claim 74, wherein the diacid is dodecanedioic acid or sebacic acid.

76. A genetically modified yeast, comprising a heterologous acyl-CoA oxidase polypeptide or a heterologous acyl-CoA dehydrogenase polypeptide, which yeast is capable of producing a diacid from a feedstock comprising one or more components from a vegetable oil.

77. The genetically modified yeast of claim 76, wherein the heterologous acyl-CoA oxidase polypeptide is a native polypeptide.

78. The genetically modified yeast of claim 76, wherein the heterologous acyl-CoA oxidase polypeptide is an active, modified polypeptide.

79. The genetically modified yeast of claim 76, wherein the heterologous acyl-CoA dehydrogenase polypeptide is a native polypeptide.

80. The genetically modified yeast of claim 76, wherein the heterologous acyl-CoA dehydrogenase polypeptide is an active, modified polypeptide.

81. The genetically modified yeast of claim 76 or 77, wherein the heterologous acyl-CoA oxidase polypeptide is selected from the group consisting of: a polypeptide having an amino acid sequence set forth in SEQ ID NO: 51 to SEQ ID NO: 3673.

82. The genetically modified yeast of claim 76 or 79, wherein the heterologous acyl-CoA dehydrogenase polypeptide is selected from the group consisting of: SEQ ID NOs: 3679 to 3683, 3686, 3689, 3691, 3693, 3695, 3697, 3699, 3701 and 3703.

83. The genetically modified yeast of any one of claims 76-82, which is selected from the group consisting of: a Candida spp. yeast, Yarrowia spp. yeast, Pichia spp. yeast, Saccharomyces spp. yeast and Kluyveromyces spp. yeast.

84. The genetically modified yeast of claim 83, which is selected from the group consisting of: C. tropicalis, C. viswanathii, Y. lipolytica, P. pastoris, P. membranifaciens, P. kluyveri, P. guilliermondii, P. heedii, P. subpelliculosa, S. cerevisiae, S. bayanus, S. pastorianus, S. carlsbergensis, K. lactis and K. marxianus.

85. The genetically modified yeast of any one of claims 76-84, which comprises a genetic modification that reduces the activity of an enoyl CoA isomerase polypeptide.

86. The genetically modified yeast of claim 85, wherein the genetic modification disrupts a polynucleotide that encodes the enoyl CoA isomerase polypeptide.

87. The genetically modified yeast of claim 85 or 86, wherein the enoyl CoA isomerase polypeptide is a polypeptide native to the yeast.

88. The genetically modified yeast of claim 87, wherein the yeast is a Candida spp. yeast.

89. The genetically modified yeast of claim 88, wherein the enoyl CoA isomerase polypeptide comprises the amino acid sequence of SEQ ID NO: 3675 or 3677.

90. The genetically modified yeast of any one of claims 76-89, which comprises a genetic modification that reduces the cytoplasmic activity of an acyl-CoA synthetase (ACS) polypeptide.

91. The genetically modified yeast of any one of claims 76-90, which comprises a genetic modification that reduces the peroxisomal activity of an acyl-CoA synthetase (ACS) polypeptide.

92. The genetically modified yeast of claim 90 or 91, wherein the genetic modification disrupts a polynucleotide that encodes the acyl-CoA synthetase (ACS) polypeptide.

93. The genetically modified yeast of claim 92, wherein the genetic modification disrupts an ACS1 polypeptide or ACS2 polypeptide.

94. The genetically modified yeast of any one of claims 90-93, wherein the genetic modification disrupts a polynucleotide that encodes a long-chain acyl-CoA synthetase polypeptide.

95. The genetically modified yeast of claim 94, wherein the genetic modification disrupts a FAT1 polypeptide.

96. The genetically modified yeast of any one of claims 90-95, wherein the acyl-CoA synthetase (ACS) polypeptide is a polypeptide native to the yeast.

97. The genetically modified yeast of claim 96, wherein the yeast is a Candida spp. yeast.

98. The genetically modified yeast of claim 97, wherein the acyl-CoA synthetase (ACS) polypeptide comprises an amino acid sequence selected from the group consisting of: SEQ ID NOs: 80, 82, 84, 158 and 159.

99. The genetically modified yeast of claim 96, wherein the FAT1 polypeptide comprises the amino acid sequence of SEQ ID NO: 90.

100. The genetically modified yeast of any one of claims 76-99, which comprises a genetic modification that reduces the activity of a PXA polypeptide.

101. The genetically modified yeast of claim 100, wherein the genetic modification disrupts a polynucleotide that encodes the PXA polypeptide.

102. The genetically modified yeast of claim 100 or 101, wherein the PXA polypeptide is a PXAl polypeptide or a PXA2 polypeptide, or a PXAl polypeptide and a PXA2 polypeptide.

103. The genetically modified yeast of any one of claims 100-102, wherein the PXA polypeptide is native to the yeast.

104. The genetically modified yeast of claim 103, wherein the yeast is a Candida spp. yeast.

105. The genetically modified yeast of claim 104, wherein the PXAl polypeptide comprises the amino acid sequence of SEQ ID NO: 92.

106. The genetically modified yeast of claim 104, wherein the PXA2 polypeptide comprises the amino acid sequence of SEQ ID NO: 94.

107. The genetically modified yeast of any one of claims 76-106, comprising an active, modified endogenous acyl-CoA oxidase polypeptide and no active, modified endogenous acyl-CoA dehydrogenase polypeptide.

108. The genetically modified yeast of any one of claims 76-106, comprising no active, modified endogenous acyl-CoA oxidase polypeptide and an active, modified endogenous acyl-CoA dehydrogenase polypeptide.

109. The genetically modified yeast of any one of claims 1-108, comprising one or more genetic modifications that reduce the activity of one or more native endogenous acyl-CoA oxidase polypeptides.

110. The genetically modified yeast of claim 109, comprising genetic modifications that reduce the activity of all native endogenous acyl-CoA oxidase polypeptides.

111. The genetically modified yeast of claim 109 or 110, wherein the genetic modifications partially block beta oxidation activity.

112. The genetically modified yeast of any one of claims 1-1 11, wherein the diacid is a C4 to C24 diacid.

113. The genetically modified yeast of claim 112, wherein the diacid is a CIO, CI2, CI4, CI6, CI8 or C20 diacid.

114. The genetically modified yeast of claim 113, wherein the diacid is a CIO diacid.

115. The genetically modified yeast of claim 113, wherein the diacid is a C12 diacid.

116. The genetically modified yeast of claim 113, wherein the diacid is a C18 diacid.

117. The genetically modified yeast of any one of claims 112-1 16, wherein the diacid contains no unsaturation.

118. The genetically modified yeast of any one of claims 112-1 16, wherein the diacid contains one or more unsaturations.

119. The genetically modified yeast of any one of claims 112-1 18, wherein the diacid is the predominant diacid in a mixture of diacids.

120. The genetically modified yeast of any one of claims 1-1 19, wherein the feedstock comprises a substantially pure oil.

121 . The genetically modified yeast of any one of claims 1-120, wherein the feedstock comprises a plurality of fatty acids.

122. The genetically modified yeast of claim 121, wherein the feedstock comprises a soapstock.

123. The genetically modified yeast of claim 121, wherein the feedstock comprises a fatty acid distillate.

124. The genetically modified yeast of any one of claims 1-123, wherein the vegetable oil is from a plant selected from the group consisting of: palm, palm kernel, coconut, soy, safflower, canola, palm, palm kernel or combination thereof.

125. A method for producing a desired diacid, comprising:

contacting a genetically modified yeast of any one of claims 1-124 with a feedstock comprising one or more components from a vegetable oil capable of being converted by the yeast to a diacid; and

culturing the yeast under conditions in which the desired diacid is produced from the feedstock.

126. The method of claim 125, wherein the desired diacid is a C4 to C24 diacid.

127. The method of any one of claims 125-126, wherein the desired diacid contains no unsaturation.

128. The method of any one of claims 125-126, wherein the desired diacid contains one or more unsaturations.

129. The method of any one of claims 125-128, wherein the desired diacid is the predominant diacid in a mixture of diacids.

130. The method of any one of claims 125-129, wherein the feedstock comprises a substantially pure oil.

131. The method of any one of claims 125-129, wherein the feedstock comprises a plurality of fatty acids.

132. The method of claim 131, wherein the feedstock comprises a soapstock.

133. The method of claim 131, wherein the feedstock comprises a fatty acid distillate.

134. The method of any one of claims 125-133, wherein the vegetable oil is from a plant selected from the group consisting of: palm, palm kernel, coconut, soy, safflower, canola, palm, palm kernel or combination thereof.

135. An isolated nucleic acid, comprising a polynucleotide that encodes a modified acyl-CoA oxidase polypeptide from a yeast.

136. The isolated nucleic acid of claim 135, wherein the modified acyl-CoA oxidase polypeptide comprises an amino acid modification in the N-terminal loop.

137. The isolated nucleic acid of claim 135, wherein the modified endogenous acyl-CoA oxidase polypeptide comprises an amino acid modification in the D alpha helix.

138. The isolated nucleic acid of claim 135, wherein the modified endogenous acyl-CoA oxidase polypeptide comprises an amino acid modification in the loop between the D alpha helix and the E' alpha helix.

139. The isolated nucleic acid of claim 135, wherein the modified endogenous acyl-CoA oxidase polypeptide comprises an amino acid modification to an amino acid in effective contact with carbons 6 to 9 in a feedstock component.

140. The isolated nucleic acid of claim 135, wherein the modified endogenous acyl-CoA oxidase polypeptide comprises an amino acid modification to an amino acid in effective contact with carbons 10 to 12 in a feedstock component.

141. The isolated nucleic acid of claim 135, wherein the modified endogenous acyl-CoA oxidase polypeptide comprises an amino acid modification in the L alpha helix.

142. The isolated nucleic acid of claim 135, wherein the modified endogenous acyl-CoA oxidase polypeptide comprises an amino acid modification in the loop C-terminal to the L alpha helix.

143. The isolated nucleic acid of claim 135, wherein the modified endogenous acyl-CoA oxidase polypeptide comprises an amino acid modification in the loop between the L alpha helix and the M alpha helix.

144. The isolated nucleic acid of any one of claims 135-143, wherein the yeast is selected from the group consisting of: a Candida spp. yeast, Yarrowia spp. yeast, Pichia spp. yeast, Saccharomyces spp. yeast and Kluyveromyces spp. yeast.

145. The isolated nucleic acid of claim 144, wherein the yeast is selected from the group consisting of: C. tropicalis, C. viswanathii, Y. lipolytica, P. pastoris, P. membranifaciens, P. kluyveri, P. guilliermondii, P. heedii, P. subpelliculosa, S. cerevisiae, S. bayanus, S. pastorianus, S. carlsbergensis, K. lactis and K. marxianus.

146. The isolated nucleic acid of claim 144, wherein the yeast is a Candida spp. Yeast selected from the group consisting of: C. tropicalis and C. viswanathii.

147. The isolated nucleic acid of claim 146, wherein the yeast is a genetically modified ATCC20336 yeast.

148. The isolated nucleic acid of any one of claims 144-147, wherein the endogenous acyl-CoA oxidase polypeptide is a POX4 polypeptide.

149. The isolated nucleic acid of claim 148, wherein the POX4 polypeptide comprises a modified amino acid sequence of SEQ ID NO: 30.

150. The isolated nucleic acid of claim 148 or 149, wherein the POX4 polypeptide comprises amino acid modifications at one or more amino acid positions selected from the group consisting of: 88, 90, 96, 98, 99, 100, 102, 103, 302, 309, 310, 473, 474, 475, 476, 477, 478, 479, 480, 481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 501, 502, 503, 504 and 505.

151. The isolated nucleic acid of any one of claims 144-147, wherein the endogenous acyl-CoA oxidase polypeptide is a POX5 polypeptide.

152. The isolated nucleic acid of claim 141, wherein the POX5 polypeptide comprises a modified amino acid sequence of SEQ ID NO: 32.

153. The isolated nucleic acid of claim 151 or 152, wherein the POX5 polypeptide comprises amino acid modifications at one or more amino acid positions selected from the group consisting of: 81, 82, 83, 84, 85, 86, 88, 93, 94, 95, 96, 98, 102, 284, 287, 290, 291, 292, 294, 295, 436, 453, 454, 455, 456, 457, 458, 459, 460, 461, 462 and 463.

154. An expression vector comprising the isolated nucleic acid of any one of claims 135-153.

155. A host cell comprising a nucleic acid of any one of claims 135-153.

156. A host cell of claim 155, which is a Candida spp. yeast cell selected from the group consisting of: C. tropicalis and C. viswanathii.

157. The host cell of claim 156, wherein the Candida spp. yeast is a genetically modified ATCC20336 yeast.

158. The host cell of claim 155, which is selected from the group consisting of: a Yarrowia spp. yeast, Pichia spp. yeast, Saccharomyces spp. yeast and Kluyveromyces spp. yeast.

159. The host cell of claim 158, which is selected from the group consisting of: Y. lipolytica, P. pastoris, P. membranifaciens, P. kluyveri, P. guilliermondii, P. heedii, P. subpelliculosa, S. cerevisiae, S. bayanus, S. pastorianus, S. carlsbergensis, K. lactis and K. marxianus.

160. A genetically modified yeast comprising (i) an active, modified endogenous acyl-coA oxidase polypeptide or an active, modified endogenous acyl-coA dehydrogenase polypeptide; (ii) a genetic modification that alters an enoyl co-A isomerase activity; and (iii) a genetic modification that alters a dieneoyl CoA reductase (DCR) activity, wherein the yeast is capable of producing a diacid from a feedstock comprising one or more components from a vegetable oil.

161. The genetically modified yeast of claim 160, wherein the feedstock comprises at least one long chain unsaturated fatty acid.

162. The genetically modified yeast of claim 161, wherein the long chain unsaturated fatty acid is linoleic acid or linolenic acid.

163. The genetically modified yeast of any one of claims 160-162, which comprises a genetic modification that reduces an enoyl CoA isomerase activity.

164. The genetically modified yeast of claim 163, which comprises a disruption, deletion or knockout of (i) a polynucleotide that encodes an enoyl CoA isomerase polypeptide or (ii) a promoter operably linked to a polynucleotide that encodes an enoyl CoA isomerase polypeptide, whereby the enoyl CoA isomerase activity is reduced or removed.

165. The genetically modified yeast of claim 164, wherein the genetic modification disrupts a polynucleotide that encodes an enoyl CoA isomerase polypeptide.

166. The genetically modified yeast of any one of claims 160-165, which comprises a genetic modification that increases an enoyl CoA isomerase activity.

167. The genetically modified yeast of claim 166, which comprises (i) an increased copy number of a polynucleotide that encodes an enoyl CoA isomerase polypeptide or (ii) a promoter

inserted and operably linked to a polynucleotide that encodes an enoyl CoA isomerase polypeptide.

168. The genetically modified yeast of any one of claims 160-167, wherein the enoyl CoA isomerase activity is provided by a polypeptide comprising the amino acid sequence of SEQ ID NO: 3675, 3677, 3800, 3802, 3803 or 3805.

169. The genetically modified yeast of any one of claims 160-167, wherein enoyl CoA isomerase activity is provided by a polypeptide present in a Yarrowia yeast.

170. The genetically modified yeast of claim 169, wherein the Yarrowia yeast is a Y. lipolytica yeast.

171. The genetically modified yeast of any one of claims 160-167, wherein enoyl CoA isomerase activity is provided by a polypeptide comprising the amino acid sequence of SEQ ID NO: 3804 or 3806.

172. The genetically modified yeast of any one of claims 160-171, which comprises a genetic modification that reduces a dieneoyl CoA reductase activity.

173. The genetically modified yeast of claim 172, which comprises a disruption, deletion or knockout of (i) a polynucleotide that encodes a dieneoyl CoA reductase polypeptide or (ii) a promoter operably linked to a polynucleotide that encodes a dieneoyl CoA reductase polypeptide, whereby the dieneoyl CoA reductase activity is reduced or removed.

174. The genetically modified yeast of claim 173, wherein the genetic modification disrupts a polynucleotide that encodes a dieneoyl CoA reductase polypeptide.

175. The genetically modified yeast of any one of claims 160-171, which comprises a genetic modification that increases a dieneoyl CoA reductase activity.

176. The genetically modified yeast of claim 175, which comprises (i) an increased copy number of a polynucleotide that encodes a dieneoyl CoA reductase polypeptide or (ii) a

promoter inserted and operably linked to a polynucleotide that encodes a dieneoyl CoA reductase polypeptide.

177. The genetically modified yeast of any one of claims 160-176, wherein the dieneoyl CoA reductase activity is provided by a polypeptide comprising the amino acid sequence of SEQ ID NO: 3789, 3791, 3792 or 3793.

178. The genetically modified yeast of any one of claims 160-177, wherein the dieneoyl CoA reductase activity is provided by a polypeptide present in a Yarrowia yeast.

179. The genetically modified yeast of claim 178, wherein the Yarrowia yeast is a Y. lipolytica yeast.

180. The genetically modified yeast of any one of claims 160-177, wherein the dieneoyl CoA reductase activity is provided by a polypeptide comprising the amino acid sequence of SEQ ID NO: 3794 or 3795.

181. The genetically modified yeast of any one of claims 160-180, wherein the endogenous acyl-CoA oxidase polypeptide is a POX4 polypeptide.

182. The genetically modified yeast of claim 181, wherein the POX4 polypeptide comprises a modified amino acid sequence of SEQ ID NO: 30.

183. The genetically modified yeast of claim 181 or 182, wherein the POX4 polypeptide comprises an amino acid modification at one or more amino acid positions selected from the group consisting of: 88, 90, 96, 98, 99, 100, 102, 103, 302, 309, 310, 473, 474, 475, 476, 477, 478, 479, 480, 481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 501, 502, 503, 504 and 505.

184. The genetically modified yeast of any one of claims 160-180, wherein the endogenous acyl-CoA oxidase polypeptide is a POX5 polypeptide.

185. The genetically modified yeast of claim 184, wherein the POX5 polypeptide comprises a modified amino acid sequence of SEQ ID NO: 32.

186. The genetically modified yeast of claim 184, wherein the POX5 polypeptide comprises an amino acid modification at one or more amino acid positions selected from the group consisting of: 81, 82, 83, 84, 85, 86, 88, 93, 94, 95, 96, 98, 102, 284, 287, 290, 291, 292, 294, 295, 428, 429, 436, 453, 454, 455, 456, 457, 458, 459, 460, 461, 462 and 463.

187. The genetically modified yeast of any one of claims 160-186, wherein the acyl-CoA dehydrogenase polypeptide is selected from the group consisting of: ACAD, VLCAD, LCAD, MCAD and SCAD polypeptides.

188. The genetically modified yeast of claim 187, wherein the acyl-CoA dehydrogenase polypeptide comprises a modified amino acid sequence of SEQ ID NO: 3685.

189. The genetically modified yeast of any one of claims 160-188, wherein the modified endogenous acyl-CoA oxidase polypeptide comprises an amino acid modification in the N-terminal loop.

190. The genetically modified yeast of any one of claims 160-189, wherein the modified endogenous acyl-CoA oxidase polypeptide comprises an amino acid modification in the D alpha helix.

191. The genetically modified yeast of any one of claims 160-190, wherein the modified endogenous acyl-CoA oxidase polypeptide comprises an amino acid modification in the loop between the D alpha helix and the E' alpha helix.

192. The genetically modified yeast of any one of claims 160-191, wherein the modified endogenous acyl-CoA oxidase polypeptide comprises an amino acid modification to an amino acid in effective contact with carbons 6 to 9 in a feedstock component.

193. The genetically modified yeast of any one of claims 160-192, wherein the modified endogenous acyl-CoA oxidase polypeptide comprises an amino acid modification to an amino acid in effective contact with carbons 10 to 12 in a feedstock component.

194. The genetically modified yeast of any one of claims 160-193, wherein the modified endogenous acyl-CoA oxidase polypeptide comprises an amino acid modification in the L alpha helix.

195. The genetically modified yeast of any one of claims 160-194, wherein the modified endogenous acyl-CoA oxidase polypeptide comprises an amino acid modification in the loop C-terminal to the L alpha helix.

196. The genetically modified yeast of any one of claims 160-195, wherein the modified endogenous acyl-CoA oxidase polypeptide comprises an amino acid modification in the loop between the L alpha helix and the M alpha helix.

197. A method for producing a diacid, comprising:

(a) contacting a genetically modified yeast of any one of claims 160-196 with a feedstock comprising a vegetable oil; and

(b) culturing the yeast under conditions in which diacid is produced from the feedstock.

198. The method of claim 197, comprising purifying the diacid from the yeast.

199. A method for producing a diacid, comprising:

(a) contacting a genetically modified yeast of any of embodiments 1-124 and 160-198 with a feedstock comprising an unsaturated vegetable fatty acid capable of being converted by the yeast to a diacid;

(b) culturing the yeast under conditions in which diacid containing one or more unsaturations is produced from the feedstock; and

(c) hydrogenating and removing one or more of the unsaturations.

200. The method of claim 199, wherein the unsaturated vegeteable fatty acid is linoleic acid or linolenic acid.

201 . A diacid obtained by the method of any one of claims 199-200.

202. The diacid of claim 201, wherein said diacid is is dodecanedioic acid or sebacic acid.

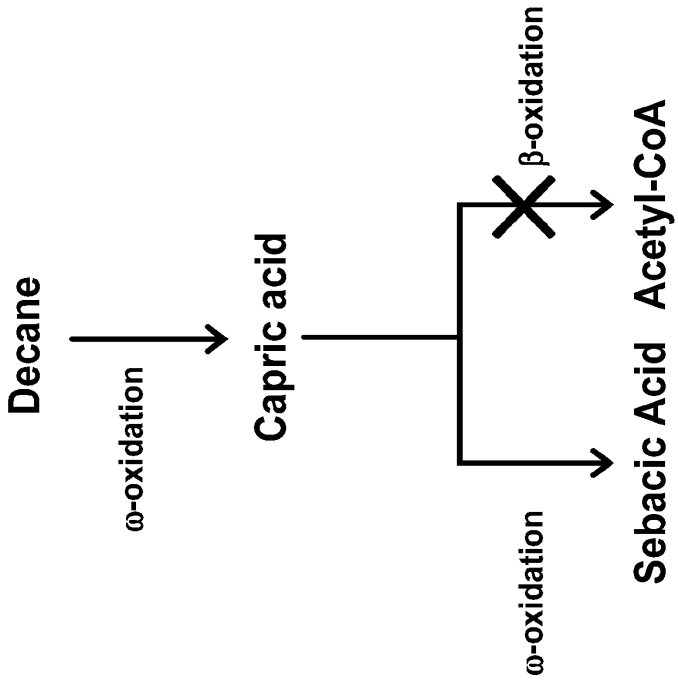

Decane
ω-oxidation
Capric acid
β-oxidation
Acetyl-CoA
ω-oxidation
Sebacic Acid

FIG. 1

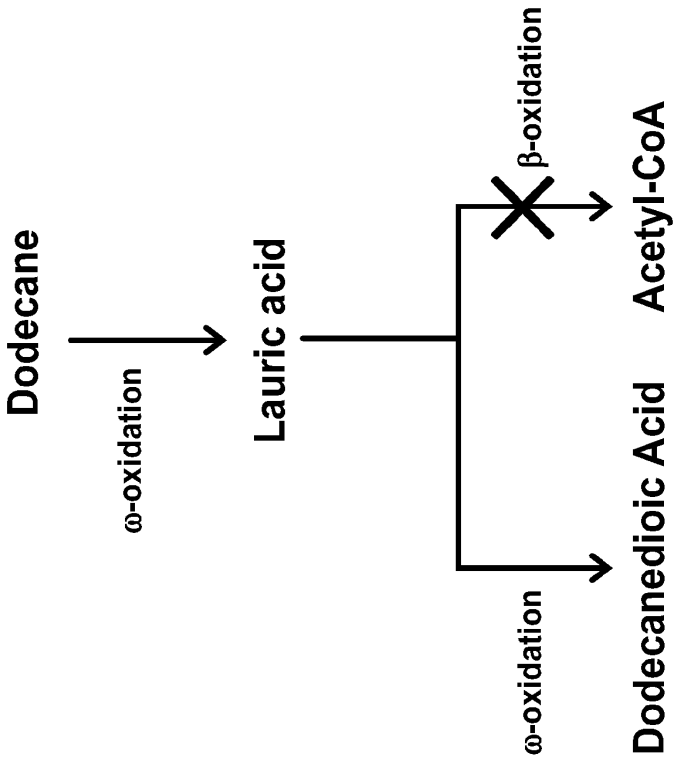

Dodecane
ω-oxidation
Lauric acid
β-oxidation
Acetyl-CoA
ω-oxidation
Dodecanedioic Acid

FIG. 2

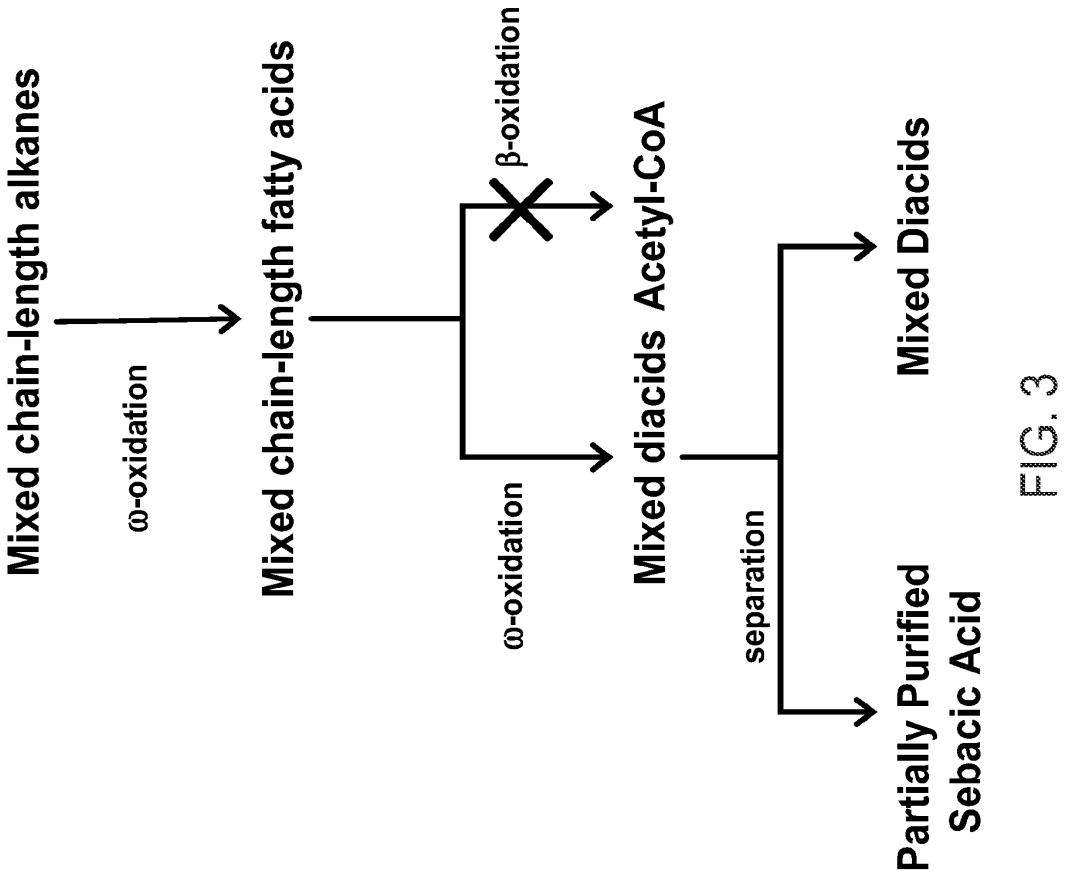

Mixed chain-length alkanes
ω-oxidation
Mixed chain-length fatty acids
β-oxidation
ω-oxidation
Mixed diacids Acetyl-CoA
separation
Mixed Diacids
Partially Purified
Sebacic Acid

FIG. 3

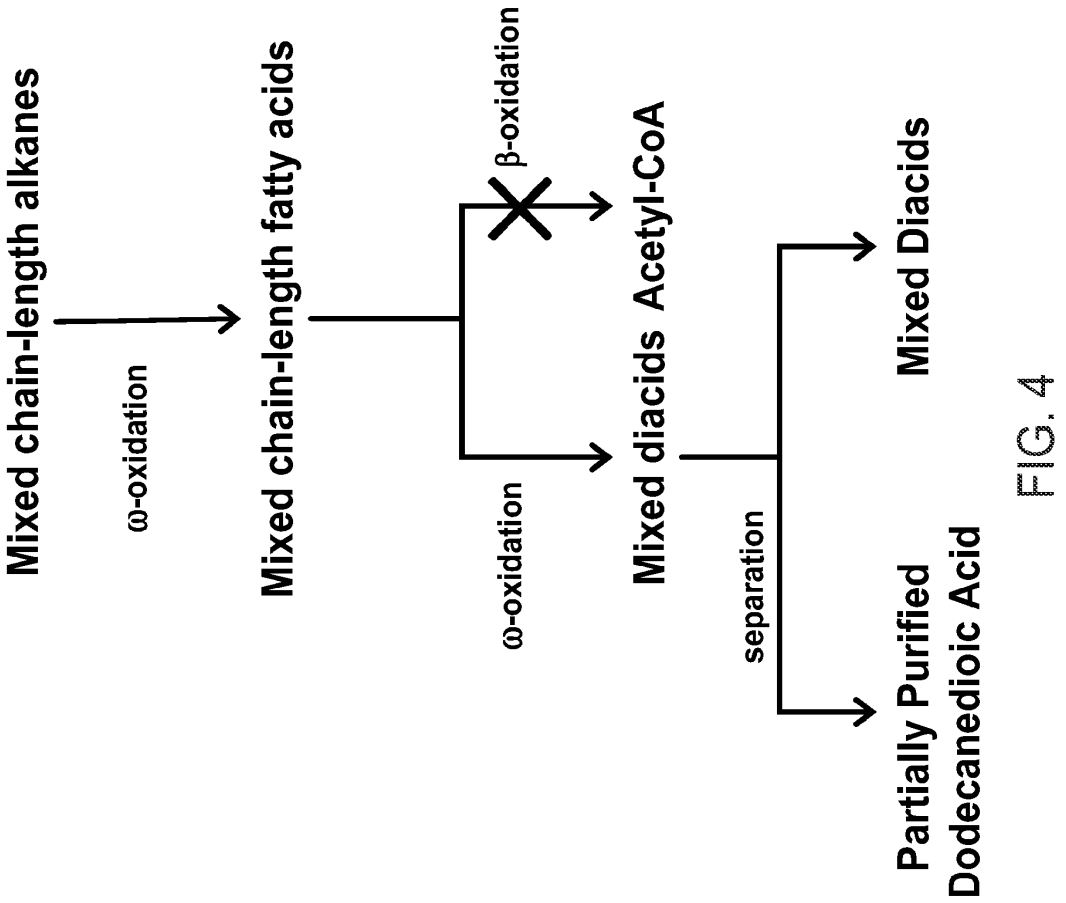

Mixed chain-length alkanes
ω-oxidation
Mixed chain-length fatty acids
β-oxidation
ω-oxidation
Mixed diacids Acetyl-CoA
separation
Mixed Diacids
Partially Purified Dodecanedioic Acid

FIG. 4

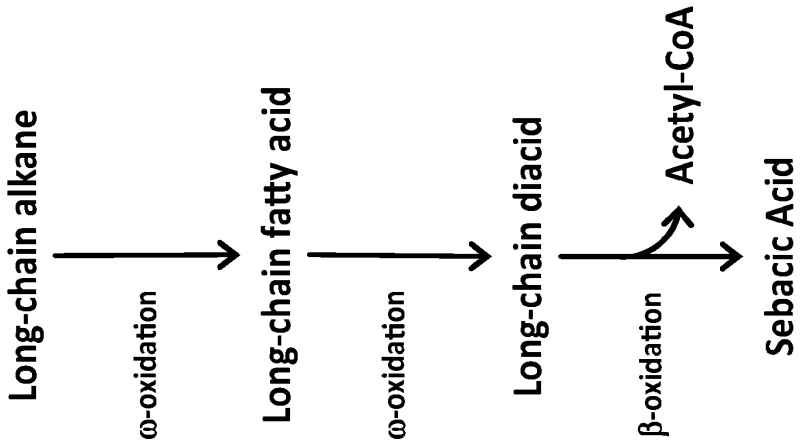

Long-chain alkane
ω-oxidation
Long-chain fatty acid
ω-oxidation
Long-chain diacid
β-oxidation
Acetyl-CoA
Sebacic Acid

FIG. 5

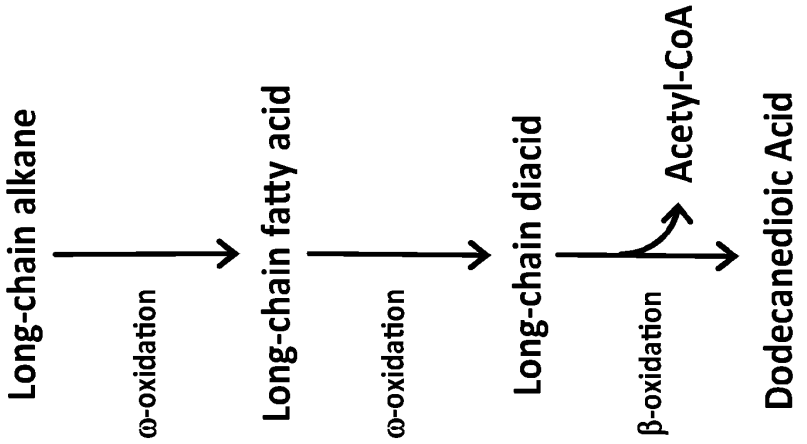

Long-chain alkane
ω-oxidation
Long-chain fatty acid
ω-oxidation
Long-chain diacid
β-oxidation
Acetyl-CoA
Dodecanedioic Acid

FIG. 6

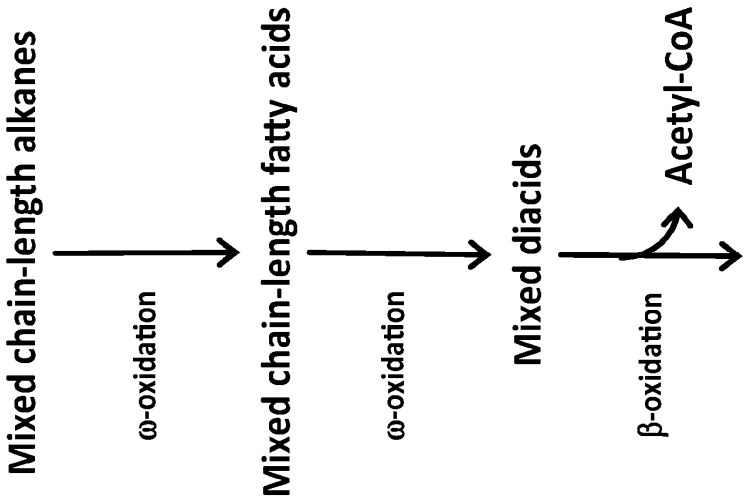

Mixed chain-length alkanes
ω-oxidation
Mixed chain-length fatty acids
ω-oxidation
Mixed diacids
β-oxidation
Acetyl-CoA
Sebacic Acid

FIG. 7

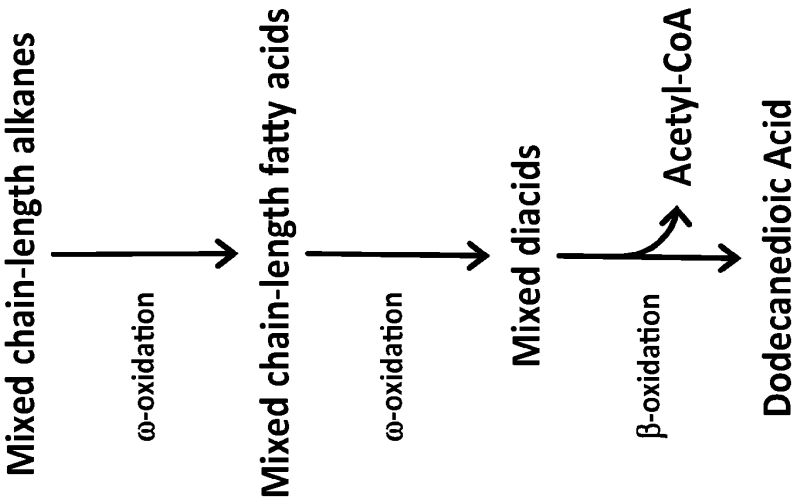

Mixed chain-length alkanes
ω-oxidation
Mixed chain-length fatty acids
ω-oxidation
Mixed diacids
Acetyl-CoA
β-oxidation
Dodecanedioic Acid

FIG. 8

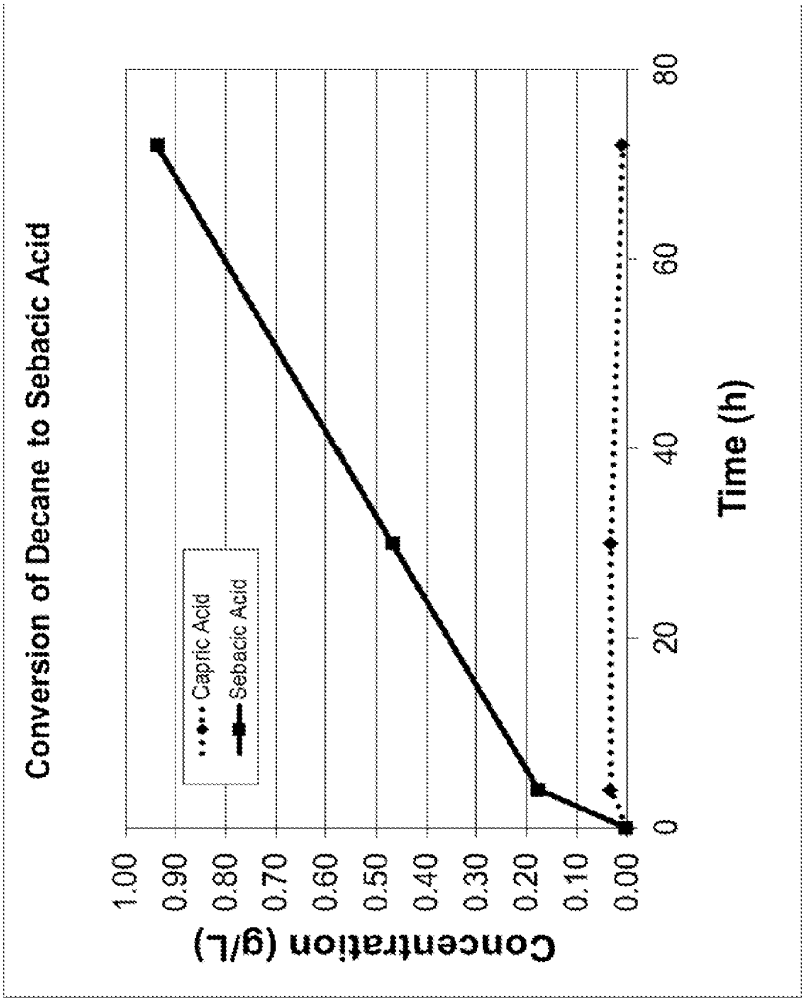

Conversion of Decane to Sebacic Acid
Concentration (g/L)
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
0
20
40
60
80
Time (h)
Capric Acid
Sebacic Acid

FIG. 9

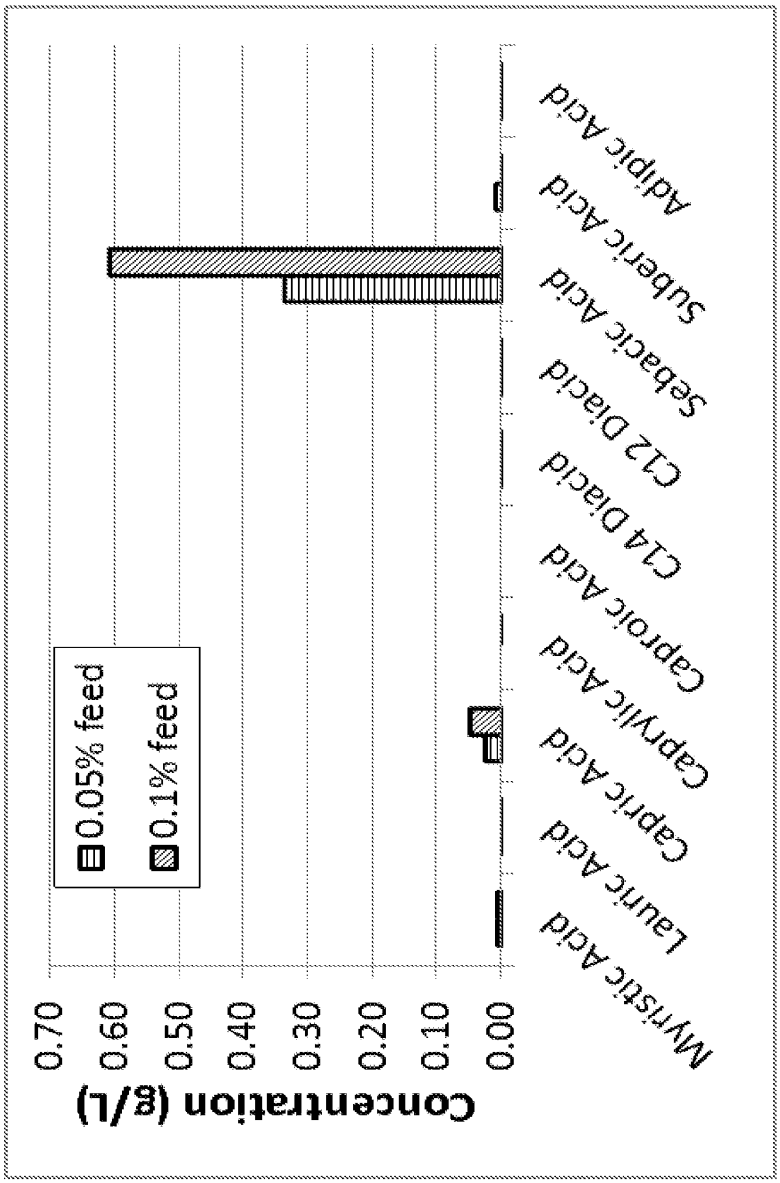

Concentration (g/L)
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
0.05% feed
0.1% feed
Myristic Acid
Lauric Acid
Capric Acid
Caprylic Acid
Caproic Acid
C14 Diacid
C12 Diacid
Sebacic Acid
Suberic Acid
Adipic Acid

FIG. 10

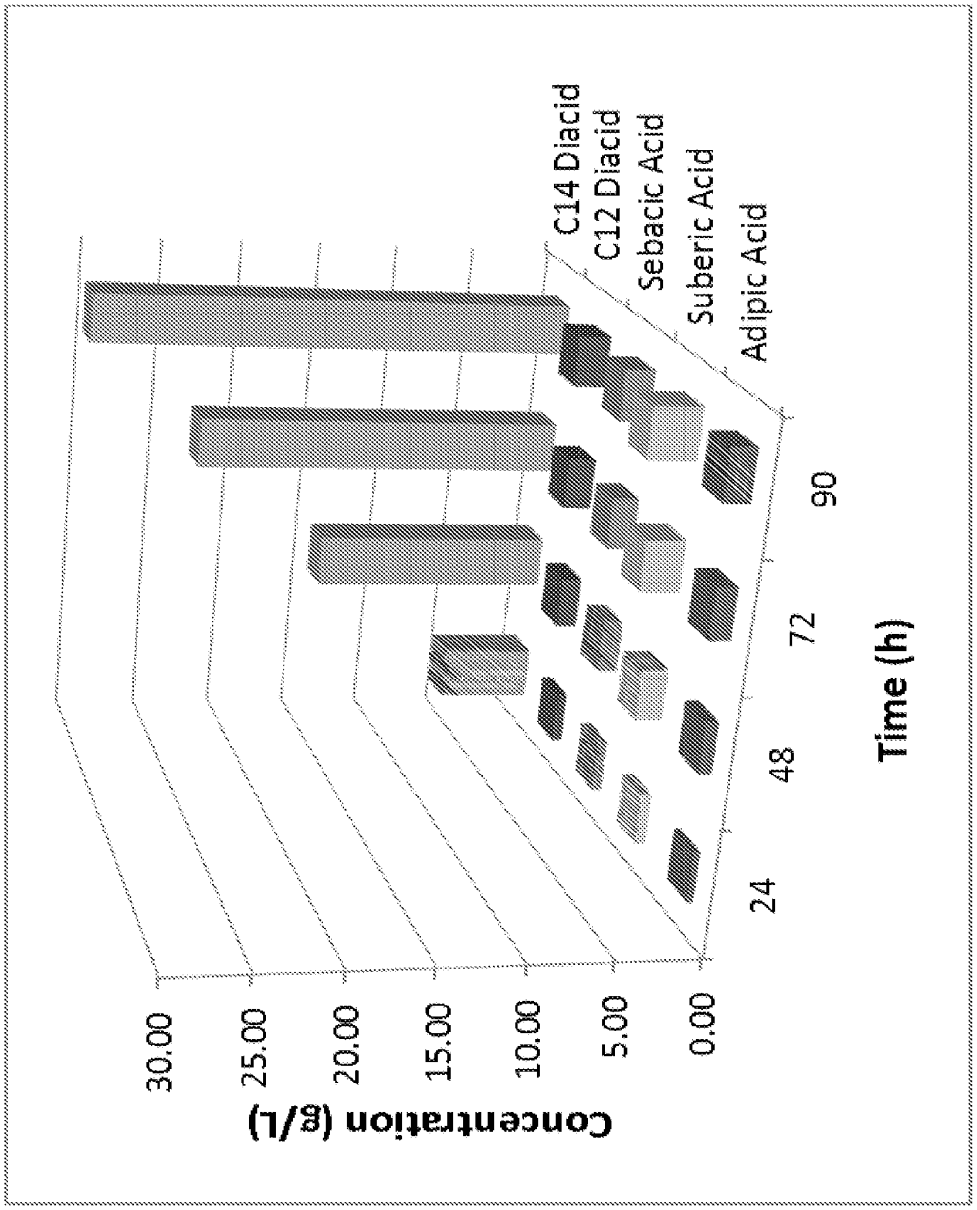

C14 Diacid
C12 Diacid
Sebacic Acid
Suberic Acid
Adipic Acid
30.00
25.00
20.00
15.00
10.00
5.00
0.00
Concentration (g/L)
24
48
72
90
Time (h)

FIG. 11

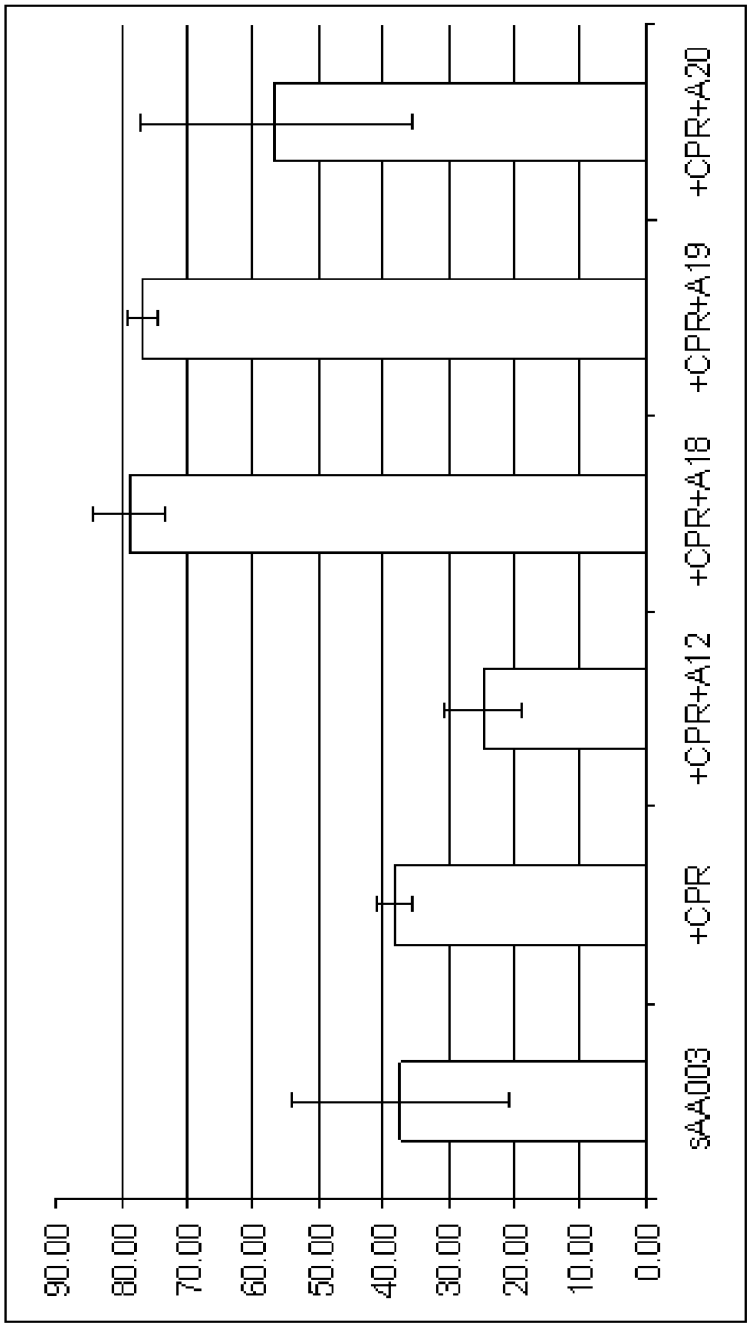

90.00
80.00
70.00
60.00
50.00
40.00
30.00
20.00
10.00
0.00
sAA003
+CPR
+CPR+A12
+CPR+A18
+CPR+A19
+CPR+A20

FIG. 12

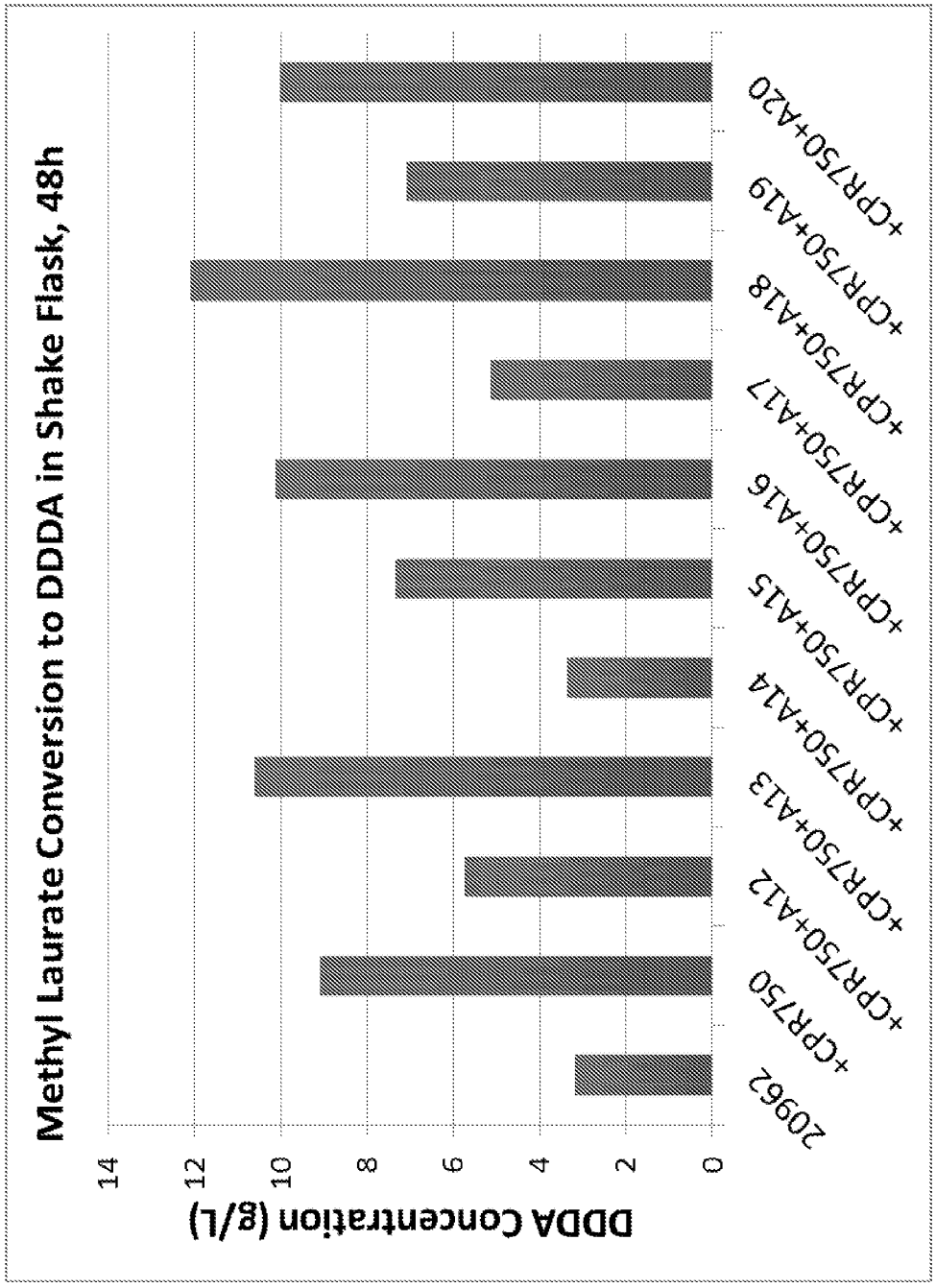

Methyl Laurate Conversion to DDDA in Shake Flask, 48h
DDDA Concentration (g/L)
0
2
4
6
8
10
12
14
20962
+CPR750
+CPR750+A12
+CPR750+A13
+CPR750+A14
+CPR750+A15
+CPR750+A16
+CPR750+A17
+CPR750+A18
+CPR750+A19
+CPR750+A20

FIG. 13

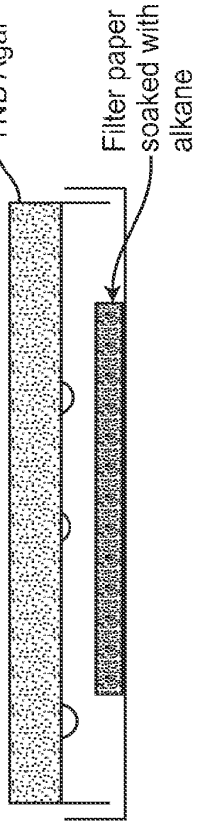


FIG. 14

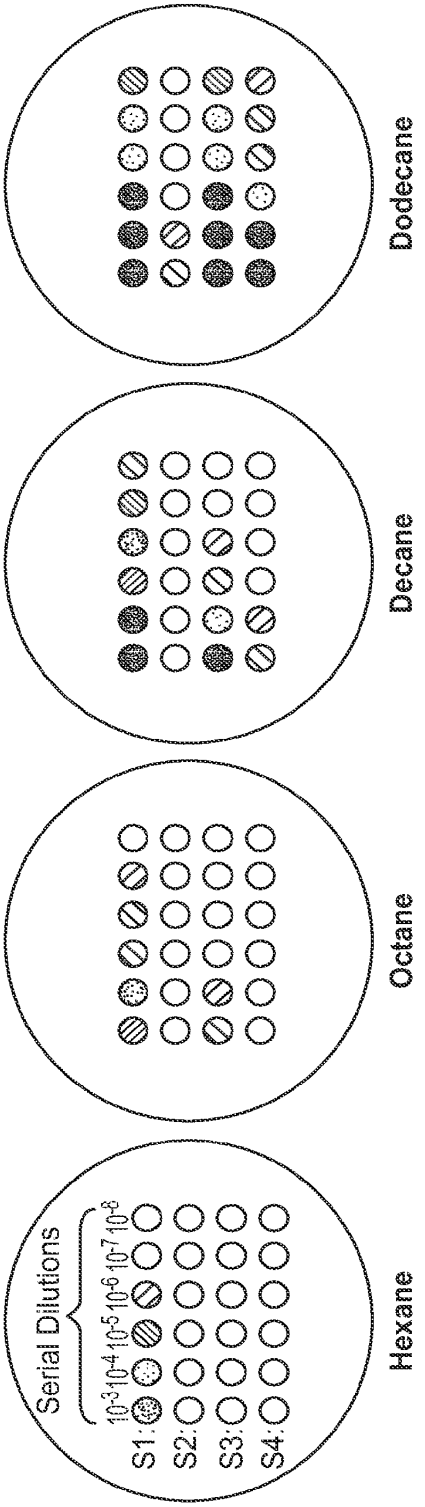

Dodecane
Decane
Octane
Hexane
Serial Dilutions
10^{-3} 10^{-4} 10^{-5} 10^{-6} 10^{-7} 10^{-8}
S1:
S2:
S3:
S4:

FIG. 15

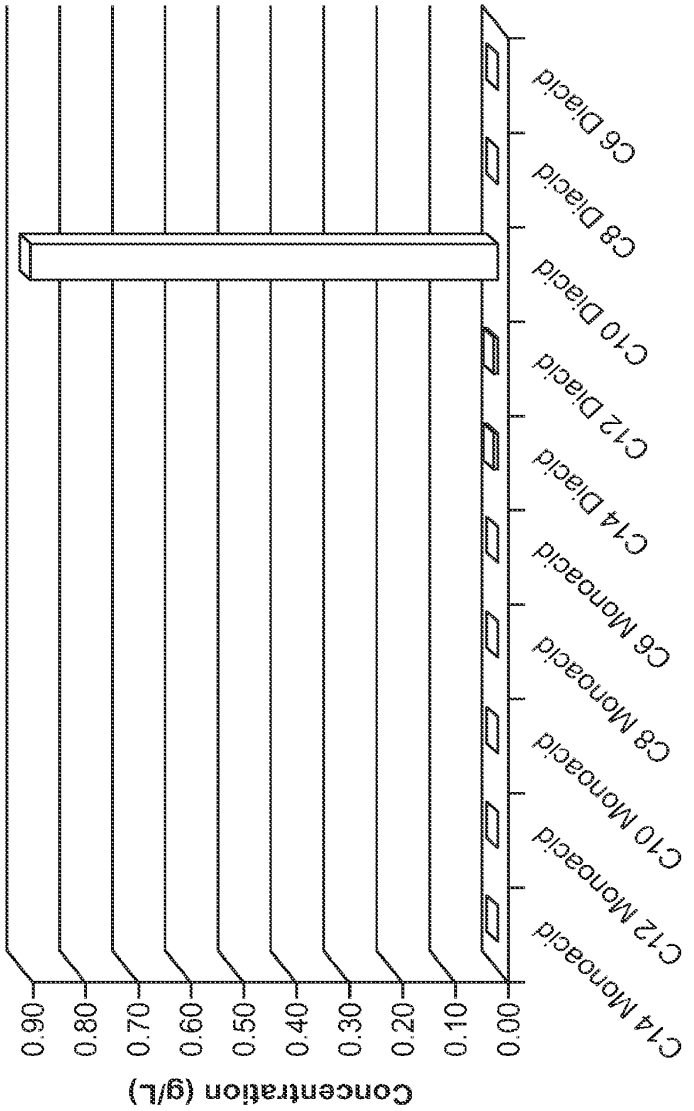

Concentration (g/L)
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
C14 Monoacid
C12 Monoacid
C10 Monoacid
C8 Monoacid
C6 Monoacid
C14 Diacid
C12 Diacid
C10 Diacid
C8 Diacid
C6 Diacid

FIG. 19

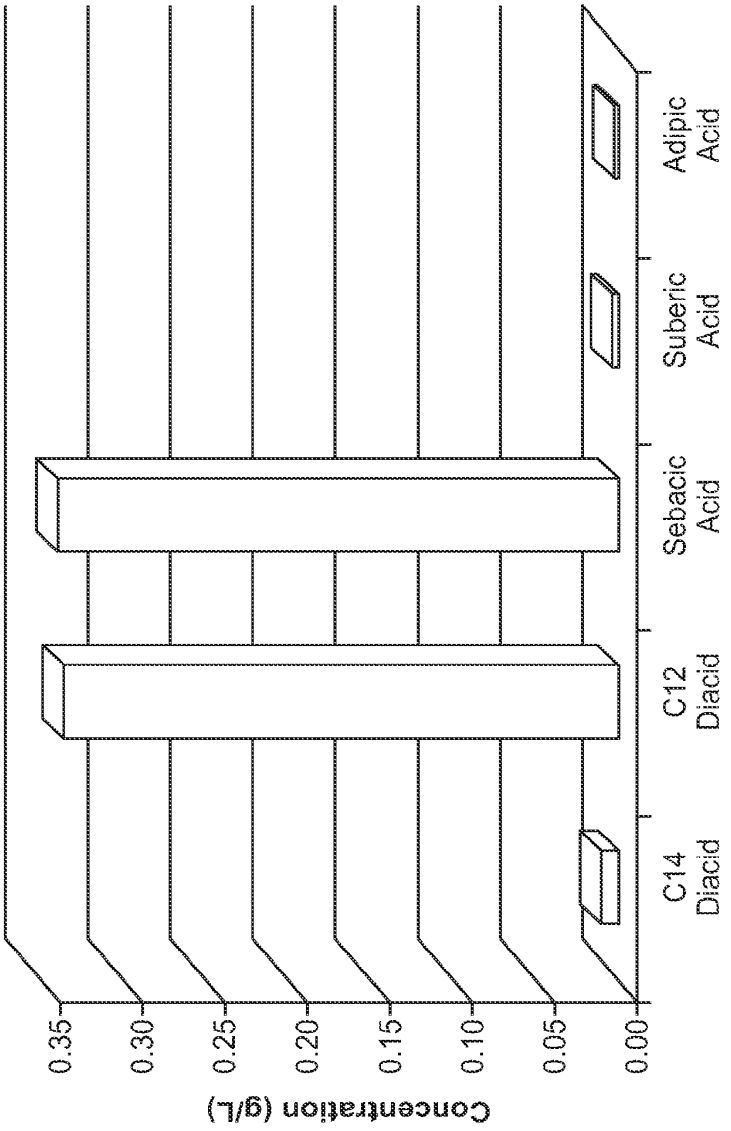

Concentration (g/L)
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
C14 Diacid
C12 Diacid
Sebacic Acid
Suberic Acid
Adipic Acid

FIG. 17

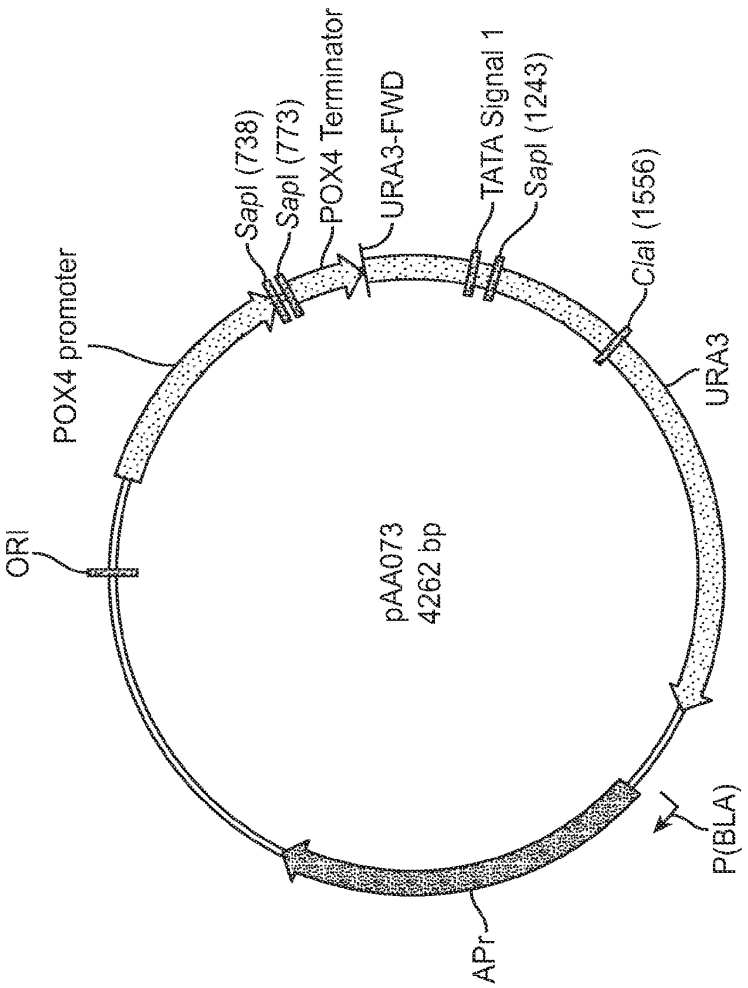

pAA073
4262 bp
ORI
POX4 promoter
SapI (738)
SapI (773)
POX4 Terminator
URA3-FWD
TATA Signal 1
SapI (1243)
ClaI (1556)
URA3
P(BLA)
APr

FIG. 18

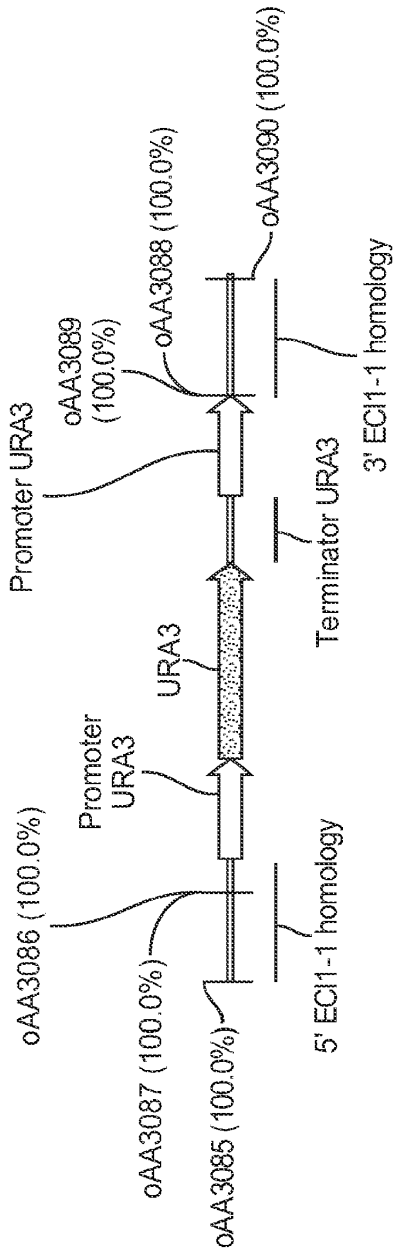

oAA3086 (100.0%)
oAA3087 (100.0%)
oAA3085 (100.0%)
Promoter URA3
URA3
5' ECI1-1 homology
Promoter URA3
oAA3089 (100.0%)
oAA3088 (100.0%)
oAA3090 (100.0%)
Terminator URA3
3' ECI1-1 homology

FIG. 19

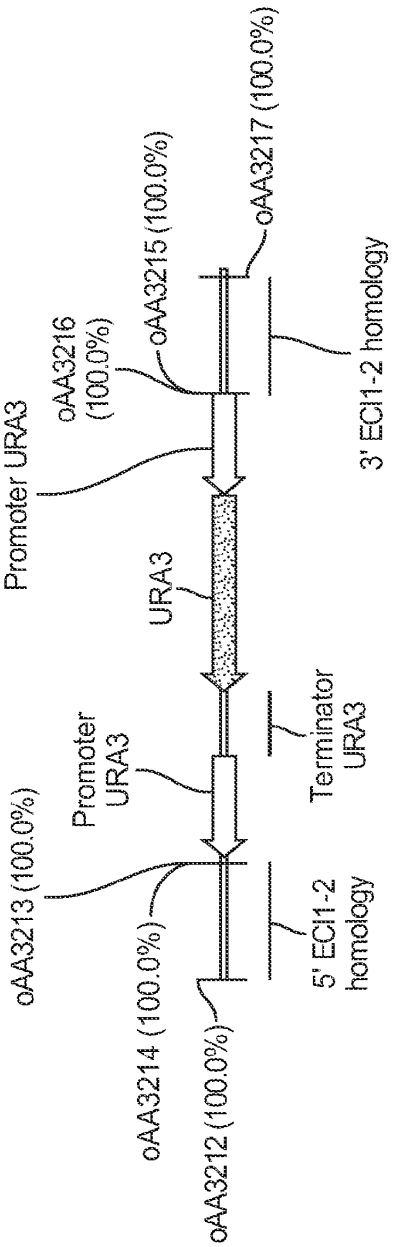

oAA3213 (100.0%)
oAA3214 (100.0%)
oAA3212 (100.0%)
Promoter URA3
5' ECI1-2 homology
URA3
Promoter URA3
Terminator URA3
oAA3216 (100.0%)
oAA3215 (100.0%)
oAA3217 (100.0%)
3' ECI1-2 homology

FIG. 20

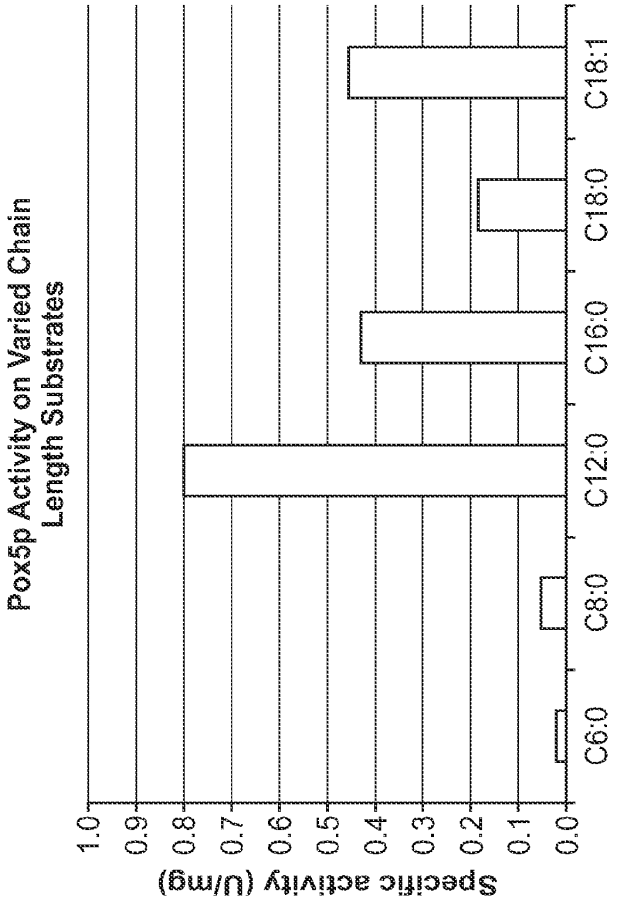

Pox5p Activity on Varied Chain Length Substrates
Specific activity (U/mg)
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
C6:0
C8:0
C12:0
C16:0
C18:0
C18:1

FIG. 21

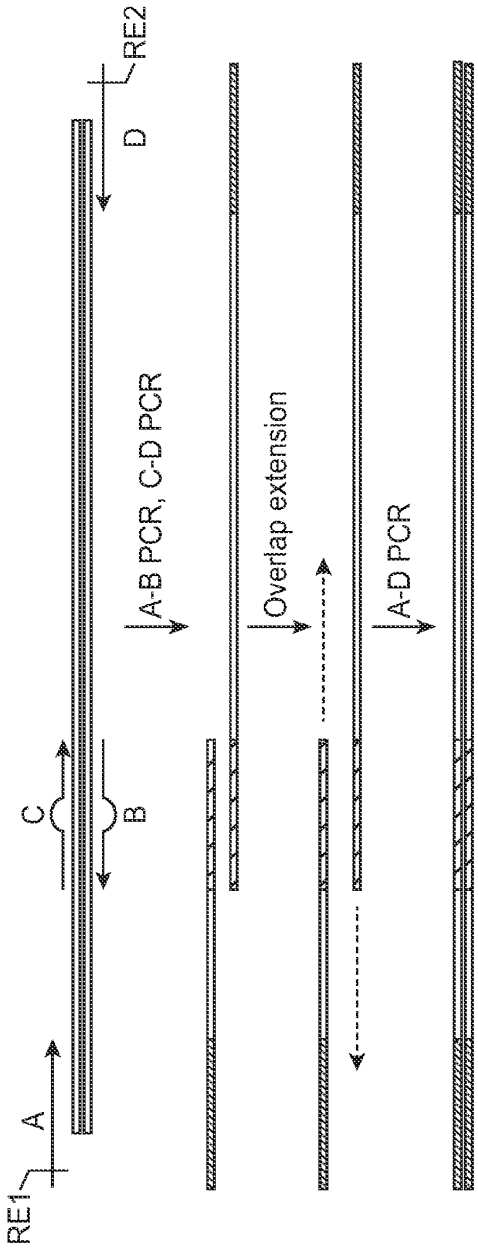


FIG. 22

```
AcoII   1   MNPDLRKERASATFNPELITHILDGSPENTRRREIENLIINDPDFQHEIMNFLTRSQRY  60
            MNPDLRKERASATFNPELITHILDGSPENTRRREIENLIINDPDFQHEDYNFLTRSQRY
AcoI    1   MNPDLRKERASATFNPELITHILDGSPENTRRREIENLIINDPDFQHEDYNFLTRSQRY  60

AcoII  61   EVAVKKSATMVKKMPEYGISDPEEIMWFKNSVIRGHPEPILDLHLGVMFLPTLLHQATAEQQ  120
            EVAVKKSATMVKKMPEYGISDPEEIMWFK          EP+ L+  MF+PTLL+Q T  QQ
AcoI   61   EVAVKKSATMVKKMPEYGTSDPEEIMWFKKLYLANFVEPVGLNYSMFIPTLLNQGTTAQQ  120

AcoII 121   ERFMPAWMEITGTYAQTEMGHGTHLRGLETTATYDPKTQEFTLNSPTVTSIKWWPGGL  180
            E++  P+  L+I GTYAQTEMGHGTHLRGLETTATYDPKTQEFTLNSPTVTSIKWWPGGL
AcoI   121   EKWMRPSQELQIIGTYAQTEMGHGTHLRGLETTATYDPKTQEFTLNSPTVTSIKWWPGGL  180
```

FIG. 23

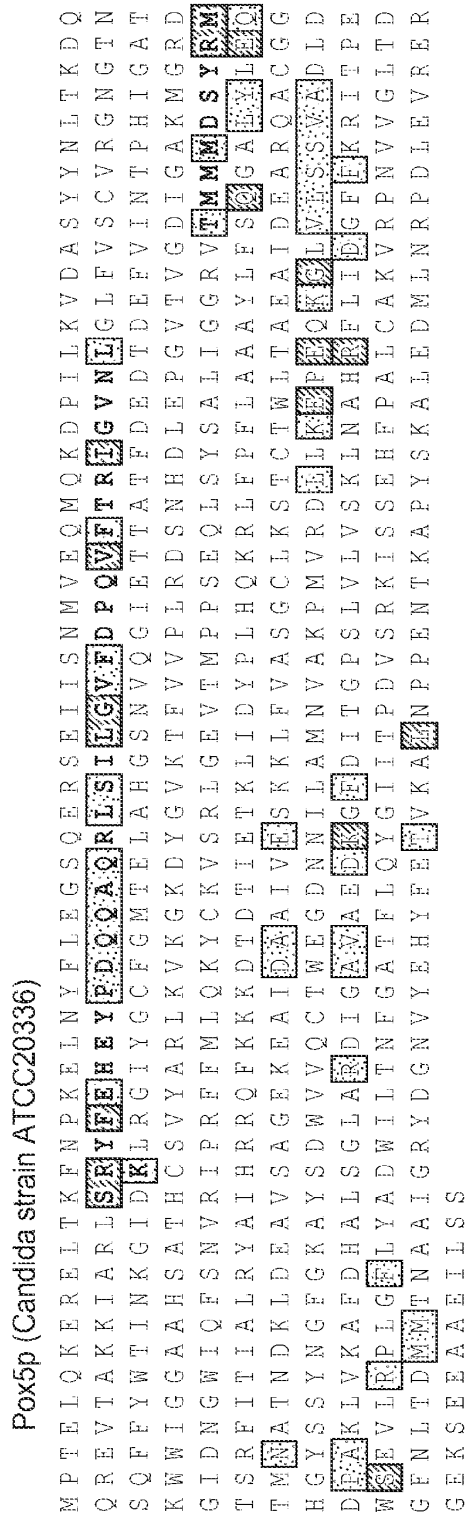


FIG. 24A

FIG. 24B

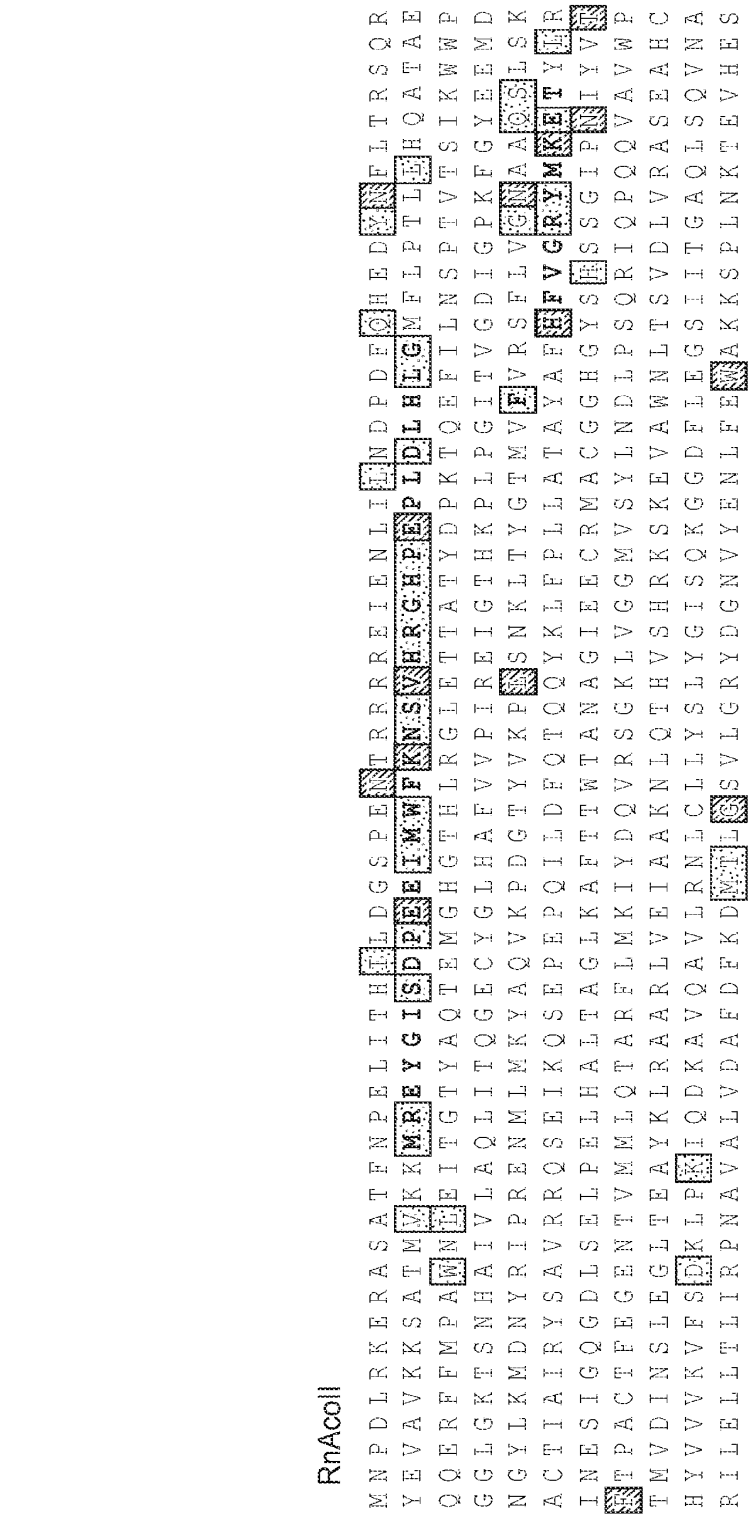


FIG. 25

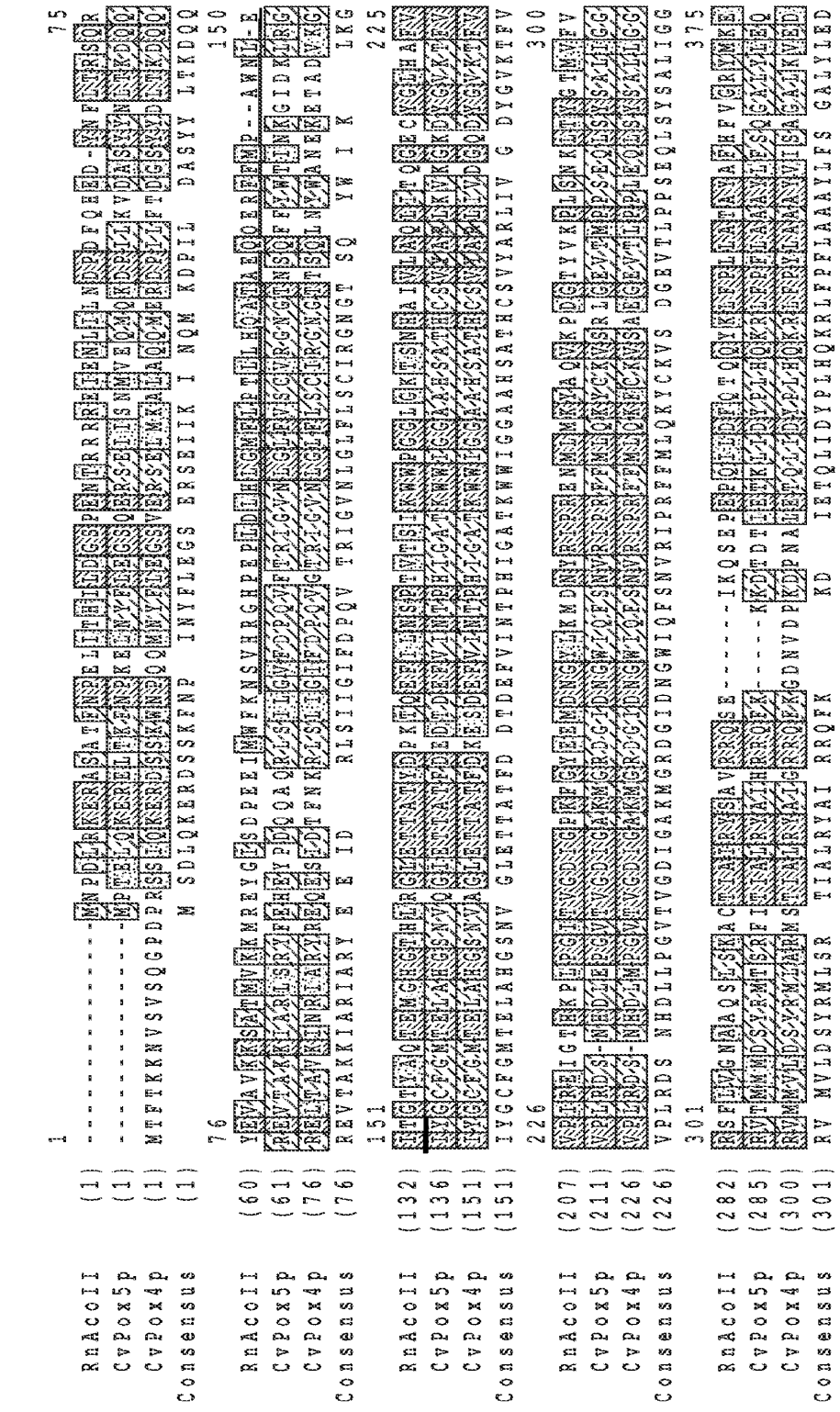


FIG. 26

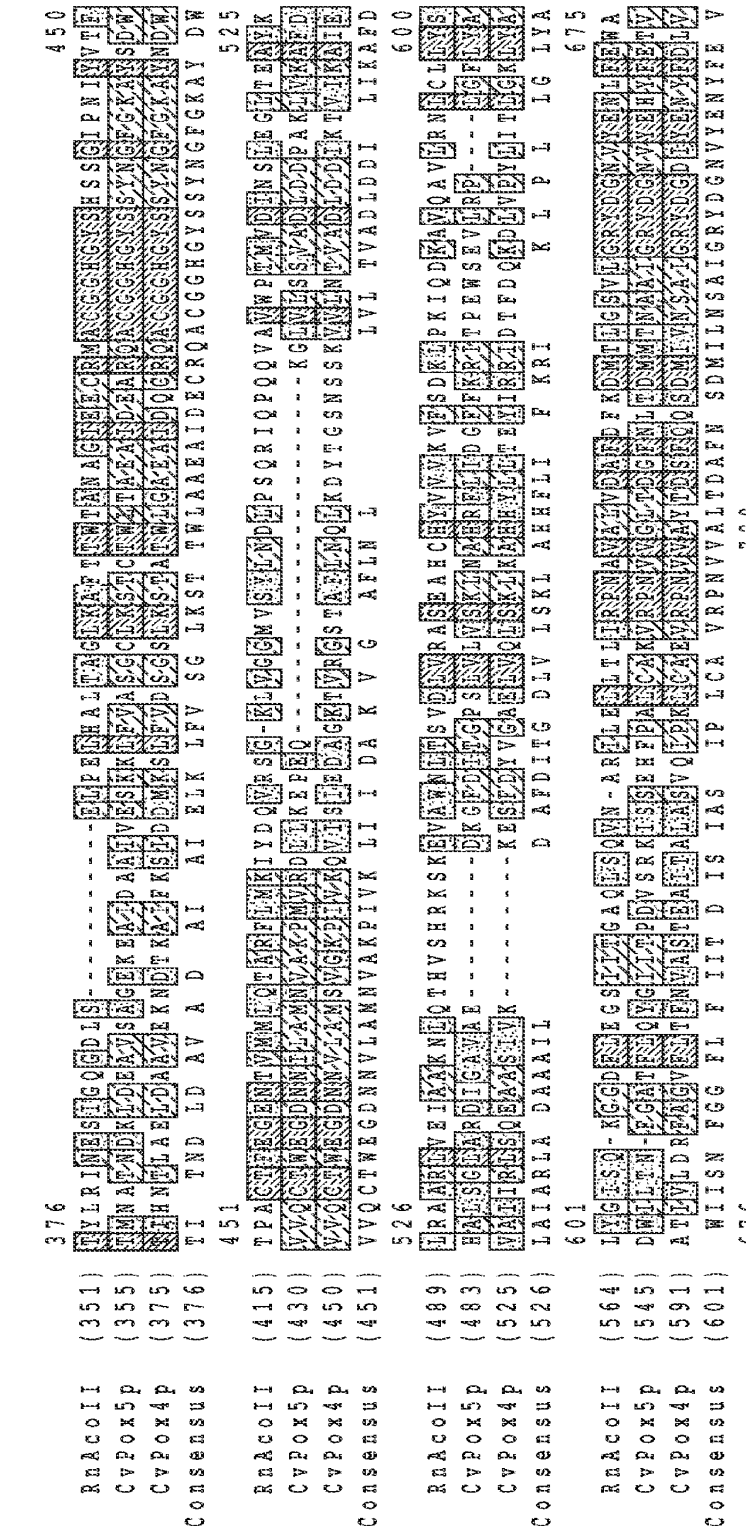


FIG. 26 (Cont.)

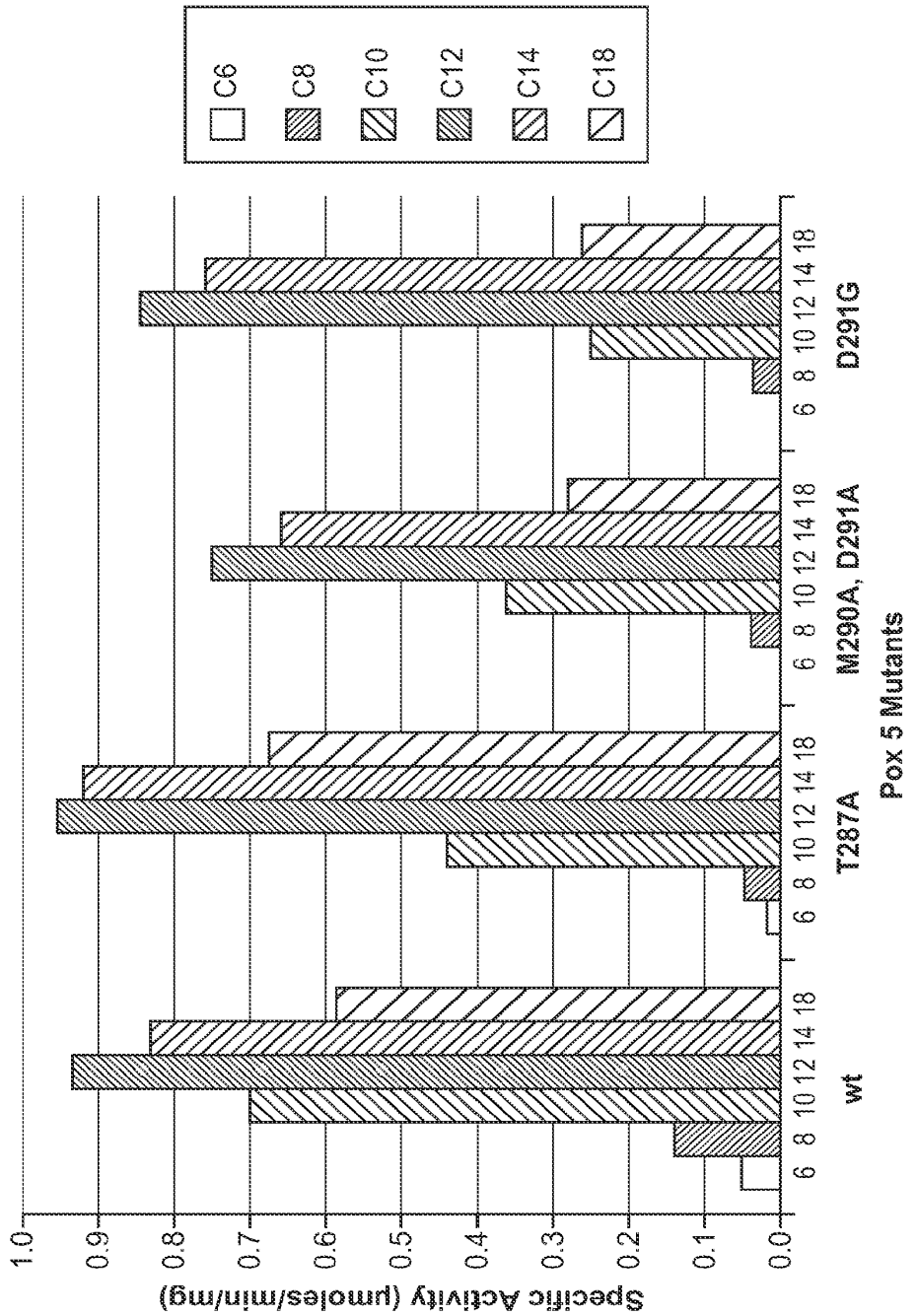

C6
C8
C10
C12
C14
C18
Specific Activity (μmoles/min/mg)
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
6 8 10 12 14 18
wt
T287A
M290A, D291A
D291G
Pox 5 Mutants

FIG. 27

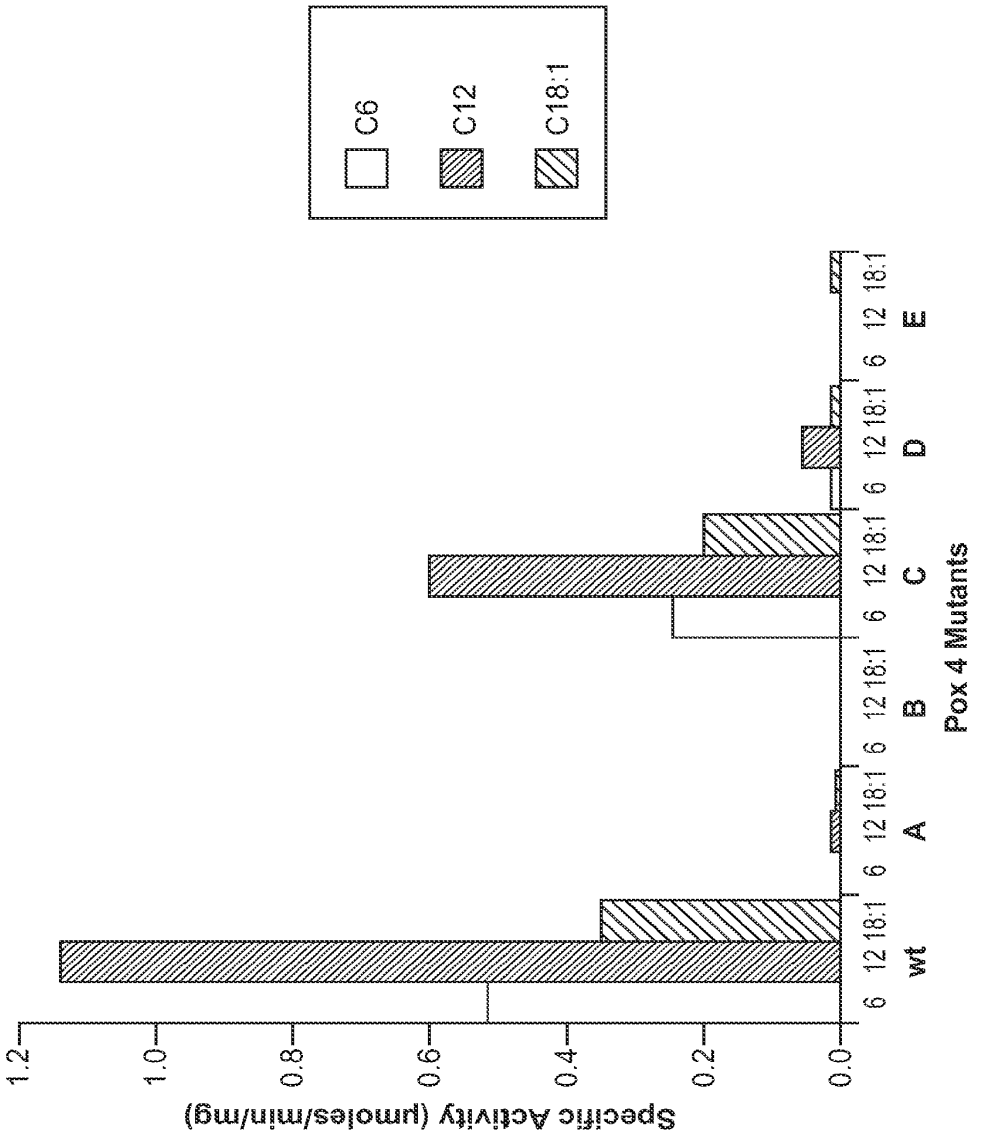

C6
C12
C18:1
Specific Activity (μmoles/min/mg)
0.0
0.2
0.4
0.6
0.8
1.0
1.2
6 12 18:1 6 12 18:1 6 12 18:1 6 12 18:1 6 12 18:1
wt A B C D E
Pox 4 Mutants

FIG. 28

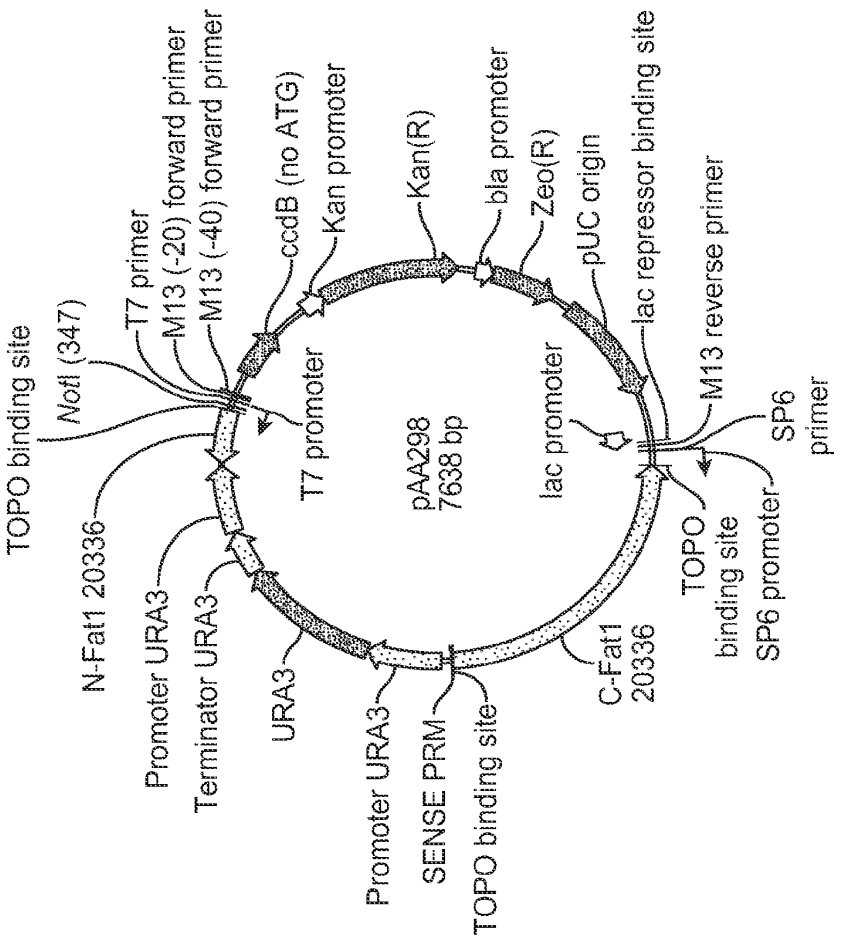

TOPO binding site
NotI (347)
T7 primer
M13 (-20) forward primer
M13 (-40) forward primer
ccdB (no ATG)
Kan promoter
Kan(R)
bla promoter
Zeo(R)
pUC origin
lac repressor binding site
M13 reverse primer
SP6 primer
TOPO binding site
SP6 promoter
C-Fat1 20336
T7 promoter
pAA298 7638 bp
lac promoter
N-Fat1 20336
Promoter URA3
Terminator URA3
URA3
Promoter URA3
SENSE PRM
TOPO binding site

FIG. 29

FIG. 30A

```
                1                                                    50
     POX5     (1) ----------------------MPTELQKERELTKFNPKELNYFLEGSQE
   YlPox2     (1) -MNPNNTGTIEINGKEYNTFTEPPVAMAQERAKTSFPVREMTYFLDGGEK
   YlPox3     (1) MISPNLTANVEIDGKQYNTFTEPPKALAGERAKVKFPIKDMTEFLHGGEE
   YlPox5     (1) --MNNNPTNVILGGKEYDTFTEPPAQMELERAKTQFKVRDVTNFLTGSEQ
     POX4     (1) -------MTFTKKNVSVSQGPDPRSSIQKERDSSKWNPQQMNYFLEGSVE
Consensus     (1)     N    V I GK Y TFTEPP AL  ERAKTKF VKDMTYFLEGSEE
                51                                                   100
     POX5    (29) RSEIISNMVEQMQKDPILKVDASYYNLTKDQQREVTAKKIARLSRYFEHE
   YlPox2    (50) NTLKNEQIMEEIERDPLFNNDN-YYDLNKEQIRELTMERVAKLSLFVRDQ
   YlPox3    (51) NVTMIERLMTELERDPVLNVSG-DYDMPKEQLRETAVARIAALSGHWKKD
   YlPox5    (49) ETLLTERIMREIERDPVLNVAG-DYDADLPTKRRQAVERIGALARYLPKD
     POX4    (44) RSELMKALAQQMERDPILFTDGSYYDLTKDQQRELTAVKINRIARYREQE
Consensus    (51)  S LIE IM EIERDPILNVDG YYDL KDQ RELTV RIARLSRY   D
                101                                                  150
     POX5    (79) YPDQQAQRLSILGVFDPQVFTRIGVNLGLFVSCVRGNGTNSQFFYWTINK
   YlPox2    (99) PEDDIKKRFALIGIADMGTYTRLGVHYGLFFGAVRGTGTAEQFGHWIS-K
   YlPox3   (100) TEKEALLRSQLHGIVDMGTRTRLGVHTGLFMGAIRGSGTKEQYDYWVR-K
   YlPox5    (98) SEKEAILRGQLHGIVDMGTRTRIAVHYGLFMGAIRGSGTKEQYDYWVA-K
     POX4    (94) SIDTFNKRLSLIGIFDPQVGTRIGVNLGLFLSCIRGNGTTSQLNYWANEK
Consensus   (101) SEDE   R SLIGI DMGT TRIGVH GLFMGAIRGSGT EQF YWV  K
                151                                                  200
     POX5   (129) GIDKLRGIYGCFGMTELAHGSNVQGIETTATFDEDTDEFVINTPHIGATK
   YlPox2   (148) GAGDLRKFYGCFSMTELGHGSNLAGLETTAIYDEETDEFIINTPHIAATK
   YlPox3   (149) GAADVKGFYGCFAMTELGHGSNVAGLETTATYIQDTDEFIINTPNTGATK
   YlPox5   (147) GAATLHKFYGCFAMTELGHGSNVAGLETTATLDKDTDEFIINTPNSGATK
     POX4   (144) ETADVKGIYGCFGMTELAHGSNVAGLETTATFDKESDEFVINTPHIGATK
Consensus   (151) GAADLKGFYGCFAMTELGHGSNVAGLETTATFD DTDEFIINTPHIGATK
                201                                                  250
     POX5   (179) WWIGGAAHSATHCSVYARLKVKGKDYGVKTFVVPLRDSN-HDLEPGVTVG
   YlPox2   (198) WWIGGAAHTATHTVVFARLIVKGKDYGVKTFVVQLRNINDHSLKVGISIG
   YlPox3   (199) WWIGGAAHSATHTACFARLLVDGKDYGVKIFVVQLRDVSSHSLMPGIALG
   YlPox5   (197) WWIGGAAHSATHTACLARLIVDGKDYGVKIFIVQLRDLNSHSLLNGIAIG
     POX4   (194) WWIGGAAHSATHCSVYARLIVDGQDYGVKTFVVPLRDSN-HDLMPGVTVG
Consensus   (201) WWIGGAAHSATHTAVFARLIVDGKDYGVKTFVVQLRDIN HSLMPGITIG
                251                                                  300
     POX5   (228) DIGAKMGRDGIDNGWIQFSNVRIPRFFMLQKYCKVSRLGEVTMPPSEQLS
   YlPox2   (248) DIGKKMGRDGIDNGWIQFTNVRIPRQNLLMKYTKVDREGNVTQPPLAQLT
   YlPox3   (249) DIGKKMGRDAIDNGWIQFTNVRIPRQNMLMKYAKVSSTGKVSQPPLAQLT
   YlPox5   (247) DIGKKMGRDAIDNGWIQFTDVRIPRQNMLMRYDRVSRDGEVTTSELAQLT
     POX4   (243) DIGAKMGRDGIDNGWIQFSNVRIPRFFMLQKFCKVSAEGEVTLPPLEQLS
Consensus   (251) DIGKKMGRDGIDNGWIQFTNVRIPRQNMLMKY KVSREGEVT PPLAQLT
                301                                                  350
     POX5   (278) YSALIGGRVTMMMDSYRMTSRFITIALRYAIHRRQFK-----KKDTDTIE
   YlPox2   (298) YGSLITGRVSMASDSHQVGKRFITIALRYACIRRQFS------TTPGQPE
   YlPox3   (299) YGALIGGRVTMIADSFFVSQRFITIALRYACVRRQFG------TTPGQPE
   YlPox5   (297) YGALLSGRVTMIAESHLLSARFLTIALRYACIRRQFG------AVPDKPE
     POX4   (293) YSALLGGRVMMVLDSYRMLARMSTIALRYAIGRRQFKGDNVDPKDPNALE
Consensus   (301) YGALIGGRVTMIADSH MSARFITIALRYACIRRQF         P  PE
```

FIG. 30B

```
                351                                                 400
     POX5  (323) TKL-DYPLHQKRLFPFLAAAYLFSQGALYLEQTMNATNDKLDEAVSAGEK
   YlPox2  (342) TKI-DYPYHQRRLLPLLAYVYALKMTADEVGALFSRTMLKMDDLKP-DDK
   YlPox3  (343) TKI-DYPYHQRRLLPLLAFTYAMKMAADQSQIQYDQTTDLLQTIDP-KDK
   YlPox5  (341) TKL-DYPYHQRRLLPLLAYTYAMKMGADEAQQQYNSSFGALLKLNPVKDA
     POX4  (343) TQL-DYPLHQKRLFPYLAAAYVISAGALKVEDTIHNTLAELDAAVEKNDT
Consensus  (351) TKL-DYPYHQRRLLPLLAY YAMKMGAD V   Y  T   LD L P  DK
                401                                                 450
     POX5  (373) EAIDAAIVESKKLFVASGCLKSTCTWLTAEAIDEARQACGGHGYSSYNGF
   YlPox2  (391) AGLNEVVSDVKELFSVSAGLKAFSTWACADVIDKTRQACGGHGYSGYNGF
   YlPox3  (392) GALGKAIVDLKELFASSAGLKAFTTWTCANIIDQCRQACGGHGYSGYNGF
   YlPox5  (391) EKFAVATADLKALFASSAGMKAFTTWAAAKIIDECRQACGGHGYSGYNGF
     POX4  (393) KAIFKSIDDMKSLFVDSGSLKSTATWLGAEAIDQCRQACGGHGYSSYNGF
Consensus  (401)  AI  AI DLK LFASSAGLKAFTTW  AEIID CRQACGGHGYSGYNGF
                451                                                 500
     POX5  (423) GKAYSDWVVQCTWEGDNNILAMNVAKPMVRDLLKEPEQ------------
   YlPox2  (441) GQAYADWVVQCTWEGDNNILTLSAGRALIQSAVALRKGE-PVGNAVSYLK
   YlPox3  (442) GQAYADWVVQCTWEGDNNVLCLSMGRGLIQSCLGHRKGK-PLGSSVGYLA
   YlPox5  (441) CQAYADWVVQCTWEGDNNVLCLSMCRSLIQSCIAMRKKKCHVCKSVEYLQ
     POX4  (443) GKAYNDWVVQCTWEGDNNVLAMSVGKPIVKQVISIEDAGKTVRGSTAFLN
Consensus  (451) GQAYADWVVQCTWEGDNNVL LSMGR LIQS IAIRKG   VG SV YL
                501                                                 550
     POX5  (461) ----------KGLVLSSVADLDDPAKLVKAFDHALSGLARDIGAVAE---
   YlPox2  (490) RY------KDLANAKLNGRSLTDPKVLVEAWEVAAGNIINRATDQYEKLI
   YlPox3  (491) N-------KGLEQATLSGRDLKDPKVLIEAWEKVANGAIQRATDKFVELT
   YlPox5  (491) RR------DELQNARVDNKPLTDPAVLITAWEKVACEAINRATDSFIKLT
     POX4  (493) QLKDYTGSNSSKVVLNTVADLDDIKTVIKAIEVAIIRLSQEAASIVK---
Consensus  (501)           L NA LS RDL DPKVLI AWE AA  LINRATD F  L
                551                                                 600
     POX5  (498) ---D---KGFDITGPSLVLVSKLNAHRFLIDGFFKRITPEWSEVLRP---
   YlPox2  (534) GEGLNADQAFEVLSQQRFQAAKVHTRRHLIAAFFSRIDTEAGEAIKQPLL
   YlPox3  (534) KGGLSPDQAFEELSQQRFQCAKIHTRKHLVTAFYERINASAKADVKPYLI
   YlPox5  (535) QEGLSPDQAFEELSQQRFECARIHTRKHLITSFYARISK-AKARVKPHLT
     POX4  (540) ---K---ESFDYVGAELVQLSKLKAHHYLLTEYIRRIDTFDQKDLVPYLI
Consensus  (551)   GL  DQAFE LSQQRFQ AKIHTRKHLITAFF RI   A   LKPYLI
                601                                                 650
     POX5  (539) -LGFLYADWILTN-FGATFLQYGIITPDVSRKISSEHFPALCAKVRPNVV
   YlPox2  (584) NLALLFALWSIEE-DSGLFLREGFLEPKDIDTVT-ELVNKYCTTVREEVI
   YlPox3  (584) NLANLFTLWSIEE-DSGLFLREGFLQPKDIDQVT-ELVNHYCKEVRDQVA
   YlPox5  (584) VLANLFAVWSIEE-DSGLFLREGCFEPAEMDEIT-ALVDELCCEAREQVI
     POX4  (584) TLGKLYAATIVLDRFAGVFLTFNVASTEAITALASVQIPKLCAEVRPNVV
Consensus  (601)  LA LFALWSIEE DSGLFLREG L P DID IT ELV  LC EVRENVI
                651                                                 700
     POX5  (587) GLTDGFNLTDMMTNAAIGRYDGNVYEHYFETVKALNPPENTKAPYSKALE
   YlPox2  (632) GYTDAFNLSDYFINAPIGCYDGDAYRHYFQKVNEQNPARDPRPPYYASTL
   YlPox3  (632) GYTDAFGLSDWFINAPIGNYDGDVYKHYFAKVNQQNPAQNPRPPYYESTL
   YlPox5  (632) GFTDAFNLSDFFINAPIGRFDGDAYKHYMDEVKAANNPRNTHAPYYETKL
     POX4  (634) AYTDSFQQSDMIVNSAIGRYDGDIYENYFDLVKLQNPPSKTKAPYSDALE
Consensus  (651) GYTDAFNLSDFFINAPIGRYDGDVYKHYFD VK QNPP NTKAPYYEA L
                701                   727
     POX5  (637) DMLNRPDLEVRERCEKSEEAAEILSS-
   YlPox2  (682) KPFLFREEEDDDICELDEE--------
   YlPox3  (682) RPFLFREDEDDDICELDEE--------
   YlPox5  (682) RPFLFRPDEDEEICDLDE---------
     POX4  (684) AMLNRPTLDERERFEKSDETAAILSK-
Consensus  (701) RPFLFREDEDDEICELDEE
```

FIG. 31

Sebacic (C10:2) from Linoleic acid

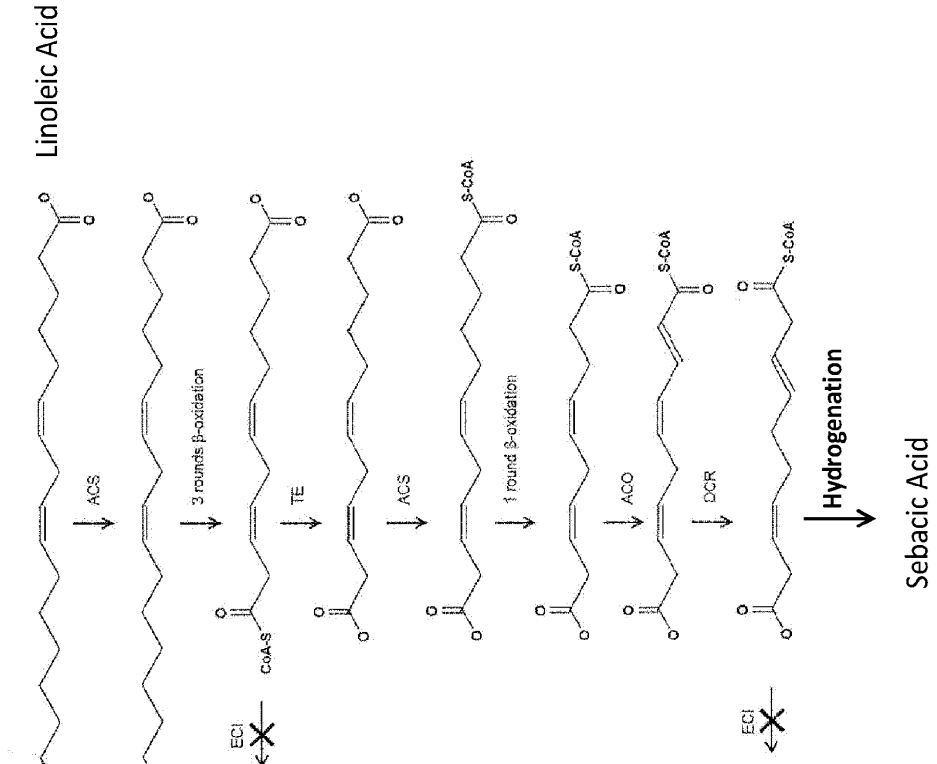


FIG. 32

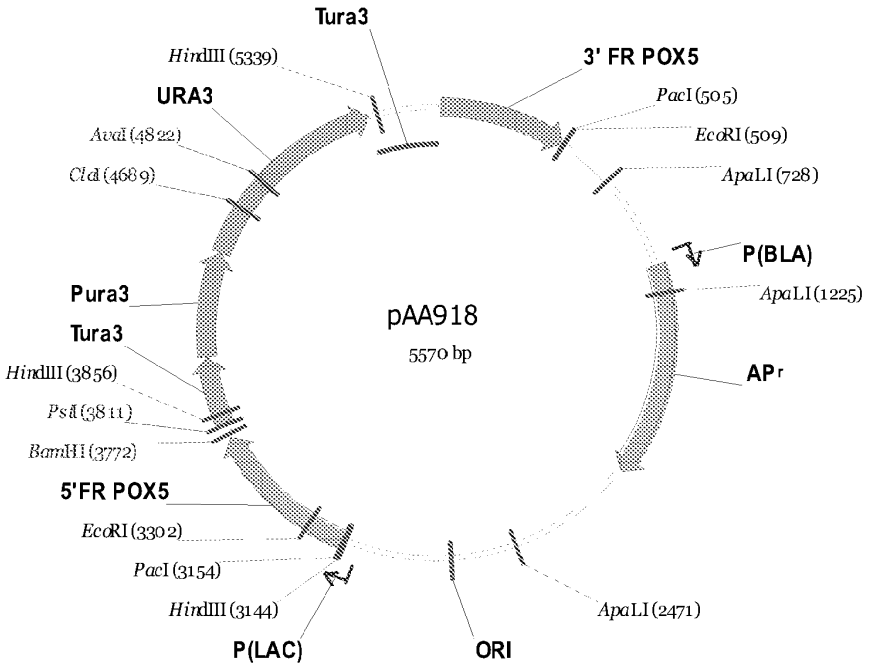

Tura3
HindIII (5339)
3' FR POX5
URA3
PacI (505)
AvaI (4822)
EcoRI (509)
ClaI (4689)
ApaLI (728)
P(BLA)
Pura3
ApaLI (1225)
pAA918
5570 bp
Tura3
APr
HindIII (3856)
PstI (3811)
BamHI (3772)
5'FR POX5
EcoRI (3302)
PacI (3154)
HindIII (3144)
ApaLI (2471)
P(LAC)
ORI

Fig. 33

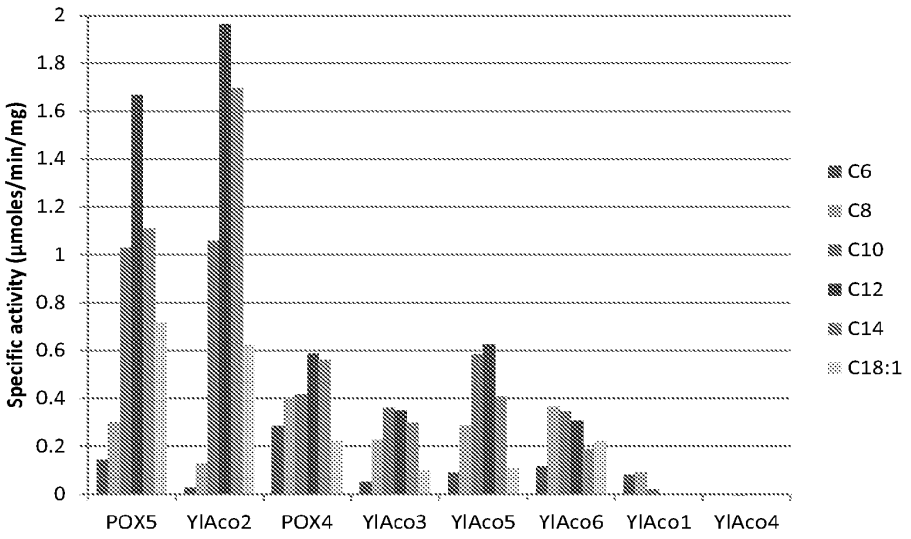
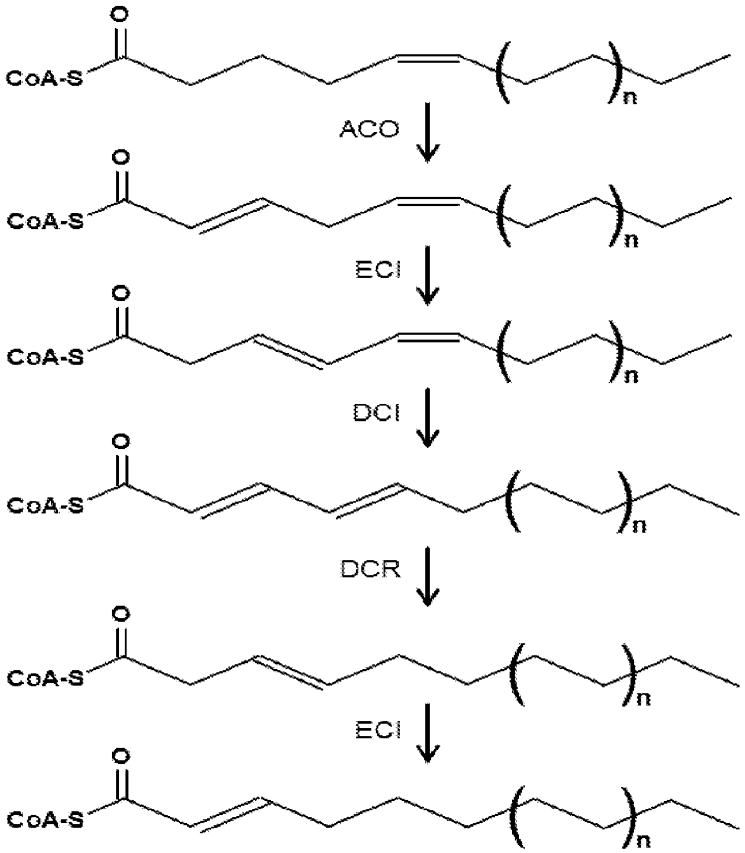

Specific activity (µmoles/min/mg)
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
POX5
YlAco2
POX4
YlAco3
YlAco5
YlAco6
YlAco1
YlAco4
C6
C8
C10
C12
C14
C18:1

FIG. 34



(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
24 May 2018 (24.05.2018)

W!P O I PCT

(10) International Publication Number
WO 2018/094110 A2

(51) International Patent Classification:
Not classified

(21) International Application Number:
PCT/US20 17/062 100

(22) International Filing Date:
16 November 201 7 (16.11.201 7)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

62/423,196	16 November 2016 (16. 11.2016)	US
15/469,452	24 March 20 17 (24 .03 .20 17)	US
62/481,601	04 April 20 17 (04 .04 .20 17)	US

(71) Applicant: **THE BOARD OF TRUSTEES OF THE LELAND STANFORD JUNIOR UNIVERSITY** [US/US]; Office Of General Counsel, Building 170, 3rd Floor, Main Quad, P.o.box 20386, Stanford, CA 94305-2038 (US).

(72) Inventors: **NAUGHTON, Brian, Thomas;** 217 Ada Avenue, #5 1, Mountain View, CA 94043 (US). **HARVEY, Colin;** 572 Franklin Street, Mountain View, CA 94041 (US). **SCHLECHT, Ulrich;** 226 Morse Avenue, Sunnyvale, CA 94086 (US). **HILLENMEYER, Maureen, Elizabeth;** 365 N California Avenue, Palo Alto, CA 94301 (US). **HORECKA, Joe;** 38448 Darnel Court, Fremont, CA 95536 (US).

(74) Agent: **THOMAS, Charles, A.;** KPPB LLP, 2190 S. Towne Centre Place, Suite 300, Anaheim, CA 92806 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available):* AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available):* ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ,

(54) Title: SYSTEMS AND METHODS FOR IDENTIFYING AND EXPRESSING GENE CLUSTERS

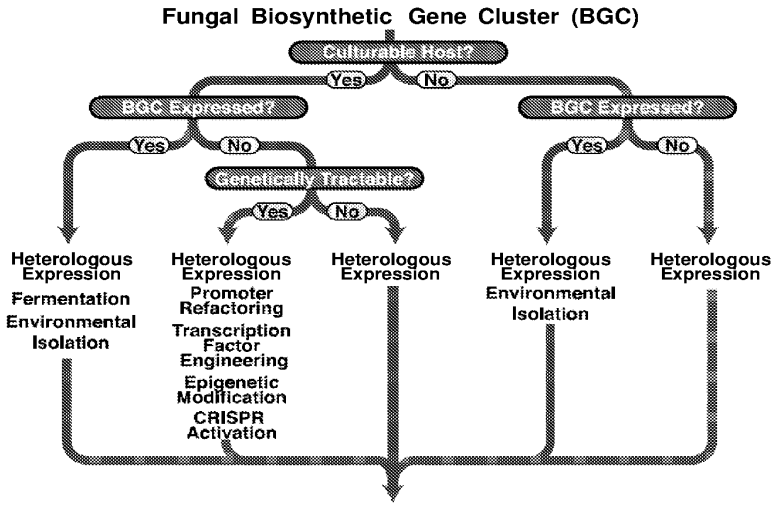


FIG. IA

(57) Abstract: Methods for identifying biosynthetic gene clusters that include genes for producing compounds that interact with specific target proteins are disclosed. Some methods relate to bioinformatics methods for identifying and/or prioritizing biosynthetic gene clusters. Related systems, components, and tools for the identification and expression of such gene clusters are also disclosed.

[Continued on next page]

WO 2018/094110 A2

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1006 of 3077

UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

— *with sequence listing part of description (Rule 5.2(a))*

SYSTEMS AND METHODS FOR IDENTIFYING AND EXPRESSING GENE CLUSTERS

CROSS-REFERENCE

[1] This application claims the benefit of U.S. Provisional Application No. 60/243,196, filed November 16, 2016, U.S. Provisional Application No. 62/481,601, filed April 4 2017, and U.S. Serial No. 15/469,452, filed March 24, 2017, which applications are incorporated herein by reference.

REFERENCE TO A SEQUENCE LISTING SUBMITTED ELECTRONICALLY VIA EFS-WEB

[2] The instant application contains a Sequence Listing which has been filed electronically in ASCII format and is hereby incorporated by reference in its entirety. Said ASCII copy, created on November 9, 2017, is named 52592-702_601_SL.txt and is 1,208,543 bytes in size.

STATEMENT AS TO FEDERALLY SPONSORED RESEARCH

[3] This invention was made with Government support under U01 GM1 10706 awarded by the National Institutes of Health. The government has certain rights to the invention.

TECHNICAL FIELD

[4] The present disclosure generally relates to the introduction of gene clusters into host organisms for the manufacture of small molecules. In some cases, the small molecules are analogs to products produced by the organism in which the gene cluster is identified and that modulate a specific protein. More specifically, the disclosure also relates to the identification of gene clusters likely to produce products that target a specific target protein and methods of expressing gene clusters in host cells. Additionally sequences of various gene clusters are provided, together with structures of compounds produced from these gene clusters.

BACKGROUND

[5] "Secondary metabolites," as used herein, are small molecules that can be produced by the expression of one or more gene clusters. Often, secondary metabolites are not critical for the survival of the organism in which the gene cluster is natively found. Secondary metabolites can be clinically valuable small molecules. Examples include: antibacterial products such as penicillin, and daptomycin; antifungal products such as amphotericin; cholesterol-lowering products such as

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1008 of 3077

lovastatin; anticancer products such as taxol, and eribulin; and immune-modulating products including rapamycin, and cyclosporine. In spite of the great success of secondary metabolites in the history of drug discovery, challenges of secondary metabolites in drug discovery and development can include (i) extremely low yields, (ii) limited supply, (iii) complex structures posing difficulty for structural modifications, and (iv) complex structures precluding practical synthesis. These difficulties have prompted the pharmaceutical industry to embrace new technologies in past decades, particularly combinatorial chemistry, as an alternative to natural product discovery. As a result, the percentage of new secondary metabolites being tested for use in medical treatment of humans has declined steadily since the 1940s due to a greater reliance on synthetic libraries that can be utilized in high throughput screening. Despite the pharmaceutical industry's preference for synthetic libraries, secondary metabolites possess enormous structural and chemical diversity that is unsurpassed by synthetic libraries. Most importantly, secondary metabolites are often evolutionarily optimized as drug-like molecules to target specific proteins and/or pathways.

[6] Now, thousands of bacterial and fungal genomes have been sequenced. These organisms are known to be rich sources of secondary metabolites. These secondary metabolites are enzymatically biosynthesized by the products of one or more genes, often grouped into gene clusters. New genome sequences have revealed that traditional approaches have tapped only a fraction of the biosynthetic potential of these organisms as, on average, fewer than 10% of the biosynthetic gene clusters (BGCs) in a microbial genome are expressed in any single culture condition. Further, millions of fungal species are believed to exist in nature, but have not been cultured in the laboratory. Accordingly, the supply bottleneck for secondary metabolites can be reduced by introducing the genes utilized in the synthesis of the secondary metabolite into a microorganism that can overproduce a desired secondary metabolite or analog thereof. In this way, the vast, untapped, ecological biodiversity of microbes holds renewed promise for the discovery of novel secondary-metabolites useful in one or more contexts, such as the treatment of disease.

SUMMARY OF THE INVENTION

[7] In some embodiments, this disclosure refers to a method for screening a plurality of compounds, the method comprising: identifying a gene cluster that includes or is within 20 kilobases of a region that encodes for a protein that is identical with or homologous to a first target protein,

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1009 of 3077

wherein the gene cluster comprises a region that encodes for a protein selected from the group consisting of (1) polyketide synthases and (2) non-ribosomal peptide synthetases, (3) terpene synthetases, (4) UbiA-type terpene cyclases, and (5) dimethylallyl transferases; introducing a plurality of genes from the gene cluster into a vector; introducing the vector into a host cell; expressing the proteins encoded by the plurality of genes in the host cell; and determining whether a compound that is formed or modified by the expressed proteins modulates the first target protein. In some cases, the host cell is a yeast cell. In some cases, the yeast cell is a yeast cell that has been modified to increased sporulation frequency and increased mitochondrial stability. In some cases, each gene of the plurality of genes is under the control of a different promoter. In some cases, the promoters are designed to increase expression when the host cell is in the presence a nonfermentable carbon source. In some cases, the plurality of genes are introduced into the vector via homologous recombination. In some cases, introducing the plurality of genes into the vector via homologous recombination comprises combining a first plurality of nucleotides with a second plurality of nucleotides, wherein: each polynucleotide of the first plurality of DNA polynucleotides encodes for a promoter and terminator, wherein each promoter and terminator is distinct from the promoter and terminator of other polynucleotides of the first plurality of nucleotides; and each polynucleotide of the second plurality of nucleotides includes a coding sequence, a first flanking region on the 5' side of the polynucleotide and a second flanking region on the 3' side of the polynucleotide; and introducing the polynucleotides into a host cell that includes machinery for homologous recombination, wherein the host cell assembles the expression vector via homologous recombination that occurs in the flanking regions of the second plurality of polynucleotides; wherein the expression vector is configured to facilitate simultaneous production of a plurality of proteins encoded by the second plurality of nucleotides. In some cases, the first flanking region and the second flanking region are each between 15 and 75 base pairs in length. In some cases, the first flanking region and the second flanking region are each between 40 and 60 base pairs in length. In some embodiments, the present disclosure provides a method for identifying a gene cluster capable of producing a small molecule for modulating a first target protein, the method comprising: selecting, from a database comprising a list of biosynthetic gene clusters, one or more gene clusters that include or are positioned proximal to a region that encodes a protein that is identical with or homologous to the first target protein. In some cases, the one or more gene clusters are selected

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1010 of 3077

from the group consisting of (1) clusters that comprise one or more polyketide synthases and (2) clusters that comprise one or more non-ribosomal peptide synthetases, (3) clusters that comprise one or more terpene synthases, (4) clusters that comprises one or more UbiA-type terpene cyclases, and (5) clusters that comprise one or more dimethylallyl transferases. In some cases, the protein that is encoded by the region that is included in or positioned proximal to the biosynthetic gene cluster is identical to or has greater than 30% homology to the first target protein. In some cases, the region that encodes the protein that is identical with or homologous to the first target protein is within 20,000 base pairs of a region of a portion of the gene cluster that encodes a polyketide synthase, a non-ribosomal peptide synthetase, a terpene synthetase, a UbiA-type terpene cyclase, or a dimethylallyl transferase. In some cases, the first target protein is BRSK1. In some cases, selecting the one or more gene clusters comprises operating a computer, wherein operation of the computer comprises running an algorithm that takes into account both an input sequence for the first target protein and sequence information from a database that includes sequence information from a plurality of species such that the computer returns information corresponding to one or more gene clusters. In some cases, the algorithm takes into account the phylogenetic relationship between gene clusters in the database. In some cases, the one or more gene clusters include a coding sequence for a protein that is an extracellular protein, a membrane-tethered protein, a protein involved in a transport or secretion pathway, a protein homologous to a protein involved in a transport or secretion pathway, a protein with a peptide targeting signal, a protein with a terminal sequence with homology to a targeting signal, an enzyme that degrades small molecules, or a protein with homology to an enzyme that degrades small molecules. In some embodiments the present disclosure provides a method for producing a compound that modulates the first target protein, the method comprising: identifying a gene cluster (e.g., via methods disclosed herein), expressing the gene cluster or a plurality of genes from the gene cluster in a host cell, and isolating a compound produced by the gene cluster. In some cases, the method further comprises screening the isolated compound for modulation of an activity of the first target protein. In some cases, the cluster-encoded protein that is homologous to the first target protein is resistant to modulation by the isolated compound when compared to modulation of the first target protein. In some cases, the compound is not toxic to the species from which the cluster originates due to one or more of (1) sequence differences between

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1011 of 3077

target protein and the cluster-encoded protein, (2) spatial separation of the compound from the cluster-encoded protein and (3) high expression levels for the cluster-encoded protein.

[8] In some embodiments the present disclosure provides a method for making a DNA vector, the method comprising: identifying a gene cluster comprising a plurality of genes that are capable of producing a small molecule for modulating a first target protein; introducing two or more genes of the plurality of genes into a vector, wherein the vector is configured to facilitate expression of the two or more genes in a host organism; wherein (1) the gene cluster encodes one or more proteins selected from the group consisting of a polyketide synthase, a non-ribosomal peptide synthetase, a terpene synthetase, a UbiA-type terpene cyclase, and a dimethylallyl transferase and (2) the gene cluster includes or is positioned proximal to a region that encodes a protein that is identical with or homologous to the first target protein. In some cases, the DNA vector is a circular plasmid. In some cases, the DNA vector comprises a plurality of promoters, wherein each promoter of the plurality of promoters is configured to, when the vector is introduced into a Saccharomyces cerevisiae cell, promote a lower level of heterologous expression when the cell exhibits predominantly anaerobic energy metabolism than when the cell exhibits aerobic energy metabolism. In some cases, each promoter of the plurality of promoters differs in sequence from one another. In some cases, each promoter of the plurality of promoters has a sequence selected from the group consisting of: SEQ ID NOs: 1-66. In some cases, each promoter of the plurality of promoters has a sequence selected from the group consisting of: SEQ ID NOs: 20-35, and SEQ ID NOs: 41-50. In some cases, when the *Saccharomyces cerevisiae* cell is exhibiting anaerobic energy metabolism, the cell is catabolizing a fermentable carbon source selected from glucose or dextrose; and when the *Saccharomyces cerevisiae* cell is exhibiting aerobic energy metabolism, the cell is catabolizing a non-fermentable carbon source selected from ethanol or glycerol.

[9] In some embodiments the present disclosure provides a method for the heterologous expression of a plurality of genes in a yeast strain, the method comprising: obtaining a yeast strain that includes a vector for expressing a plurality of genes from a single gene cluster of a non-yeast organism; and inducing expression of the plurality of genes; wherein the gene cluster in the non-yeast organism includes or is positioned proximal to a region that encodes a protein that is at least 30% homologous to a target protein. In some cases, the method further comprises introducing the plurality of genes from the single gene cluster into the vector. In some cases, the method further

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1012 of 3077

comprises introducing the vector into the yeast strain. In some cases, expression of the plurality of genes results in the formation of small molecule, wherein the small molecule modulates the activity of the target protein. In some cases, the gene cluster is a gene cluster of a non-yeast fungus. In some cases, the yeast strain is from *Saccharomyces cerevisiae*. In some cases, the target protein is a human protein.

[10] In some embodiments the present disclosure provides a system for identifying a gene cluster for introduction of a plurality of genes from the gene cluster into a host organism, the system comprising: a processor; a non-transitory computer-readable medium comprising instructions that, when executed by the processor, cause the processor to perform operations, the operations comprising: loading the identity or sequence of a first target protein into memory; loading the identity or sequence of a plurality of biosynthetic gene clusters into memory; identifying, from the plurality of biosynthetic gene clusters, one or more gene clusters that encode or are positioned proximal to a region that encodes a protein that is identical with or homologous to the first target protein; and scoring the one or more gene clusters based on the likelihood of each gene cluster being capable of use to produce a small molecule that modulates the first target protein. In some cases, scoring the one or more gene clusters comprises comparing the sequence of the first target protein (or a DNA sequence encoding the first target protein) to a sequence of a protein encoded in or proximal to the gene cluster (or to a DNA sequence encoding the protein that is in or proximal to the gene cluster).

[11] In some embodiments the present disclosure provides a system for identifying one or more biosynthetic gene clusters for introduction into a host organism to produce one or more compounds that modulate a specific target protein, the system comprising: a processor; a memory containing a gene cluster identification application; wherein the gene cluster identification application directs the processor to: load data describing at least one target protein into the memory; load data describing a plurality of biosynthetic gene clusters into the memory; score each of the plurality of biosynthetic gene clusters based upon: performing a homolog search for each biosynthetic gene cluster to determine a presence of at least one homolog of a target protein within or adjacent the biosynthetic gene cluster; confidence of homology of the at least one target protein to at least one gene in a biosynthetic gene cluster; a fraction of a homologous gene that meets an identity threshold; a total number of genes homologous to the at least one target protein present in the entire genome of an

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1013 of 3077

organism; homology of the at least one homolog of at least one target protein within or adjacent the biosynthetic gene cluster to genes in the target protein's genome; phylogenetic relationship of the at least one target protein to a gene in a cluster; expected number of homologs of the at least one target protein in or adjacent to a biosynthetic cluster; and a likelihood that at least one target protein is essential for cellular process in the natural environment; and output a report identifying one or more biosynthetic gene clusters that are most likely to produce a compound that modulates the at least one target protein.

[12] In some embodiments the present disclosure provides a method for selecting a biosynthetic gene cluster that produces a secondary metabolite, the method comprising: obtaining a list of gene clusters; performing a phylogenetic analysis of the genes within the clusters compared to known genes from known biosynthetic gene clusters; and selecting the biosynthetic gene cluster based on its phylogenetic relationship with the known genes. In some cases, the biosynthetic gene cluster with the most distant phylogenetic relationship from the known genes is selected.

[13] In some embodiments the present disclosure provides a method for identifying a gene cluster that produces a compound that binds a protein of interest, the method comprising: obtaining sequence information for a plurality of contiguous sequences, wherein each contiguous sequence includes a biosynthetic gene cluster and flanking genomic sequences; analyzing the contiguous sequences for the presence of a gene that encodes a protein with homology to the protein of interest, and selecting a biosynthetic gene cluster which includes, or is proximal to, a gene that encodes a protein that is homologous to the protein of interest. In some cases, the contiguous nucleotide sequence is less than 40,000 base pairs in length.

[14] In some embodiments, the present disclosure provides a modified yeast cell having a BY background, wherein relative to unmodified BY4741 and BY4742, the modified yeast cell has both (1) increased sporulation frequency and (2) increased mitochondrial stability. In some cases, the modified yeast cell grows faster on non-fermentable carbon sources than unmodified BY4741 and BY4742. In some cases, the yeast cell comprises one or more of the following genotypes: MKT1(30G), RME1(INS-308A), and TA03(1493Q). In some cases, the yeast cell comprises one or more of the following genotypes: CAT5(91M), MIP1(661T), SAL1+, and HAP1+. In some embodiments the present disclosure provides a method of forming an expression vector, the method comprising: combining a first plurality of DNA polynucleotides with a second plurality of

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1014 of 3077

polynucleotides, wherein: each polynucleotide of the first plurality of DNA polynucleotides encodes for a promoter and terminator, wherein each promoter and terminator is distinct from the promoter and terminator of other polynucleotides of the first plurality of nucleotides; and each polynucleotide of the second plurality of nucleotides includes a coding sequence, a first flanking region on the 5' side of the polynucleotide and a second flanking region on the 3' side of the polynucleotide, wherein each flanking region is between 15 and 75 base pairs in length; and introducing the polynucleotides into a host cell that includes machinery for homologous recombination, wherein the host cell assembles the expression vector via homologous recombination that occurs in the flanking regions of the second plurality of polynucleotides; wherein the expression vector is configured to facilitate simultaneously production of a plurality of proteins encoded by the second plurality of nucleotides. In some cases, the host cell is a yeast cell. In some cases, each flanking region is between 40 and 60 base pairs in length. In some cases, at least one polynucleotide of the first plurality of nucleotides encodes a selection marker.

[15] In some embodiments, the present disclosure provides a system for generating a synthetic gene cluster via homologous recombination, the system comprising 1 though N unique promoter sequences, 1 through N unique terminator sequences, and 1 through N unique coding sequences, wherein: coding sequence 1 is attached to an additional 30-70 base pair sequence on each end such that a first end portion is identical to the last 30-70 base pairs of promoter 1 and a second end portion is identical to the first 30-70 base pairs of terminator 1; coding sequence 2 is attached to an additional 30-70 base pair sequence on each end such that a first end portion is identical to the last 30-70 base pairs of promoter 2 and a second end portion is identical to the first 30-70 base pairs of terminator 2; and coding sequence N is attached to an additional 30-70 base pair sequence on each end such that a first end portion is identical to the last 30-70 base pairs of promoter N and a second end portion is identical to the first 30-70 base pairs of terminator N. In some cases, terminator 1 and promoter 2 are portions of the same double-stranded oligonucleotide.

[16] In some embodiments, the present disclosure provides a method for assembling a synthetic gene cluster, the method comprising: obtaining 1 through N unique promoters, 1 through N unique terminators, and 1 through N unique coding sequences, wherein: coding sequence 1 is attached to an additional 30-70 base pair sequence on each end such that a first end portion is identical to the last 30-70 base pairs of promoter 1 and a second end portion is identical to the first 30-70 base pairs of

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1015 of 3077

terminator 1; coding sequence 2 is attached to an additional 30-70 base pair sequence on each end such that a first end portion is identical to the last 30-70 base pairs of promoter 2 and a second end portion is identical to the first 30-70 base pairs of terminator 2; and coding sequence N is attached to an additional 30-70 base pair sequence on each end such that a first end portion is identical to the last 30-70 base pairs of promoter N and a second end portion is identical to the first 30-70 base pairs of terminator N; transforming the 1 through N promoters, terminators and coding sequences into a yeast cell; isolating a plasmid containing the 1 through N promoters, terminators and coding sequences from the yeast cell. In some cases, the method further comprises a coding sequence for a selection marker. In some cases, the selection marker is an auxotrophic marker. In some cases, the yeast cell has a deficiency in a DNA ligase gene.

[17] In some embodiments the present disclosure provides a yeast strain which allows for both (1) homologous DNA assembly and (2) production of heterologous genes in the same strain. In some cases, the yeast strain is a DHY strain. In some cases, the strain allows DNA assembly via homologous recombination with an efficiency of at least 80% as compared to DNA assembly in BY. In some cases, production of heterologous compounds in the strain is accomplished with an efficiency of at least 80% as compared to heterologous compound production in BJ5464. In some cases, the strain allows production of heterologous proteins with an efficiency of at least 80% as compared to heterologous protein production in BJ5464.

[18] In some embodiments the present disclosure provides a method for isolating a plasmid from a yeast cell, the method comprising: isolating total DNA from a yeast cell that includes a plasmid; incubating the DNA with an exonuclease such that the exonuclease degrades substantially all of the linear DNA in the isolated total DNA from the yeast cell; optionally inactivating the exonuclease; and recovering the plasmid DNA. In some cases, the isolated plasmid DNA is of sufficient purity for use in a sequencing reaction. In some cases, the plasmid DNA is further prepared for a sequencing reaction.

[19] In some embodiments, the present disclosure provides pharmaceutical composition comprising Compound 6 and a pharmaceutically acceptable excipient. In some embodiments, the present disclosure provides pharmaceutical composition comprising Compound 7 and a pharmaceutically acceptable excipient. In some embodiments, the present disclosure provides pharmaceutical composition comprising Compound 8 and a pharmaceutically acceptable excipient.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1016 of 3077

In some embodiments, the present disclosure provides pharmaceutical composition comprising Compound 9 and a pharmaceutically acceptable excipient. In some embodiments, the present disclosure provides pharmaceutical composition comprising Compound 10 and a pharmaceutically acceptable excipient. In some embodiments, the present disclosure provides pharmaceutical composition comprising Compound 11 and a pharmaceutically acceptable excipient. In some embodiments, the present disclosure provides pharmaceutical composition comprising Compound 12 and a pharmaceutically acceptable excipient. In some embodiments, the present disclosure provides pharmaceutical composition comprising Compound 13 and a pharmaceutically acceptable excipient. In some embodiments, the present disclosure provides pharmaceutical composition comprising Compound 14 and a pharmaceutically acceptable excipient. In some embodiments, the present disclosure provides pharmaceutical composition comprising Compound 15 and a pharmaceutically acceptable excipient. In some embodiments, the present disclosure provides pharmaceutical composition comprising Compound 16 and a pharmaceutically acceptable excipient. In some embodiments, the present disclosure provides method of producing Compound 3, the method comprising: providing a vector or vectors comprising the coding sequences of **SEQ ID NOs: 200-206;** transforming a host cell with the vector or vectors; incubating the host cell in culture media under conditions suitable for the expression of the coding sequences; and isolating the compound produced by the host cell.

INCORPORATION BY REFERENCE

[20] All publications, patents, and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication, patent, or patent application was specifically and individually indicated to be incorporated by reference.

BRIEF DESCRIPTION OF THE DRAWINGS

[21] The written disclosure herein describes illustrative embodiments that are non-limiting and non-exhaustive. Reference is made to certain of such illustrative embodiments that are depicted in the figures, in which:

[22] **Figure 1A** illustrates strategies that may be used to obtain a secondary metabolite from a fungal strain given different properties of the fungal strain.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1017 of 3077

[23] **Figure IB** illustrates examples of characterized gene clusters which produce known chemicals and a novel gene cluster with a product.

[24] **Figure 2** illustrates a phylogenetic analysis of enzymes that produce secondary metabolites.

[25] **Figure 3** illustrates self-resistance mechanisms for potentially toxic secondary metabolites.

[26] **Figure 4** illustrates a gene cluster that produces lovastatin.

[27] **Figure 5A** illustrates an exemplary process to extract and utilize compound products in accordance with an embodiment of the invention.

[28] **Figure 5B** illustrates an exemplary process to produce and extract heterologous, biosynthetic compound products in accordance with an embodiment of the invention.

[29] **Figure 5C** illustrates an example production pipeline for producing secondary metabolites from gene clusters.

[30] **Figure 5D** illustrates an example work flow for producing secondary metabolites from gene clusters.

[31] **Figure 6A** illustrates a yeast phase chart displaying yeast cell concentration in relation to time to provide reference for various embodiments of the disclosure.

[32] **Figure 6B** illustrates a yeast phase chart displaying glucose or dextrose concentration in relation to time to provide reference for various embodiments of the disclosure.

[33] **Figure 6C** illustrates a yeast phase chart displaying ethanol or glycerol concentration in relation to time to provide reference for various embodiments of the disclosure.

[34] **Figure 7A** illustrates a DNA vector having a production-phase promoter in accordance with an embodiment of the disclosure.

[35] **Figure 7B** illustrates a DNA vector having multiple production-phase promoters in accordance with an embodiment of the disclosure.

[36] **Figure 8A** illustrates a DNA expression vector having a production-phase promoter within an expression cassette in accordance with an embodiment of the disclosure.

[37] **Figure 8B** illustrates a DNA expression vector having multiple production-phase promoters, each within an expression cassette in accordance with an embodiment of the disclosure.

[38] **Figure 9** illustrates a method to construct and utilize production-phase promoter DNA vectors in accordance with various embodiments of the disclosure.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1018 of 3077

[39] **Figure 10A** illustrates an overview of an approach involving yeast homologous recombination assembly.

[40] **Figure 10B** illustrates the homologous recombination in yeast of the parts from FIG. 10A.

[41] **Figure IOC** illustrates the plasmid which results from the parts of **FIG. 10A,** homologously recombined as in **FIG. 10B.**

[42] **Figure 10D** illustrates improved sequencing results obtained via disclosed methods.

[43] **Figure 10E** illustrates assembly of plasmid DNA from up to 14 individual fragments.

[44] **Figure 11A** illustrates improved assembly efficiency in a background lacking the DNL4 ligase.

[45] **Figure 11B** illustrates equivalent sequencing efficiencies using DNA prepared from both colonies (red) and liquid cultures (blue) for four test assemblies.

[46] **Figure 11C** illustrates sequencing efficiencies observed using both standard and modified NexteraXT library preparation methods.

[47] **Figure 11D** illustrates a workflow for sequencing plasmids from yeast via a step of transforming the plasmids into E. coli.

[48] **Figure H E** illustrates a workflow for sequencing plasmids from yeast in accordance with methods described herein.

[49] **Figure 12** illustrates repaired SNPs in yeast strain DHY674 relative to BY4741 .

[50] **Figure 13** illustrates DHY213, BJ5464, and X303 (a W303 derivative) grown on glucose (fermentation) and ethanol/glycerol (respiration) media.

[51] **Figure 14A** illustrates relative growth rates of strains described here in YPD culture. The dotted line denotes the diauxic shift (the point at which the culture exhausts all glucose and transitions from fermentation to respiration). The DHY derived strain JHY692 shows significantly improved growth in the respiration phase of the culture.

[52] **Figure 14B** illustrates expression of eGFP driven by the PADH2 promoter in the strains from **Figure 14A.**

[53] **Figure 14C** illustrates expression of eGFP driven by the PPCKl promoter in the strains from **Figure 14A.**

[54] **Figure 15A** illustrates an example gene cluster.

[55] **Figure 15B** illustrates a polyketide produced by a 5-gene gene cluster.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1019 of 3077

[56] **Figure 16** is a heat map graphic generated in accordance with various embodiments of the disclosure with data of expression of enhanced-green fluorescent protein driven by various *S. cerevisiae* promoters.

[57] **Figure 17** is a data graph of enhanced-green fluorescent protein expression driven by various *S. cerevisiae* promoters.

[58] **Figure 18** illustrates fluorescence intensity of 10^5 cells expressing enhanced-green fluorescent protein driven by various promoters.

[59] **Figure 19** illustrates a phylogenetic tree of *Saccharomyces sensu stricto* subgenus.

[60] **Figure 20** illustrates a multiple sequence alignment of various *Saccharomyces sensu stricto* species' upstream activating sequences in ADH2 promoters.

[61] **Figure 21** illustrates homology between various *Saccharomyces sensu stricto* species' ADH2 promoters.

[62] **Figure 22** is a heat map graphic generated in accordance with various embodiments of the disclosure with data of expression of enhanced-green fluorescent protein driven by various *S. sensu stricto* ADH2 promoters.

[63] **Figure 23** is a data graph of enhanced-green fluorescent protein expression driven by various *S. sensu stricto* ADH2 promoters.

[64] **Figure 24** illustrates four multi-gene expression vector constructs and a data graph of the resultant compound production, in accordance with an embodiment of the invention.

[65] **Figure 25** illustrates a biosynthetic process that produces the compound emindole SB via a fungal four-gene cluster.

[66] **Figure 26** is a data graph of the production results of two product compounds generated.

[67] **Figure 27A** illustrates two plasmid vector constructs in accordance with an embodiment of the disclosure.

[68] **Figure 27B** illustrates a further vector construct in a yeast cell in accordance with an embodiment of the disclosure.

[69] **Figure 28A** illustrates a phylogenetic analysis of further gene clusters. Abbreviations used may include Adenylation domain (A), a,b-hydrolase (a,b-h), ATP-binding cassette transporter (ABC), Acyl carrier protein (ACP), Alcohol dehydrogenase (ADH), Aldo-keto reductase (AK-red), Aminooxidase (AmOx), Aminotransferase (AmT), Arylsulfotransferase (ArST), Acyltransferase

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1020 of 3077

domain (AT), C-mehtyltransferase (cMT), Terpene cyclase (Cyc), Dehydratase (DH), Domain of unknown function 4246 (DUF4246), Flavin adenine dinucleotide (FAD) binding protein (FAD), Iron dependent alcohol dehydrogenase (Fe-ADH), Flavin-dependent monooxygenase (FMO), Geranyl-geranyl pyrophosphate synthase (GGPPS), Glycosidase (glycos.), Glucose-methanol-choline oxidoreductase (GMC), Halogenase (Halo), Highly reducing polyketide synthase (HR-PKS), Hypothetical protein (Hyp), Indole-3-acetic acid-amido synthetase (IAS), Ketosynthase domain (KS), metallo-B-lactamase (mBla), Mitochodrial phosphate carrier protein (MCP), Major facilitator superfamily transporter (MFS), Metallohydrolase (MH), Methyl transferase (MT), Nicotine adenine dinucleotide dependent dehydrogenase (NAD-DH), Nicotine adenine dinucleotide phosphate (NADP) dependent reductase (NADP-R), N-mehtyltransferase (nMT), Non reducing polyketide synthase (NR-PKS), O-succinylhomoserine sulfhydrylase (O-suc-SH), O-methyltransferase (oMT), Oxidoreductase (OxR), Cytochrome p450 (p450), Dipeptidyl peptidase (Pep), Prenyl transferase (PrT), Product template domain (PT), Riboflavin biosynthesis protein RibD (RibD), RNA helicase (RNAh), starter unit: ACP transacylase domain (SAT), Short-chain dehydrogenase (SDH), Short-chain dehydrogenase/reductase (SDR), Serine hydrolase (SH), Sugar transport protein (ST), Thiolation domain (T), Terminal domain (TD), Thioesterase domain (TE), Transcription factor (TF), and UbiA-type terpene cyclase (UTC).

[70] **Figure 28B** illustrates various gene clusters and biosynthetic compound products in accordance with various embodiments of the invention.

[71] **Figure 28C** illustrates various gene clusters and biosynthetic compound products in accordance with various embodiments of the invention.

[72] **Figure 28D** illustrates various gene clusters and biosynthetic compound products in accordance with various embodiments of the invention.

[73] **Figure 28E** illustrates various gene clusters and biosynthetic compound products in accordance with various embodiments of the invention.

[74] **Figure 28F** illustrates various gene clusters and biosynthetic compound products in accordance with various embodiments of the invention.

[75] **Figure 29A** illustrates a phylogenetic analysis of gene clusters.

[76] **Figure 29B** illustrates correction of a gene cluster.

[77] **Figure 30A** illustrates schematics of PKS enzyme containing BGCs examined herein.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1021 of 3077

[78] **Figure 30B** illustrates schematics of UTC containing BGCs examined herein.

[79] **Figure 31** illustrates a volcano plot of all spectral features identified in the automated analysis of strains expressing PKS containing BGCs. All features determined to be specific to the BGC expressing strain were identified by comparison to a negative vector control.

[80] **Figure 32** illustrates features produced by BGC PKS1.

[81] **Figure 33** illustrates features produced by BGC PKS2.

[82] **Figure 34** illustrates features produced by BGC PKS4.

[83] **Figure 35** illustrates features produced by BGC PKS6.

[84] **Figure 36** illustrates features produced by BGC PKS8.

[85] **Figure 37** illustrates features produced by BGC PKS 10.

[86] **Figure 38** illustrates features produced by BGC PKS 13.

[87] **Figure 39** illustrates features produced by BGC PKS 14.

[88] **Figure 40** illustrates features produced by BGC PKS 15.

[89] **Figure 41** illustrates features produced by BGC PKS 16.

[90] **Figure 42** illustrates features produced by BGC PKS 17.

[91] **Figure 43** illustrates features produced by BGC PKS 18.

[92] **Figure 44** illustrates features produced by BGC PKS20.

[93] **Figure 45** illustrates features produced by BGC PKS22.

[94] **Figure 46** illustrates features produced by BGC PKS23.

[95] **Figure 47** illustrates features produced by BGC PKS24.

[96] **Figure 48** illustrates features produced by BGC PKS28.

[97] **Figure 49** illustrates NMR data and structure of Compound 6.

[98] **Figure 50** illustrates NMR data and structure of Compound 7.

[99] **Figure 51** illustrates NMR data and structure of Compound 8.

[100] **Figure 52** illustrates NMR data and structure of Compound 9.

[101] **Figure 53** illustrates NMR data and structure of Compound 10.

[102] **Figure 54** illustrates NMR data and structure of Compound 11.

[103] **Figure 55** illustrates NMR data and structure of Compound 12.

[104] **Figure 56** illustrates NMR data and structure of Compound 13.

[105] **Figure 57** illustrates NMR data and structure of Compound 14.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1022 of 3077

[106] **Figure 58** illustrates NMR data and structure of Compound 15.

[107] **Figure 59** illustrates NMR data and structure of Compound 16.

DETAILED DESCRIPTION

[108] Systems and methods in accordance with various embodiments of the disclosure identify, refactor, and express biosynthetic gene clusters in host organisms to produce secondary metabolites. Systems and methods in accordance with various embodiments of the disclosure utilize host organisms to produce secondary metabolites. Using host organisms allows for production of secondary metabolites from biosynthetic gene clusters regardless of whether the native host cell can be cultured, or whether the cluster is expressed in the native host, see **Figure 1A.** In some cases, the secondary metabolites may bind to one or more specific proteins. In some embodiments, the host organism ordinarily does not produce the secondary metabolite. The host organism obtains the ability to produce the secondary metabolite due to the introduction of a biosynthetic gene cluster identified using a cluster identification process performed in accordance with an embodiment of the disclosure. In one embodiment, a cluster identification process identifies a biosynthetic gene cluster that possesses characteristics suggesting that it is responsible for producing a secondary metabolite with novel chemistry. In another embodiment, a cluster identification process (discussed further below) identifies a biosynthetic gene cluster that possesses characteristics suggesting that it is responsible for producing a secondary metabolite that binds to a specific protein of interest. In some embodiments this disclosure provides the inclusion of a biosynthetic gene cluster identified using a cluster identification process, in accordance with various embodiments of the disclosure, within a host organism enabling the host organism to express a secondary metabolite. In some cases, the secondary metabolite produced in the host cell may be identical to a secondary metabolite that is naturally produced by the organism in which the biosynthetic gene cluster was originally identified. In some cases, the secondary metabolite produced by the host cell is an analog of a secondary metabolite produced by the organism from which the cluster was identified. In other cases, the secondary metabolite produced in the host cell may be structurally distinct from the secondary metabolite produced by the cluster in the originating species. Differences in the product produced may arise from differences in expression and localization of the coding sequences from the cluster.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1023 of 3077

Additionally coding sequences, or expression products thereof, from the cluster may interact with other coding sequences, or expression products thereof, not contained in the cluster. Host cell produced secondary metabolites which are distinct from those produced in the originating species of the cluster may be termed non-naturally occurring secondary metabolites or non-natural secondary metabolites. The secondary metabolite produced by the host organism can be isolated and then can be used, for example, in a treatment. In some cases, the secondary metabolites may be used in a treatment of a disease or disorder which involves aberrant activity of a specific protein. This disclosure also describes a panel of sesquiterpenoid and polyketide products.

Cluster identification

[109] Cluster identification processes, in accordance with many embodiments of the disclosure, utilize specific properties of a biosynthetic gene cluster to identify gene clusters of interest from sequence data. Sequence data used with the methods of this disclosure may comprise genomic sequence data, transcriptome sequence data, or other sequence data. In some cases, sequence data may be generated by sequencing a DNA sample obtained from an environmental sample. In other cases, sequence data may be obtained from publicaliy available genome sequence libraries, or may be purchased. Genome sequences may be derived from any organism. In some cases, genome sequences may be derived from a fungal, bacterial, archaeal or plant species.

[110] In some cases, the genome sequences may be derived from fungi. In some cases, the genome sequences may be derived from fungi which are poorly characterized or difficult to culture. In some cases, the genome sequences may be derived from fungi which are well characterized, or partially characteri zed. Examples of types of fungi from which sequences may be deri ved include fungi from one of the following Phyla: Basidiomycota, Ascomycota, Neocallimastigomycota, Blastocladiomycota, Glomeromycota, Chytridiomycota and Microsporidia. Examples of fungal species which may be the source of sequence data include, but are not limited to: *Aspergillus tubingensis, Hypomyces subiculosus, Coniothyrium sporulosum, Acremonium Sp. KY4917, Aspergillus niger, Thielavia terrestris, Trichoderma virens, Pseudogymnoascus pannorum, Scedosporium apiospermum, Metarhizium anisopliae, Cochliobolus heterostrophus, Verruconis gallopava, Moniliophthora roreri, Punctularia strigosozonata, Hydnomerulius pinastri, Arthroderma gypseum, Setosphaeria turcica, Pyrenophora teres, Cladophialophora yegresit,*

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1024 of 3077

Talaromyces cellulolyticus, Endocarpon pusillum, Hypholoma sublateritium, Ceriporiopsis subvermispora, Botryotonia cinerea, Formitiporia mediterranea, Heterobasidion annosum, Ge/atoporia subvermispora, Dichomitus squalens, Pleurotus ostreatus, Schizophyllum commune, Stereum hirsutum, Sternum hirsutum, and *Dacryopinax primogenitus.*

[Ill] Provided in **Figure IB** is a flow chart showing a process embodiment that can be implemented using computer systems for identifying and ranking BSGs that are likely to produce secondary metabolites capable for anthropogenic use (e.g., medicinal). As shown, Process 1000 can begin by obtaining genetic sequence data from a biological source or sequence database (1001). In some cases, the sequence data may be derived from single cell sequencing of a fungal cell of unknown species. For example an environmental sample may contain many different cells which may be separated and sequenced. In some cases, the sequence data is obtained from a publically available genetic data library, or may be purchased. Often the genetic sequence data is a genomic sequence of an organism, however, any genetic data sequence, including partial genome sequence data, may be used.

[112] Process 1000 also identifies biosynthetic gene clusters (1003). Clusters may be identified using bioinformatics methods to scan genome sequences. Characteristics of a gene cluster may include a grouping of two or more genes within about 5 kb, 10 kb, 15 kb, 20 kb, 25 kb, 30 kb, 35 kb, 40kb or 45 kb of each other. Genes may be bioinformatically identified by the presence of known promoter sequences, transcription initiation sequences, or homology to known genes or gene features. The term homology as used herein refers to sequences with high sequence identity, for example a sequence identity of at least about 50%, 60%, 70%, 80%, 90%, 95%, 97%, 98%, 99% or more than 99%. Sequence identity may be determined using alignment tools such as Basic Local Alignment Search Tool (BLAST), available via the National Center for Biotechnology Information (NCBI), to determine areas of conserved sequence. Biosynthetic gene clusters may be identified using a bioinformatics tool, such as ClustScan, SMURF, CLUSEAN, and/or antiSMASH. Biosynthetic gene clusters typically contain one or more core biosynthetic genes and one or more tailoring genes. Clusters which contain the same, or highly similar, core enzymes with different tailoring genes may produce quite different compounds, as seen in **Figure 1C.** Expressing a subset of genes from a cluster may also result in production of a different compound compared with the compound produced by expression of all the genes in the cluster.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1025 of 3077

[113] Process 1000 also scores and/or ranks gene clusters utilizing various factors (1005). In some cases, scores and rankings are based on the type of secondary metabolite to be produced. In other cases, scores and rankings are based on a protein or domain existing within the cluster. In even more cases, the level of homology of a particular protein within the cluster to a protein of interest is considered. It should be understood that many different factors can be used, as determined by the application and use of the biosynthetic gene cluster data.

[114] An embodiment of a process for identifying biosynthetic gene clusters using computer systems is provided in **Figure** ID. Process 2000 can begin with obtaining genetic sequence data of an organism having gene clusters (2001). In many cases, the genetic sequence data is a genomic sequence of an organism. In other cases, the genetic data sequence is a partial genome sequence data. In addition to genetic sequence data, Process 2000 also obtains target sequence data keying to biosynthetic gene clusters of interest (2003). The target sequence is any sequence the user wishes to define and identify clusters. In some cases, the target sequence is a particular protein domain of interest. In other cases, the target sequence is a particular protein, protein homolog or protein class.

[115] In some embodiments, clusters are scanned for the presence of genes encoding proteins known to be involved in biosynthetic pathways. Key proteins involved in biosynthetic pathways include terpene synthases, polyketide synthases (PKSs, both highly-reducing and non-reducing), non-ribosomal peptide synthetases, UbiA-type terpene cyclases (UTCs), polyketide synthase non-ribosomal peptide synthetase hybrids, and dimethyl allyl transferases (see **Figure 2).**

[116] Process 2000 also identifies the target sequence within genetic sequence data using homologous alignment scores (2005). Using an appropriate application, the sequence of the target is used to align to the genetic sequence data, looking for a threshold of homology. In some cases, a positive homologous event occurs when the target sequences aligns with the a portion of the genetic sequence having homology of at least about 50%, 60%, 70%, 80%, 90%, 95%, 97%, 98%, 99% or more than 99%. Sequence homology may be determined using any alignment tool, such as, for example BLAST.

[117] Using the homologous alignment scores, candidate biosynthetic gene clusters may be identified in the region surrounding the homologous event (2007). In many cases, the proximal upstream and downstream genes are examined to determine and define a gene cluster. In some of these cases, 5, 6, 7, 8, 9, 10 or more proximal genes in either direction are examined. In addition or

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1026 of 3077

alternatively, a gene cluster may be defined by genetic distance of the homologous event. Gene clusters may also be defined using a bioinformatics tool, such as ClustScan, SMURF, CLUSEAN, and/or antiSMASH. Once identified and defined, clusters may be stored as data and/or reported via an output interface (2009).

[118] An embodiment for ranking biosynthetic gene clusters using computer systems is provided in **Figure IE.** As shown, Process 3000 can begin with obtaining genet sequence data of multiple biosynthetic gene clusters. In this process, the clusters obtained each have a homolog protein of interest. The homolog of interest depends on the user's desired result. In many cases, the homolog of interest has a human ortholog that is known to be involved in a human condition, disorder, or disease. In some of these cases, the human ortholog is known to have mutations, either congenital or somatic, that lead to a condition, disorder, or disease. In some other of these cases, the human ortholog is involved in biological pathways that are involved in a condition, disorder, or disease. In other cases, the homolog of interest has an ortholog in an infectious species, including, but not limited to bacterial, fungal, protozoan species. In many of these cases, the ortholog in the infectious species is essential to the vitality of the organisms of the species. In some other of these cases, the ortholog in the infectious species is involved in producing a toxin. Furthermore, in many cases, the user has a desired result to prioritize clusters that will produce a secondary metabolite that may target the of human or infectious species ortholog.

[119] Identifying a gene cluster that may produce a secondary metabolite that binds a target protein may involve multiple different steps. In some cases, such a gene cluster may be identified by the presence of one or more genes encoding a homolog of the target protein, within or adjacent to the cluster, as determined by a homology search (for example, using the tblastn algorithm, with a maximum score granted when one homolog is found). In some cases, a gene cluster may be identified by the confidence in homology of the target to a gene or genes in a cluster. For example, according to the tblastn algorithm, gene clusters containing a gene with an e value less than about 10^{-10}, 10^{-20}, 10^{-30}, 10^{-35} or 10^{-40} may be selected. Gene clusters may also be selected using a protein blast method such as blastp to compare a predicted protein sequence against either known protein sequence or other predicted protein sequence. For blast searches using a known or predicted protein sequence gene clusters containing a gene with an e value less than about 10^{-10}, 10^{-20}, 10^{-30}, 10^{-35} or 10^{-40} may be selected. Selected gene clusters may be prioritized with increasing priority scores for

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1027 of 3077

lower e-values. In some cases, a gene cluster may be identified by the fraction of the homologous gene that meets a certain threshold of identity (for example, with an increasing score for more identity, and a lower bound threshold of at least about 99%, 95%, 80%, 85%, 75%, 70%, 65%, 60%, 50%, 45%, 40%, 35%, 30%, 25%, 20% or 15% coverage). In some cases, a gene cluster may be selected if it contains a gene which produces a protein with at least about 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97%, 98%, 99% or 100% homology to the target protein. In some cases, a protein that is encoded by the region that is included in, or positioned proximal to, a BGC may be identical to, or have at least about 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 99% or about 100% homology to a target protein. In some cases, a gene cluster may be identified by the total number of genes homologous to the target protein present in the entire genome of the organism (for example, with a maximum score granted to cases with 2-4 homologs per genome). In some cases, a gene cluster may be identified by the homology of the gene in, or adjacent to, the cluster to the target protein (for example, using the blastx algorithm, with a maximum score granted when the gene in the biosynthetic gene cluster's closest homolog in the target protein's genome is the target protein itself). In some cases, a gene cluster may be identified by the phylogenetic relationship of the target protein to the gene in the cluster (for example, with an increasing score for homologs in the gene cluster that clade with the target protein, with confidence assigned by a bootstrap test or Bayesian inference of phylogeny, and a lower bound threshold defined as homologs in a phylogenetic context that appear in a clade with bootstrap value of 0.7 or Bayesian posterior probability of 0.8). In some cases, a gene cluster may be identified by the expected number of homologs of the target in or adjacent to the biosynthetic cluster (for example, with a greater score the lower the probability of a homolog of the target being present in or adjacent to a biosynthetic cluster of a certain size, given the number of total homologs in the genome, as determined by a permutation test). In some cases, a gene cluster may be identified by the likelihood that the target protein is essential for viability, growth, or other cellular processes in the native environment (for example, through evidence that deletion of homologs in related organisms (such as *S. cerevisiae)* render the organism inviable). In some cases, a gene cluster may be identified by one or more of the above methods, or by any two or more of the above methods.

[120] Utilizing identification steps, Process 3000 also scores each biosynthetic gene cluster based on factors indicative of secondary metabolite synthesis related to the homolog of interest (3003).

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1028 of 3077

Accordingly, a score is constructed for each biosynthetic cluster based on one or more of the following:

(a) the presence of one or more homologs of target orthologs within or adjacent to the cluster, as determined by a homology search (e.g., the BLAST algorithm);

(b) the degree of homology of one or more target orthologs to genes in a cluster (e.g., as defined by the e-value);

(c) the fraction of the homologous gene that meets a certain threshold (e.g., the number of homologous protein domains);

(d) the total number of genes homologous to the target ortholog present in the entire genome of the organism;

(e) the degree of homology of a gene in or adjacent to the cluster to the target ortholog

(f) the number of homologs in the gene family (e.g., the number of homologs of a human target gene in the human genome); and/or

(g) the expected number of homologs of the target ortholog in, or neighboring, a biosynthetic cluster (e.g., the probability of a homolog of the target being present in or adjacent to a biosynthetic cluster of a certain size, given the number of total homologs in the genome, and as determined from a permutation test)

(h) the synteny of the gene cluster with related species (e.g., conservation of gene cluster)

(i) the function class of the target ortholog

(j) the presence of specific promoters adjacent to the homolog(s) within the gene cluster (e.g., identification of bidirection promoter upstream the homolog and biosynthetic gene)

(k) the presence of specific regulatory elements in the biosynthetic cluster (e.g., the number of transcription factor binding sites shared among target orthologs and homologs/biosynthetic genes in the cluster)

(l) the presence of homologs outside the cluster that are co-regulated with some or all the genes within the biosynthetic cluster

(m) the presence of protein- and DNA-sequence derived features within the clusters that have successfully been shown to produce secondary metabolites. It should be understood that a particular user may desire to use one, some, or all the factors listed here, and/or other factors not listed. The factors utilized depend on the user's application and desired result.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1029 of 3077

[121] Process 3000 also has the option to calibrate the score to by referencing a set of "true positives" (i.e., cases where there is one or more known targets in or adjacent to a biosynthetic cluster that produces a small molecule targeting the target, such as the lovastatin BGC) (3005). The output of this process may be a ranked and/or scored list of biosynthetic clusters, which may be used to identify the clusters that will produce therapeutic small molecules targeting the products of a disease-related gene (3007). The ranked and/or scored list of clusters can be stored as data or reported via an output interface (3009).

[122] Turning now to **Figure IF,** computer systems (4001) may be implemented on a single or multiple computing devices in accordance with some embodiments of the invention. Computer systems (4001) may be personal computers, laptop computers, and/or any other computing devices with sufficient processing power for the processes described herein. The computer systems (401) include a processor (403), which may refer to one or more devices within the computing devices that can be configured to perform computations via machine readable instructions stored within a memory (4007) of the computer systems (4001). The processor may include one or more microprocessors (CPUs), one or more graphics processing units (GPUs), and/or one or more digital signal processors (DSPs). According to other embodiments of the invention, the computer system may be implemented on multiple computers.

[123] In a number of embodiments of the invention, the memory (4007) may contain a gene cluster identification and/ scoring application (4009) that performs all or a portion of various methods according to different embodiments of the invention described throughout the present application. As an example, processor (4003) may perform a gene cluster identification and/or scoring method similar to any of the processes described above with reference to Figures ID and IE, during which memory (4007) may be used to store various intermediate processing data such as the genetic sequence alignment data (e.g., BLASTn) (4009a), identification of key target sequences with gene clusters (4009b), identification of homolog(s) to a protein of interest (4009c), characterization of homologs (4009d), scores and/or ranks of gene clusters (4009e), and calibration of gene cluster scores (4009f).

[124] In some embodiments of the invention, computer systems (4001) may include an input/output interface (4005) that can be utilized to communicate with a variety of devices, including but not limited to other computing systems, a projector, and/or other display devices. As can be

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1030 of 3077

readily appreciated, a variety of software architectures can be utilized to implement a computer system as appropriate to the requirements of specific applications in accordance with various embodiments of the invention.

[125] Although computer systems and processes for chimeric sequence unveiling and performing actions based thereon are described above with respect to Figure IF, any of a variety of devices and processes for data associated with cluster identification and/or scoring as appropriate to the requirements of a specific application can be utilized in accordance with many embodiments of the invention.

[126] In some embodiments, gene sequences within a novel cluster, or a set of novel clusters, may be compared to gene sequences from known and characterized biosynthetic gene clusters. In some cases, phylogenetic comparisons may be carried out between gene sequences in a novel cluster and gene sequences from characterized gene clusters. As shown in **Figure 2,** of the many biosynthetic enzymes which have been identified from sequence data, only a small fraction have been characterized, suggesting potential for many novel chemistries. Phylogenetic analysis may be performed on the core biosynthetic gene or genes, or on one or more tailoring genes of the novel cluster. Preferred clusters may be clusters containing one or more gene sequences which do not share a close phylogenetic relationship with a sequence from a characterized gene cluster. In some cases, gene clusters may be ordered according to their phylogenetic relationship to characterized gene clusters, clusters with the most distant relationships may be preferred for further analysis.

[127] In several embodiments, a specific secondary metabolite is used as a weapon against another organism (i.e., is toxic to or inhibits the growth of specific type of organism). In many instances, the toxic secondary metabolite may also be toxic to the organism that produces the secondary metabolite. Accordingly, the producing organism may defend against self-harm in a number of ways including (but not limited to) pumping the secondary metabolite out of the cell, enzymatically negating the secondary metabolite, or producing an additional version of the protein targeted by the secondary metabolite that is less sensitive or insensitive to the secondary metabolite, (see Figure 3). In instances in which an organism produces an additional version of the target protein, the "protective" version of the gene that encodes the additional version of the target protein that is less sensitive to the secondary metabolite is often colocalized with the biosynthetic gene cluster, for example the HMGR gene in the lovastatin cluster shown in Figure 4. Although different to the gene

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1031 of 3077

that produces the target protein, the protective version of the gene should produce a protein that maintains detectable homology to the target protein. In several embodiments, the cluster identification process takes advantage of this homology to identify those biosynthetic gene clusters that contain or are adjacent to a gene that encodes a protective homolog of the target protein. In a number of embodiments of the disclosure, the genetic sequences of multiple organisms are analyzed to detect biosynthetic gene clusters possessing this characteristic.

[128] A target protein may be any protein of interest which has a homolog in the genome sequence/s from which the gene clusters were obtained. In some cases, the target protein is an enzyme. In some cases, the target protein is a signaling protein. In some cases, the target protein is one which is required by the species of origin. For example, the target protein may be one which contributes to viability of growth of the cell, and deletion or inactivation of the target protein from the cell may have deleterious effects on the viability or growth of the cell. In some cases, the target protein may have a vertebrate or mammalian homolog. In some cases, the target protein has a human homolog. The human homolog may be a protein which is dysregulated in a disease. An example of a gene cluster which was identified using methods disclosed herein, and which comprises a homolog of the BRSK1 gene is shown in Figure 15A.

[129] In some embodiments, a biosynthetic cluster of interest which produces a secondary metabolite that interacts with a target protein may also produce a protein for inactivating the secondary metabolite or for secreting the secondary metabolite from the cell in which it is produced. A protein which inactivates the secondary metabolite may be omitted when designing an expression construct to express this cluster in a host cell. A protein involved in secreting the secondary may be included or omitted when designing an expression construct to express this cluster in a host cell. To identify such clusters several approaches may be used. For example a set of biosynthetic gene clusters may be identified from genome data and the identified clusters may be analyzed for the presence of enzymes with activities that may be useful for degradation of a secondary metabolite. In another example, a set of biosynthetic gene clusters may be analyzed for the presence of proteins involved in transport or secretion pathways. In another example a homology search may be identified across one or more genome sequences to find genes encoding proteins homologous to an enzyme known to degrade a toxic compound. Once such genes have been identified they may be

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1032 of 3077

analyzed for proximity to a biosynthetic gene cluster. Proximity to a gene cluster may be defined as within about 50 kb, 40 kb, 30 kb, 20 kb, 10 kb, 5 kb, lkb, or less than 1 kb.

[130] In some embodiments, a gene cluster which produces a secondary metabolite that may be toxic to the host cell may also contain signals to direct the production of the secondary metabolite to a specific cellular location. In some cases, the enzymes may be membrane tethered to the intracellular or extracellular side of a cell membrane, or may be secreted through a membrane to the intracellular or extracellular side (including the insides of organelles and vacuoles). For example, the enzymes of the cluster may contain membrane-targeting signals to target the enzymes to either the extracellular membrane or to an intracellular organelle or vacuole. In some cases, the enzymes may be targeted to the extracellular membrane in an orientation which results in the active region of the enzyme being inside of an organelle or vacuole. In some cases, the enzymes may be targeted to the extracellular membrane in an orientation which results in the active region of the enzyme being outside of the cell. Such clusters may be identified by analyzing the predicted proteins of the cluster for the presence of peptide targeting signals or of terminal sequences with homology to targeting signals.

[131] In some cases, a genome or genomes may be searched for gene clusters and the set of identified gene clusters may be searched for genes which are homologous to a target gene, or which produce proteins homologous to a target protein. In other cases, a genome or genomes may be searched for a gene or genes homologous to a target gene, and the identified genes may be analyzed for association with a gene cluster. In other cases, a genome or genomes may be searched for gene clusters and the set of identified gene clusters may be phylogenetically analyzed to determine relationships between the identified gene clusters and known, characterized, gene clusters. In yet other cases, a novel genome or genomes may be searched for sequences distantly homologous to a known biosynthetic enzyme, and the identified genes may be analyzed for association with a gene cluster.

[132] Specific secondary metabolites synthesized using methods in accordance with a number of embodiments of the disclosure and the proteins utilized to identify the biosynthetic gene clusters used to synthesize the secondary metabolites and targeted by the secondary metabolites are described herein. An example of a method which may be used to produce a secondary metabolite is shown in **Figure 5.** As shown in an exemplified embodiment process in **Figure 5A,** Process 100

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1033 of 3077

generates and extracts a heterologous compound for use. Exemplary Process 100 can begin by searching genetic data of various organismal species for pathways or BGCs that produce a compound product (101). The organismal species used in this step can be any species. For example, fungal and bacterial species often contain multiple BGC pathways that are encoded in its DNA. Likewise, the genetic data to be searched can be any genetic data available or determinable by the user. Accordingly, on one end of the spectrum, the genetic data may be a fully annotated, publicly available genome of a well-studied species (e.g., Penicillium notatum). On the other end of the spectrum, the genetic data may be a publicly unavailable, partial genomic sequence of a newly discovered species incapable of anthropogenic cultivation, wherein the partial sequence is found to have a gene cluster that may produce a compound.

[133] Exemplary embodiment Process 100 may continue by using the genetic data to reconstruct the compound product pathway in an acceptable genetic expression system (103). Often, to reconstruct the compound pathway, the genetic data is used to create nucleic acid molecules (e.g., DNA) comprising the coding sequences of the pathway genes sufficient to produce the product in the acceptable genetic expression system. The nucleic sequences are to be transferred into the expression system. Expression systems are any organism capable of producing the heterologous compound by heterologous expression of the pathway genes. Typical expression systems include (but are not limited to) *E. coli* and *S. cerevisiae.*

[134] Once the compound product pathway is reconstructed and transferred within an expression system, the expression system produces the compound (105) in exemplary Process 100. Typically, production of the compound results from coordinated expression of the pathway genes in the expression system. The coordinated expression of the pathway genes results in the production of the enzymes primarily responsible for constructing the heterologous compound product.

[135] **Figure 5B** depicts another exemplary embodiment process. Exemplary Process 200 produces, extracts, and characterizes a biosynthetic compound derived from heterologous expression of a gene cluster. The process may begin with the identification and selection of a gene cluster with an identifiable trait that is indicative of compound production (201). Several indicative processes to select gene clusters exist, including several computer-implemented programs. One such program is antiSMASH2.0, which is platform for mining BGC clusters for production of secondary metabolites that searches for core structures to identify putative BGCs (K. Blin, et al. Nucleic Acids

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1034 of 3077

Res. 41:W204-12, 2013, the disclosure of which is incorporated herein by reference in its entirety). Another such method is described herein which utilizes homolog sequences within a proximal region in the genome to identify putative BGCs. It should be noted that many other methodologies could be used to select BGCs.

[136] Once a gene cluster has been selected, Process 200 continues by appropriating nucleic acid molecules with the coding sequences of the various genes within the cluster (203 in **Figure** 5B). Typically, the nucleic molecules are DNA, but other nucleic molecules (e.g., RNA) can be used for certain applications. When extracting gene sequence data from the host organism, it is usual to remove the non-translated portions (e.g., UTRs, introns) from the gene, leaving only the coding sequence, but the non-translated portions may also be used, especially if they provide a beneficial characteristic. Appropriation of the nucleic molecules can be performed by many different methods including (but not limited to) direct extraction from the host, chemical synthesis, and/or cDNA generation methods (e.g., reverse transcription of host RNA). Regardless of the method used, the resulting nucleic acid molecules may be available to build into expression vectors for heterologous expression.

[137] Exemplary Process 200 utilizes the appropriated nucleic acid molecules to assemble expression vectors for expression in an appropriate organismal expression system (e.g., *E. coli, S. cerevisiae).* Expression vectors are nucleic acid molecules that have the necessary components to express a heterologous gene in the expression system. Common expression vectors are plasmid DNA and viral vectors, but also include kits of DNA molecules that can be joined together to form a longer DNA molecule by a recombination methodology (e.g., yeast homologous recombination (YHR)).

[138] To express a heterologous gene from an expression vector, an expression cassette is may be used, which comprises the sequences of an appropriate promoter and an appropriate terminator along with the heterologous gene sequence. The promoter is typically located upstream of the heterologous gene and can regulate the gene's expression. Many different types of promoters can be used. The selection of the appropriate promoter depends on the application and expression profile desired. For example, in the *S. cerevisiae* expression system, production-phase promoters may express heterologous genes only in the production-phase of the yeast culture's life cycle, which may have desirable properties. For more description of production-phase promoters, please refer to the related

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1035 of 3077

U.S. Patent Application No. 15/469,452 ("Inducible Production-Phase Promoters For Coordinated Heterologous Expression in Yeast"), the disclosure of which is incorporated herein by reference in its entirety. However, it should be understood, that constitutive promoters and other response-driven promoters could be used within the system.

[139] The sequences of promoters to be used in an expression vector can be derived from various sources. *E. coli* expression systems often use the T7 promoter derived from the T7 bacteriophage because the promoter reliably produces high expression in *E. coli.* Endogenous promoter sequences (e.g., the lac operon in *E. coli)* are expected to perform well within the organismal expression system.

[140] Expression vectors often have other sequences that benefit duplication, selection and stability of the vector within the organismal expression system, in addition to the expression cassette. In several instances, some of these sequences are necessary for maintenance in the expression system. For example, plasmid vectors within an *E. coli* or *S. cerevisiae* host require a host origin of replication and a selectable marker. The origin of replication signals the host expression system to replicate the plasmid vector in order to produce more copies of the plasmid as the host cells duplicate and divide. The selectable marker ensures that only the host cells that contain the vector continue to survive and propagate. Accordingly, these sequences may be necessary for viable heterologous expression.

[141] Once the expression vector is assembled, the heterologous BGC genes are expressed using the organismal expression system (207). Accordingly, the expression vector is to exist within the organismal host such that the host will express the heterologous BGC genes to produce the encoded enzymes. The enzymes produce a biosynthetic compound. This compound is be extracted from the expression system (209).

[142] Extracted heterologous biosynthetic compounds can be characterized to determine their various structures and conformations. Some resultant products may have a solitary structure and conformation while other products will have several different structures with multiple conformations. The various structures and conformations can be determined using mass spectrometry, chromatography, and/or other methods.

[143] There are numerous classes of biosynthetic compounds. For example, polyketides and terpenes are a class of compounds derived from various organismal species. Many novel

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1036 of 3077

biosynthetic compounds are likely to have beneficial properties, as a multitude of biosynthetic compounds have been found to be useful in several industries.

[144] Illustrated in **Figure 5C** is an exemplary pipeline to produce heterologous, biosynthetic compounds. Exemplary Pipeline 300 takes advantage of a yeast expression system to reproduce a fungal BGC in order to produce a heterologous compound product.

[145] Pipeline 300 begins with selection of a biosynthetic gene cluster (301). Depicted, as an example, is a phylogenetic tree of numerous fungal BGCs. Using the phylogenetic data, a BGC having a desired trait is selected. The coding sequences of the various BGC genes are then used to chemically synthesize DNA molecules (303). The synthetized BGC DNA molecules are then to be assembled into a heterologous expression construct (305). In this example, the DNA molecules are assembled by yeast homologous recombination. Accordingly, the synthesized DNA molecules have overlapping homologous sequences that the yeast will use to recombine the various DNA molecules into a plasmid DNA vector.

[146] Pipeline 300 then utilizes the assembled expression vectors to maintain and express the BGC in the yeast (307). The expression of the various heterologous genes results in expression of a number of heterologous enzymes that then can produce the heterologous biosynthetic compounds. Once a sufficient titer of compound is produced, it can be characterized to determine its structures and properties (309).

[147] Briefly, the method comprises gene cluster selection as discussed herein, synthesis of coding sequence, promoters and terminators, assembly of the cluster coding sequence, expression in a fungal host, and isolation and characterization of compounds produced. An example of a gene cluster identified using the methods herein, and the compound created by expression of the identified genes in yeast, is shown in Figure 15B.

Cluster engineering

[148] Once a gene cluster is identified according to the methods described herein, the cluster may be prepared for expression in a heterologous host cell. Editing of a putative gene cluster may involve steps such as: removal of introns, replacement of promoters, replacement of terminators, gene shuffling, and codon optimization. For example, if a gene cluster is to be expressed in *S.*

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1037 of 3077

cerevisiae then coding sequences from the gene cluster may be codon optimized for *S. cerevisiae,* and operably linked to *S. cerevisiae* promoters and terminators.

[149] The gene cluster editing may rely on automatic annotation of expressed sequences, introns, and exons, or on manual inspection of the cluster. The gene cluster editing may be done *in silico* using the sequence data, or may be done *in vitro* or *in vivo* using the DNA sequence in a suitable vector such as a cloning vector. In some cases an initial edited gene cluster may not produce a product and reanalysis of the predicted coding sequences, and introns of the same, may reveal errors in the predicted transcription start sites, transcription termination sites and/or splice sites.

[150] In an embodiment, this disclosure provides sequences (SEQ **ID NOs: 67-483)** of cryptic BGCs which encode various products. These BGCs may also be reengineered to provide the coding sequences without the endogenous regulatory sequences. The coding sequences from these clusters may be isolated and cloned into one or more expression vectors for expression in a model host system. The expression vectors may be plasmids, viruses, linear DNA, bacterial artificial chromosomes or yeast artificial chromosomes. The expression vectors may be designed to integrate into the host genome, or to not integrate. In some cases, the expression vector may be a high copy number plasmid.

Promoters

[151] Expression of a refactored gene cluster in a host organism may require coordinated expression of several different coding sequences. In some cases, the expression of multiple different coding sequences in a host organism may require the use of multiple different promoters suitable for that organism. This disclosure provides methods for discovering multiple promoters with similar activities and expression patterns but with differing DNA sequence. Such methods may involve use of closely related species, such as different species of Saccharomyces. Saccharomyces (S.) is a genus of fungi composed of different yeast species. The genus can be divided into two further subgenera: S. sensu stricto and S. sensu lato. The former have relatively similar characteristics, including the ability to interbreed, exhibiting uniform karyotype of sixteen chromosomes, and their use in the fermentation industry. The later are more diverse and heterogeneous. Of particular importance is the *S. cerevisiae* species within the S. sensu stricto subgenus, which is a popular model organism used for genetic research.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1038 of 3077

[152] The yeast *S. cerevisiae* is a powerful host for the heterologous expression of biosynthetic systems, including production of biofuels, commodity chemicals, and small molecule drugs. The yeast's genetic tractability, ease of culture at both small and large scale, and a suite of well-characterized genetic tools make it a desirable system for heterologous expression. Occasionally, production systems require coordinated expression of two or more heterologous genes. Coordinated expression systems in bacteria (e.g., *E. coli)* has long exploited the operon structure of bacterial gene clusters (e.g., lac operon), allowing a single promoter to control the expression of multiple genes.

[153] The construction of synthetic operons therefore allows a single inducible promoter to control the timing and strength of expression of an entire synthetic system. In yeast, many heterologous-expression systems do not rely on the operon system, but instead rely on a one-promoter, one-gene paradigm. Accordingly, multi-gene heterologous expression is generally performed using multiple expression cassettes with a well-characterized promoter and terminator, each on a single expression vector (e.g., plasmid DNA) (See D. Mumberg, R. Muller, and M. Funk Gene 156: 119-22, 1995). With traditional restriction-ligation cloning, it is also possible to recycle a promoter on a single plasmid by the serial cloning of multiple genes (M. C. Tang, et al., J Am Chem Soc 137:13724-27, 1995).

[154] Turning now to the drawings and data, disclosed embodiments are generally directed to systems and constructs of heterologous expression during the production phase of yeast. In many of these embodiments, the expression system involves coordinated expression of multiple heterologous genes. More embodiments are directed to production-phase promoter systems having promoters that are inducible upon an event in the yeast's growth or by the nutrients and supplements provided to the yeast. Specifically, a number of embodiments are directed to the promoters that are capable of being repressed in the presence of glucose and/or dextrose. In more embodiments, the promoters are capable of being induced in the presence of glycerol and/or ethanol. In additional embodiments, at least one production-phase promoter exists within an exogenous DNA vector, such as (but not limited to), for example, a shuttle vector, cloning vector, and/or expression vector. Embodiments are also directed to the use of expression vectors for the expression of heterologous genes in a yeast expression system.

[155] Controlled gene expression is desirable in heterologous expression systems. For example, it would be desirable to express heterologous genes for production during a longer stable phase.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1039 of 3077

Accordingly, decoupling the anaerobic growth and aerobic production phases of a culture allows the yeast to grow to high density prior to introducing the metabolic stress of expressing unnaturally high amounts of heterologous protein. In accordance with many embodiments, the anaerobic growth phase is defined by the yeast culture's energy metabolism in which the yeast cells predominantly catabolize fermentable carbon sources (e.g., glucose and/or dextrose), and a high growth rate (i.e., short doubling-time). In contrast, and in accordance with several embodiments, the aerobic production phase is defined by the yeast culture's energy metabolism in which the yeast cells predominantly catabolize nonfermentable carbon sources (e.g., ethanol and/or glycerol), and a steady growth rate (i.e., long doubling-time). Accordingly, each yeast cell's energy metabolism can be predominantly in aerobic or nonaerobic phase, and dependent on the local concentration of the carbon source.

[156] **Figure 6A** depicts the phases of a yeast culture when provided a fermentable sugar, such as glucose or dextrose sugar, at a concentration of around 2-4% as its main carbon source. Initially, a yeast culture will predominantly catabolize the fermentable sugar, which correlates with an exponential growth with very high doubling rates. The growth phase typically lasts approximately 4-10 hours. During this phase, the catabolism of the fermentable sources results in the production of ethanol and glycerol.

[157] Once glucose becomes scarce, the growth of a yeast culture passes a diauxic shift and begins to predominantly catabolize nonfermentable carbon sources (e.g., ethanol and/or glycerol) **(Figure 6B).** The predominant catabolism of nonfermentable carbon source correlates with a longer and more stable production phase that can last for several days, or even weeks in an industrial-like setting **(Figure 6A).** During the production phase, yeast cultures reach and maintain a high concentration, but have a much lower doubling time **(Figure 6A).** Due to the decrease in doubling rate, yeast cultures no longer expend a great amount of energy and resources on rapid growth and thus can reallocate that energy and those resources to other biological activities, including heterologous expression. Accordingly, it is hypothesized that limiting the transcription of heterologous genes to the production phase would allow a yeast culture to reach a high, healthy confluency that would in turn allow better heterologous protein expression and biosynthetic production.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1040 of 3077

[158] In yeast, transcriptional regulation can be achieved in several ways, including inducement by chemical substrates (e.g., copper or methionine), the tetON/OFF system, and promoters engineered to bind unnatural hybrid transcription factors. Perhaps the most commonly employed inducible promoters are the promoters controlled by the endogenous GAL4 transcription factor. GAL4 promoters are strongly repressed in glucose, and upon switching to galactose as a carbon source, strong induction of transcription is observed (M. Johnston and R. W. Davis, Mol. Cell Biol. 4:1440-48, 1984). While this system leads to high-level transcription, only four galactose-responsive promoters are known, and galactose is both a more expensive and a less efficient carbon source as compared to glucose (S. Ostergaard, et al., Biotechnol. Bioeng. 68:252-59, 2000). Other carbon-source dependent promoters have also been used for heterologous gene expression. The *S. cerevisiae* ADH2 gene exhibits significant derepression upon depletion of glucose as well as strong induction by either glycerol or ethanol (K. M. Lee & N. A. DeSilva Yeast. 22:43 1-40, 2005). Once induced, genes driven by the ADH2 promoter (pADH2) display expression levels equivalent to those driven by highly expressed constitutive counterparts. This induction profile was found to work in heterologous expression studies, as the system auto-induces upon glucose depletion in the late stages of fermentative growth after cells have undergone diauxic shift. The ADH2 promoter has been used extensively for yeast heterologous expression studies, resulting in high-level expression of several heterologous biosynthetic proteins (For example, see C. D. Reeves, et al., Appl. Environ. Microbiol. 74:5121-29, 2008).

[159] As shown in **Figure 6C,** the concentration of ethanol and glycerol increases as glucose and dextrose sugar decreases, due to anaerobic glycolysis (i.e., breaking down the fermentable sugar) and subsequent fermentation (i.e., converting the broken-down glucose into alcohol) and glycerol biosynthesis (i.e., converting the broken-down glucose into glycerol). Upon fermentable sugar depletion, yeast cultures undergo a diauxic shift and begin to use ethanol and glycerol as a carbon source instead of glucose. A diauxic shift, as understood in the art, is defined as a point in time when an organism switches from primarily consumption of one source for energy, to primarily another source. This shift typically elicits significant changes to a yeast culture's gene-expression pattern. Accordingly, it is hypothesized that higher concentrations of ethanol, (e.g., -2-4%) and or glycerol (e.g., -2%) could be used to stimulate promoters that either directly or indirectly respond to these concentrations (See Figures 6A and 6C).

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1041 of 3077

[160] Various disclosed embodiments are based on the discovery of inducible promoters that can be used for the coordinated expression of multiple genes (e.g., gene cluster pathway) in Saccharomyces yeast. Described below are sets of inducible promoters from *S. cerevisiae* and related species that are inactive during anaerobic growth, activating transcription only after a diauxic shift when glucose is near-depleted and the yeast cells are respiring (i.e., the production phase). As portrayed in various embodiments, various production-phase promoters are auto-inducing and allow automatic decoupling of the growth and production phases of a culture and thus initiate heterologous expression without the need for exogenous inducers. It should be noted, however, that many embodiments include production-phase promoters that are also inducible in the presence of nonfermentable carbon-sources (e.g., ethanol and/or glycerol) supplied to the yeast. As such, multiple embodiments employ recombinant production-phase promoters that act much like constitutive promoters when the host yeast cultures are constantly maintained in ethanol and/or glycerol-containing media.

[161] Once activated, the strength of various production-phase promoters can vary as much as 50-fold. The strongest production-phase promoters stimulate heterologous expression greater than that observed from strong constitutive promoters. The production-phase promoters could be employed in many different applications in which high expression of multiple genes is beneficial. Accordingly, the promoters can be used, for example, in multiple subunit protein production or for the production of biosynthetic compounds that are produced by multiple proteins within a pathway. Discussed in an exemplary embodiment below, some embodiments are used to express multiple proteins involved in production of indole diterpene compound product. When compared to constitutive promoters, the production-phase promoters produced greater than a 2-fold increase in titer of the exemplary diterpene compounds. In other exemplary embodiments, it was found that the production-phase promoter system outperformed constitutive promoters by over 80-fold. Thus, these promoters can enable heterologous expression of biosynthetic systems in yeast.

[162] The practice of several embodiments will employ, unless otherwise indicated, conventional methods of chemistry, biochemistry, and molecular biology and recombinant DNA techniques within the skill of the art. Such techniques are explained fully in the literature. See, e.g., A.L. Lehninger, Biochemistry (Worth Publishers, Inc., 30 current addition); Sambrook, et al., Molecular

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1042 of 3077

Cloning: A Laboratory Manual (3rd Edition, 2001); Methods In Enzymology (S. Colowick and N. Kaplan eds., Academic Press, Inc.).

Inducible Production-Phase Promoters for Heterologous Expression in Yeast

[163] In accordance with several embodiments, inducible production-phase promoters can be constructed into exogenous expression vectors for production of at least one protein in Saccharomyces yeast. In many embodiments, the constructed expression vectors have multiple inducible production-phase promoters in order to express multiple heterologous genes. Several embodiments are directed to production-phase promoters and DNA vectors incorporating these promoters. Promoters, in general, are defined as a noncoding portion of DNA sequence situated proximately upstream of a gene to regulate and promote its expression. Typically, in *S. cerevisiae* and similar species, the promoter of a gene can be found within 500-bp upstream of a gene's translation start codon. In some cases, a promoter may be about 500 bp, 600 bp, 700 bp, 800 bp, 900 bp, 1 kb, 1.5 kb, 2 kb or more than 2 kb upstream of a gene's transcription start site.

[164] In accordance with several embodiments, production-phase promoters have two defining characteristics. First, production-phase promoters are capable of repressing heterologous expression of a gene in *S. cerevisiae* and similar species when the yeast is exhibiting anaerobic energy metabolism. As described previously, yeast exhibit anaerobic metabolism in the presence of a nontrivial concentration of fermentable carbon sources such as, for example, glucose or dextrose. In addition, production-phase promoters are also capable of inducing heterologous expression of a gene in *S. cerevisiae* and similar species when the yeast is exhibiting aerobic energy metabolism. As described previously, yeast exhibit aerobic metabolism when fermentable carbon sources are near depleted and the yeast cells switch to a catabolism of nonfermentable carbon sources such as glycerol or ethanol. These characteristics correspond to the phase charts in Figures 6A-6C. Tables 1 and 2 provide several examples of production-phase promoters in accordance with several embodiments.

[165] The production-phase promoters can be characterized based on their level of transgene expression relative to each other and to constitutive promoters. As described in an exemplary embodiment below, it was found that the sequence of endogenous promoters of the *S. cerevisiae* genes ADH2, PCK1, MLS1, and ICL1 exhibited high-level expression and thus can be characterized

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1043 of 3077

as strong production-phase promoters **(Table 1).** Sequences of the endogenous promoters of the *S. cerevisiae* genes YLR307C-A, ORF-YGR067C IDP2, ADY2, CACI, ECM13, and FAT3 exhibited mid-level expression and thus can be characterized as semi-strong production phase promoters **(Table 1).** In addition, sequences of the endogenous promoters of the *S. cerevisiae* genes PUT1, NQMl, SFCl, JENl, SIP18, AT02, YIGl, and FBPl exhibited low-level expression and thus can be characterized as weak production-phase promoters **(Table 1).**

Table 1. Production-Phase Promoters Expression Phenotype

Gene Name	Systematic Name	Expression Phenotype	Sequence ID Number
ADH2	YMR303C	Strong	1
PCK1	YKR097W	Strong	2
MLS1	YNL117W	Strong	3
ICL1	YER065C	Strong	4
YLR307C-A	YLR307C-A	Semi-Strong	5
YGR067C	YGR067C	Semi-Strong	6
IDP2	YLR174W	Semi-Strong	7
ADY2	YCR010C	Semi-Strong	8
GAC1	YOR178C	Semi-Strong	9
ECM13	YBL043W	Semi-Strong	10
FAT3	YKL187C	Semi-Strong	11
PUT1	YLR142W	Weak	12
NQM1	YGRO43C	Weak	13
SFC1	YJR095W	Weak	14
JEN1	YKL217W	Weak	15
SIP18	YMR175W	Weak	16
ATO2	YNR002C	Weak	17
YIG1	YPL201C	Weak	18
FBP1	YLR377C	Weak	19

[166] The closely related S. sensu stricto species have similar genetics and growth characteristics. Accordingly, the phase charts provided in **Figures 6A-6C** apply generally to S. sensu stricto species. **Table 2** provides a list of strong production-phase exogenous promoters of similarly related species in accordance with numerous embodiments of the disclosure.

Table 2. Strong Production-Phase Promoters of *S. sensu stricto* species		
Species	Gene Name	Sequence ID Number
S. paradoxus	ADH2	36
S. kudriavzevii	ADH2	37

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1044 of 3077

S. bayanus	ADH2	38
S. paradoxus	PCK1	41
S. kudriavzevii	PCK1	42
S. bayanus	PCK1	43
S. paradoxus	MLS1	44
S. kudriavzevii	MLS1	45
S. bayanus	MLS1	46
S. paradoxus	ICL1	47
S. kudriavzevii	ICL1	48
S. bayanus	ICL1	49

[167] It should be noted that substantially similar sequences to the production-promoter sequences are expected to regulate heterologous expression in *S. cerevisiae* and achieve similar results. Accordingly, a substantially similar sequence of a production-phase promoter, in accordance with numerous embodiments, is any sequence with a high functional equivalence such that when regulating heterologous expression in *S. cerevisiae* that it achieves substantially similar results. For example, in an exemplary embodiment below, it was found that the ADH2 promoter of *S. bayanus* is only 61% homologous, yet achieved strong heterologous expression in *S. cerevisiae,* similar to the endogenous ADH2 promoter. In some cases, a substantially similar sequence may be homologous to the promoter sequences identified herein (e.g., have a nucleotide BLAST e value of less than or equal to 10^{-10}, 10^{-20}, 10^{-30}, 10^{-35} or 10^{-40}).

[168] In **Figure 7A,** an exemplary schematic of a section of an exogenous DNA vector (e.g., cloning vector, expression vector, and/or shuttle vector) having a production-phase promoter sequence embedded within. A vector is capable of transferring nucleic acid sequences to target cells (e.g., yeast). Typical DNA vectors include, but are not limited to, plasmid or viral constructs. DNA vectors are also meant to include a kit of various linear DNA fragments that are to be recombined to form a plasmid or other functional construct, as is common in yeast homologous recombination methods (See e.g., Z. Shao, H. Zhao & H. Zhao, 2009, Nucleic Acids Research 37:el6, 2009, the disclosure of which is incorporated herein by reference). Often, embodiments of cloning vectors will incorporate other sequences in addition to the production-phase promoter. As depicted in **Figure 7A,** the exemplary cloning vector has a terminator sequence and cloning/recombination sequence in addition to the production-phase promoter, each of which can assist with expression vector construction. Furthermore, other sequences necessary for growth and amplification can be

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1045 of 3077

incorporated into the promoter vector. Embodiments of these sequences may include, for example, at least one appropriate origin of replication, at least one selectable marker, and/or at least one auxotrophic marker. It should be noted, however, that various embodiments of the disclosure are not required to contain cloning, terminator, or either sequences. For example, embodiments of a typical shuttle vector may only contain the production-phase promoter sequence along with the necessary sequences for amplification in a biological system.

[169] For purposes of this application, an exogenous DNA vector is any DNA vector that was constructed, at least in part, exogenously. Accordingly, DNA vectors that are assembled using the yeast's own cell machinery (e.g., yeast homologous recombination) would still be considered exogenous if any of the DNA molecules transduced within yeast for recombination contain exogenous sequence or were produced by a non-host methodology, such as, for example, chemical synthesis, PCR amplification, or bacterial amplification.

[170] As shown in **Figure 7B,** various embodiments of the disclosure are directed to DNA vectors having multiple production-phase promoters. In these various embodiments, multiple different production-phase promoters are incorporated, preferably each having a unique sequence and derived from a different gene and/or S. sensu stricto species. Having unique promoter sequences can prevent complications that can arise during product production in yeast, such as, for example, unwanted DNA recombination at sites similar to the promoter sequences that render the DNA vector constructs undesirable. In many embodiments, the DNA vector has at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or more than 20 production-phase promoters. As the size of the DNA vector increases, the utility may decrease, as larger vectors may become unwieldly for the intended organism to handle. For example, plasmids for amplification in *E. coli* are often somewhere between 2,000 and 10,000 base pairs (bp) but can handle up to 20,000 bp or so. Likewise, plasmids for amplification and growth in yeast can vary from approximately 10,000 to 30,000 bp. Viral vectors, on the other hand, often have a limited construct size and thus may require a more precise vector size. Thus, depending on vector and intended use, the number of production-phase promoters within a DNA vector will vary.

[171] Although **Figure 7B** depicts recombination sites, cloning sites, and terminator sequences, it should be noted that these sequences may or may not be included in various embodiments of DNA vectors having multiple production-phase promoters. The incorporation of these sequences or other

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1046 of 3077

various sequences is often dependent on the purpose of the DNA vector. For example, cloning vectors may not include a terminator sequence if that sequence is to be incorporated into an expression construct at another stage of assembly.

[172] **Figure 8A** depicts an exemplary heterologous expression vector having a production-phase promoter for expression in yeast, in accordance with various embodiments of the disclosure. Expression constructs contain an expression cassette that has a promoter, a heterologous gene, and a terminator sequence in order to produce an RNA molecule in an appropriate host. Expression cassette in accordance with numerous embodiments will have a production-phase promoter situated proximately upstream of a heterologous gene of which the promoter is to regulate expression. It should be understood, that the precise location of the production-phase promoter upstream of the heterologous gene may vary, but the promoter generally is within a certain proximity to adequately function.

[173] In many embodiments of the disclosure, a heterologous gene is any gene driven by a production-phase promoter, wherein the heterologous gene is different than the endogenous gene that the promoter regulates within its endogenous genome. Accordingly, a *S. cerevisiae* production-phase promoter could regulate another *S. cerevisiae* gene provided that the gene to be regulated is not the gene endogenously regulated. For example, the *S. cerevisiae* ADH2 promoter should not regulate the *S. cerevisiae* ADH2 gene; however, the *S. cerevisiae* ADH2 promoter can regulate any other *S. cerevisiae* gene or the ADH2 gene from any other species. Often, in accordance with many embodiments, the heterologous gene is from a different species than the species from which the production-promoter sequence was obtained.

[174] Although not depicted, various embodiments of expression cassettes may include other sequences, such as, for example, intron sequences, Kozak-like sequences, and/or protein tag sequences (e.g., 6x-His) that may or may not improve expression, production, and/or purification. In yeast, various embodiments of expression vectors will also minimally have a yeast origin of replication (e.g., 2-micron) and an auxotrophic marker (e.g., URA3) in addition to the expression cassette. Other nonessential sequences may also be included, such as, for example, bacterial origins of replication and/or bacterial selection markers that would render the expression capable of amplification in a bacterial host in addition to a yeast host. Accordingly, various embodiments of

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1047 of 3077

expression vectors would include the essential sequences for heterologous expression in yeast and other various embodiments would include additional nonessential sequences.

[175] In accordance with various embodiments, a DNA vector having a production-phase promoter expression cassette can be transformed into a yeast cell. Or alternatively, and in accordance with numerous embodiments, a DNA vector having a production-phase promoter expression cassette can be assembled within yeast using homologous recombination techniques. Once existing within a yeast cell, the production-phase promoter can regulate the expression of a heterologous gene in accordance with the yeast cell's energy metabolism. As described previously, and in accordance with many embodiments, production-phase promoters repress heterologous expression when the yeast cell is in an anaerobic energy metabolic state. Alternatively, and in accordance with a number of embodiments, production-phase promoters induce heterologous expression when the yeast cell is in an aerobic energy metabolic state

[176] Depicted in **Figure 8B** are alternative exemplary heterologous expression vectors having multiple production-phase promoters for expression of multiple genes in yeast in accordance with numerous embodiments. In some embodiments, the expression vectors will include at least two expression cassettes, each with a unique promoter, gene, and terminator sequence in order to prevent unwanted recombination. The number of expression cassettes will vary based on vector construct design and application. For heterologous expression in *S. cerevisiae,* it has been found that plasmid expression vectors of approximately 30,000 bp are still tolerated. Thus, vectors containing up to seven production-phase promoter expression cassettes can be incorporated into an expression vector and have been found to be able to maintain adequate gene expression and protein production. Larger vectors with more expression cassettes may be tolerated.

[177] Although **Figure 8B** depicts multiple expression cassettes sequentially in the same orientation (5' to 3'), it should be understood that the combination of two or more expression cassettes is not limited to sequential linear organization in the same orientation. Expression cassettes in accordance with many embodiments exist within the expression vector in any orientation and in any sequential order. Furthermore, it should be understood that other sequence elements of an expression vector (e.g., an auxotrophic marker) may be among and/or between the multiple expression cassettes. Optimal vector design is likely to depend on various factors, such as, for

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1048 of 3077

example, optimizing the location of the auxotrophic marker to enable the final expression vector to include each expression cassette to be incorporated.

[178] DNA heterologous expression vectors are a class of DNA vectors, and thus the description of general DNA vectors above also applies to the expression vectors. Accordingly, many embodiments of the expression vectors are formulated into a plasmid vector, a viral vector, a circular vector, or a kit of linear DNA fragments to be recombined into a plasmid by yeast homologous recombination. In several of these embodiments, the end-product vector contains at least one expression cassette having a production- phase promoter. It should be understood, that in addition to the at least one production-phase promoter, some vector embodiments incorporate expression cassettes that include other promoters, such as (but not limited to), constitutive promoters that maintain high expression during the growth and production phases.

[179] The various embodiments of heterologous expression vectors having at least one production-phase promoter can be used in numerous applications. For example, high expression in the production phase can lead to better, prolonged expression, as compared to constitutive promoters. In many applications, the end product is a protein from a single gene or a protein complex of multiple genes to be purified from the culture. For these applications, high, prolonged expression using production-phase promoters can lead to better yields of proteins. Furthermore, when the heterologous protein is toxic to the host yeast cells, the use of production-phase promoters prevents the expression of the toxic protein during growth phase, allowing the yeast to reach a healthy confluency before mass protein production.

[180] The production-phase promoter vectors can also benefit the production of a biosynthetic compound from a gene cluster. Many products derived from various natural species are produced from a cluster of genes with sequential enzymatic activity. For example, the antibiotic emindole SB is produced from a cluster of four genes that is expressed in *Aspergillus tubingensis*. To reproduce this gene cluster in a yeast production model, a production-promoter vector system with four different expression cassettes could work. This system would allow the yeast to reach a healthy confluency before the energy-draining expression of four heterologous proteins begins, leading to better overall yields of the antibiotic product. In fact, experimental results provided in an exemplary embodiment described in Example 1 below demonstrate that a production-phase promoter vector

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1049 of 3077

outperformed a constitutive promoter vector approximately 2-fold to produce the emindole SB product.

[181] **Figure 9** depicts an exemplary process (Process 400) to implement various embodiments of production-phase promoters. To begin, Process 400 identifies and selects at least one gene for heterologous expression in yeast (401). The choice of gene(s) for expression would depend on the desired outcome. For example, to produce a biosynthetic compound, one would likely select to express all, or a subset, of the genes within a biosynthetic gene cluster of a particular organism. Once the gene(s) have been selected, Process 400 then appropriates DNA molecules having the coding sequence of the selected genes (403). As is well known in the art, there are many ways to appropriate DNA molecules, which include chemical synthesis, extraction directly from the biological source, or amplification of a gene by polymerase chain reaction (PCR).

[182] Process 400 then uses the appropriated DNA molecules to assemble these molecules into an expression vector having production-phase promoters (405). There are many ways to assemble DNA expression vectors that are well known in the art, which include popular methodologies such as homologous recombination and restriction digestion with subsequent ligation. After assembly, the resultant expression vectors can be expressed in Saccharomyces yeast to obtain the desired outcome (407).

Yeast homologous recombination for plasmid construction and direct plasmid sequencing

[183] Each of the one or more expression vectors may contain one or more promoters suitable for expression of a heterologous gene in a model host system. Each expression vector may contain a single coding sequence or multiple coding sequences. Multiple coding sequences may be functionally linked to a single promoter, for example via an internal ribosome entry site, or may be linked to multiple promoters. The expression vectors may also contain additional elements to regulate or increase the transcriptional activity, for example enhancers, polyA sequences, introns, and posttranscriptional stability elements. The expression vectors may also contain one or more selectable markers.

[184] To improve high-throughput assembly and characterization of orphan biosynthetic systems or other systems, an automated DNA assembly pipeline using yeast homologous recombination, (YHR), as its core technology was developed. An example of the design strategy for assembling

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1050 of 3077

DNA parts for this pipeline is illustrated in **Figure 10A.** In one embodiment a method is provided for assembling synthetic gene clusters with heterologous regulation. DNA polynucleotides coding for a series of distinct promoters and terminators may be obtained in bulk and used for a variety of different synthetic gene clusters. Once a gene cluster of interest is identified the coding sequences are determined and then all coding sequences are synthesized with a flanking sequence (assembly overhang) on each side. The assembly overhangs, encode either for the flanking promoter and terminator if the gene is small enough to be ordered as a single piece, or for the adjacent gene fragments for longer sequences. The length of the flanking sequences may vary, In some cases, the flanking sequences may be about 30 bp, 40 bp, 50 bp, 60 bp, 70 bp, 80 bp, 90 bp, 100 bp, 30-100 bp, 30-70 bp, 40-60 bp, 40-80 bp, or 45-55 bp in length. Placing assembly overhangs exclusively on the unique coding sequence fragments allows for all regulatory cassettes to be generated in bulk and stockpiled as the same fragments are used in all assemblies. For example, in an assembly involving three or more genes, an auxotrophic marker may be placed between the second terminator and third terminator while no marker is present on the vector. By providing the auxotrophic marker and origin of replication on separate fragments, reaction background was significantly reduced. Additional modest increases in efficiency were observed when the assembly host is lacking a DNA ligase, such as the DNL4 DNA ligase.

[185] In some embodiments, this disclosure provides a system for generating a synthetic gene cluster via homologous recombination. The system comprises 1 though N unique promoter sequences, 1 through N unique terminator sequences, and 1 through N unique coding sequences. Each terminator sequence may be linked to the following promoter sequence, for example terminator 1 is linked to promoter 2, terminator 2 is linked to promoter 3, and so forth till terminator N-l which is linked to promoter N. In some cases, promoter 1 and terminator N may be attached to a linear plasmid backbone. Coding sequence 1 is attached to an additional 30-70 base pair sequence on each end such that a first end portion is identical or homologous to the last 30-70 base pairs of promoter 1 and a second end portion is identical or homologous to the first 30-70 base pairs of terminator 1. Coding sequence 2 is attached to an additional 30-70 base pair sequence on each end such that a first end portion is identical or homologous to the last 30-70 base pairs of promoter 2 and a second end portion is identical or homologous to the first 30-70 base pairs of terminator 2. Coding sequence N is also attached to an additional 30-70 base pair sequence on each end such that a first end portion

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1051 of 3077

is identical or homologous to the last 30-70 base pairs of promoter N and a second end portion is identical or homologous to the first 30-70 base pairs of terminator N. These DNA fragments may be assembled transforming the 1 through N promoters, terminators and coding sequences into a yeast cell where they are combined through yeast homologous recombination, and then isolating a plasmid containing the 1 through N promoters, terminators and coding sequences from the yeast cell.

[186] An example of this system is shown in **FIG. 10A.** In this example N equals four. The system comprises four unique promoters (110, 120, 130 and 140), four unique terminator sequences (210, 220, 230 and 240), and four unique coding sequences (310, 320, 330, and 340). Each of the coding sequences is created with an additional 30-70 base pair sequence that is homologous or identical to the sequence of the preceding promoter and an additional 30-70 base pair sequence that is homologous or identical to the sequence of the subsequent terminator. Thus, coding sequence 310 is flanked by sequence 111 which is identical or homologous to at least a part of sequence 110, and sequence 211 which is identical or homologous to an least a part of sequence 210. Coding sequence 320 is flanked by sequence 121 which is identical or homologous to at least a part of sequence 120, and sequence 221 which is identical or homologous to an least a part of sequence 220. Coding sequence 330 is flanked by sequence 131 which is identical or homologous to at least a part of sequence 130, and sequence 231 which is identical or homologous to an least a part of sequence 230. Coding sequence 340 is flanked by sequence 141 which is identical or homologous to at least a part of sequence 140, and sequence 241 which is identical or homologous to an least a part of sequence 240. In this example promoter sequence 110 and terminator sequence 240 are attached to the ends of a linearized plasmid backbone, and the DNA fragment comprising terminator 210 and promoter 120 further comprises an auxotrophic marker (400). Terminator 210 is linked to promoter 120, terminator 220 is linked to promoter 130, and terminator 230 is linked to promoter 140.

[187] As shown in **FIG. 10B,** once the DNA sequences from **FIG. 10A** are transfected into a yeast cell the homologous sequences are paired up and the fragments are linked together through yeast homologous recombination. The resultant DNA plasmid is illustrated in **FIG. IOC.**

[188] Traditionally, for yeast homologous recombination plasmid assemblies, plasmid DNA is isolated from assembly clones and transformed into *E. coli* in order to obtain sufficiently pure DNA to enable sequencing. The necessity of this step arises from the relatively low plasmid yields from yeast and the large amounts of contaminating genomic DNA in every sample. This disclosure

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1052 of 3077

provides a method by which plasmid DNA may be sequenced directly out of yeast. This may be achieved by a modified plasmid prep in which the majority of contaminating DNA is removed by treatment with an exonuclease enzyme. Any enzyme with exonuclease activity and free of endonuclease activity may be used in this step. Examples of exonuclease enzymes include but are not limited to: Lambda Exonuclease, RecJf, Exonuclease III *(E. coli),* Exonuclease I *(E. coli),* Exonuclease T, Exonuclease V (RecBCD), Exonuclease VIII, truncated, Exonuclease VII, T5 Exonuclease, and T7 Exonuclease. In some cases, the exonuclease is Exonuclease V. If an exonuclease with activity for single stranded DNA (ssDNA) and not double stranded DNA (dsDNA) is to be used, then the DNA may first be heated to denature the dsDNA. In some cases, the DNA may be treated with a topoisomerase to relax supercoiled plasmids. In some cases, the DNA may not be treated with a topoisomerase. Once the plasmid DNA has been purified by this method sequencing libraries can be prepared. **Figure 10D** demonstrates the increase in purity observed with the exonuclease treatment. Overall, this pipeline has been applied to the sequencing of >1000 clones. Assemblies of up to 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 or more than 30 unique DNA fragments can be achieved with high efficiency. **Figure 10E** shows efficient assembly of 2, 3, 4, 5, 6, 8, 10, 12, and 14 DNA fragments. In some cases a strain as described herein allows DNA assembly via homologous recombination with an efficiency of at least 60%, 70%, 80%, 90%, 100%, 110%, 120%, 130%, 140%, 150%, or more than 150% as compared to DNA assembly in BY.

[189] To increase the efficiency of assemblies by yeast homologous repair, the relative efficiencies of BY4743 and BY4743ADNL4 were tested; a strain in which the DNL4 ligase, involved in non-homologous end joining, has been deleted. **Figure 11A** illustrates efficiencies of several plasmid assemblies done in both strains, demonstrating the deletion of the DNL4 DNA ligase does consistently serve as the more efficient assembly background.

[190] The sequencing of plasmid DNA directly out of yeast as in **Figure H E** (e.g. without transforming into another host such as *E. coli* as in **Figure 11D**) is an advantage of the methods described herein. In establishing these methods, multiple means of preparation for both the plasmid DNA and the next-generation sequencing (NGS) library prep were tested. Shown in **Figure 1IB** is a comparison of sequencing efficiency using DNA prepared both from colonies picked from plates

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1053 of 3077

and cell pellets collected from 1 ml of liquid culture. These data show that these approaches generate samples of equivalent purity.

[191] Initially, the platform utilized an NGS library preparation in which the purified, exonuclease treated plasmid DNA was ultrasonically sheared, followed by end repair, A-tailing, and adaptor ligation. In order to decrease labor and increase throughput, a recently published modification of the Illumina NexeraXT transposase based prep was performed (M. Baym, et al., *PLoS One* 10:e01280367, 2015). Ultrasonic shearing necessitated the serial processing of multiple plates of clones while tagmentation allows for parallel processing of multiple plates. **Figure 11C** demonstrates that this modified Nextera preparation provides equivalent efficiency as compared to the standard approach.

[192] This approach may be suitable for multiple DNA preparation methods in multiple strain backgrounds. Additionally, it was shown that this approach is compatible with various library preparations for sequencing on Illumina platforms. It is anticipated that this approach could be easily modified to function in sequencing workflows using alternate sequencing platforms such as those provided by Pacific Bioscience and Oxford Nanopore technologies.

Host cells

[193] The expression vectors may be transfected into a host cell to produce secondary metabolites. The host cell may be any cell capable of expressing the coding sequences from the expression vectors. The host cell may be a cell which can be grown and maintained at a high density. For example the host cell may be one which may be grown and maintained in a bioreactor or fermenter. The host cell may be a fungal cell, a yeast cell, a plant cell, an insect cell or a mammalian cell.

[194] In some cases the host cell is bacterial. The bacteria may be a Proteobacteria such as a Caulobacteria, a phototrophic bacteria, a cold adapted bacteria, a Pseudomonads, or a Halophilic bacteria; an Actinobacteria such as Streptomycetes, Norcardia, Mycobacteria, or Coryneform; a Firmicutes bacteria such as a Bacilli, or a lactic acid bacteria. Examples of bacteria which may be used include, but are not limited to: *Caulobacter crescentus, Rodhobacter sphaeroides, Pseudoalteromonas haloplanktis, Shewanella sp. strain Ac10, Pseudomonas fluorescens, Pseudomonas putida, Pseudomonas aeruginosa, Halomonas elongata, Chromohalobacter salexigens, Streptomyces lividans, Streptomyces griseus, Nocardia lactamdurans, Mycobacterium*

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1054 of 3077

smegmatis, Corynebacterium glutamicum, Corynebacterium ammoniagenes, Brevibacterium lactofermentum, Bacillus subtilis, Bacillus brevis, Bacillus megaterium, Bacillus licheniformis, Bacillus amyloliquefaciens, Lactococcus lactis, Lactobacillus plantarum, Lactobacillus casei, Lactobacillus reuteri, and *Lactobacillus gasseri.*

[195] In some cases the host cell is a fungal cell. In some cases, the host cell is a yeast cell. Examples of yeast cells include, but are not limited to *Saccharomyces cerevisiae, Saccharomyces pombe, Candida albicans,* and *Cryptococus neoformans.* In some cases the host cell may be a filamentous fungi, such as a mold. Examples of molds include, but are not limited to Acremonium, Alternaria, Aspergillus, Cladosporium, Fusarium, Mucor, Penicillium, and Rhizopus. In some cases, the host cell may be an Acremonium cell. In some cases, the host cell may be an Alternaria cell. In some cases, the host cell may be an Aspergillus cell. In some cases, the host cell may be an Cladosporium cell. In some cases, the host cell may be an Fusarium cell. In some cases, the host cell may be a Mucor cell. In some cases, the host cell may be a Penicillium cell. In some cases, the host cell may be a Rhizopus cell.

[196] In some cases, the host cell is an insect cell. In some cases the host cell is a mammalian cell. Examples of mammalian cell lines include HeLa cells, HEK293 cells, B16 melanoma cells, Chinese hamster ovary cells, or HT1080. In some cases, the host cell is a plant cell. In some cases, the host cell may be part of a multicellular host organism.

[197] In some cases, the host cell is a genetically engineered cell. The yeast strain BJ5464 has historically been a workhorse strain for expression of heterologous proteins. BJ5464 lacks two vacuolar proteases genes (PEP4 and PRB1), which makes the strain useful for biochemical studies, owing to reduced protein degradation. However, BJ5464 has several problems that limit its utility. It has a high rate of petite cell formation, which results in offspring that cannot respire (grow on ethanol as a carbon source) and cannot express the respiration-induced promoters used in this project. It is not genetically tractable because it cannot sporulate, and its non-deletion auxotrophic markers prevent facile genome editing. Finally, BJ5464 is slow growing.

[198] This disclosure includes a new yeast super host based on the BY background. BY is a direct descendent of the yeast genome sequence reference strain and contains the complete deletions of auxotrophic markers that facilitate genome editing. It is the basis of the barcoded deletion collection, which has led to a wealth of genetic and chemico-genomic data. However, it also has major

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1055 of 3077

problems limiting its utility. In particular, it has the poorest sporulation frequency and highest petite frequency of all common lab strains.

[199] The petite phenotype arises due to a defect in aerobic respiration. Petite yeasts are unable to grow on non-fermentable carbon sources (for example glycerol or ethanol), and form small anaerobic-sized colonies when grown in the presence of fermentable carbon sources (for example glucose). The phenotype results from mutations in the mitochondrial genome, loss of mitochondria, or mutations in the host cell genome.

[200] The genes and single nucleotide polymorphisms (SNPs) responsible for the sporulation and respiration defects have been identified (**Figure 12).** The sporulation defect may be repaired by a series of genetic crosses to a previously repaired strain. The respiration problem caused by mitochondrial genome instability may also be corrected using genome editing. Genome editing may be performed with any method know in the art, such as the 50:50 method. (J. Horecka and R. W. Davis. *Yeast* 31:103-12, 2014). In some cases, an improved version of Mega 50:50 may be used in which a double stranded break is introduced into the genomic locus to be modified, increasing efficiency by several orders of magnitude (J. D. Smith, , et al., *Mol. Syst. Biol.* 13:913, 2017).

[201] In some cases, the host cell may be a cell which has been engineered to repair a sporulation defect. For example, the host cell may be a fungal cell with a repaired sporulation defect. In some cases, the host cell is a yeast cell with a repaired sporulation defect. In some cases, the host cell is a BY yeast cell in which the sporulation defect has been repaired, as in **Figure 12.**

[202] In some cases, the host cell may be a cell which has been engineered to repair a respiratory defect or a mitochondrial genome instability defect. For example, the host cell may be a fungal cell with a repaired mitochondrial stability defect. In some cases, the host cell is a yeast cell with a repaired mitochondrial stability defect. In some cases, the host cell is a BY yeast cell in which the mitochondrial genome instability defect has been repaired, as in **Figure 12.** In some cases, the host cell may be a cell in which both a sporulation defect and a mitochondrial genomic instability defect have been repaired. Using the genetic crosses and genome engineering methods discussed above and the genomic repairs outlined in **Figure 12** the BY strain was engineered to repair the mitochondrial genome instability. An unexpected benefit of repairing the mitochondrial genome instability defect was that the strains grew faster on non-fermentable carbon sources, such as ethanol (see **Figure 13 and 14A).** This is commonly the growth condition of choice for expression of heterologous genes,

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1056 of 3077

which are often linked to production-phase promoters activated by growth on non-fermentable carbon sources.

[203] In some cases, the host cell may be genetically engineered to lack a gene involved in non-homologous end joining. The lack of such a gene may increase the efficacy of homologous recombination in such an engineered cell. The appropriate genes to delete may vary in each host. As an example, an engineered yeast host cell may lack a ligase such as the DNL4 DNA ligase. An engineered bacterial host cell may lack one or both of a Ku homodimer and the multifunctional ligase/polymerase/nuclease LigD. Other genes which may be involved in non-homologous double stranded break repair, depending on species, include: Mrel 1, Rad50, Xrs2, Nbsl, DNA-PKcs , Ku70, Ku80, DNA ligase IV, XLF, Artemis, XRCC4, Dnl4, Lifl, XLF also known as Cernunnos, Nej 1 and Sir2.

[204] DHY strains have utility for expression of heterologous genes for heterologous compound production, as well as ability to perform homologous recombination and DNA assembly. This combination of abilities in one strain allows DNA assembly and production of heterologous compounds in the same strain, whereas previously these two steps were separated in previous types of yeast (BY for DNA assembly and BJ5464 for expression and small molecule production).

[205] These improvements can provide the DHY strain collection with a number of advantages over previous strains, particularly the BY strains: DHY can be faster-growing, result in fewer petite colonies (respiration-deficient), be genetically tractable, allow better expression from ADH2-like promoters, and allow both DNA assembly and production of heterologous products in the same strain **(Figure 14B and 14C)**.

[206] In some embodiments, a genetically engineered host cell lacks one or more conditionally essential genes which can be provided by a plasmid or other DNA vector. This allows for the selection of cells which are expressing the DNA vector. Examples of genes which may be used for this are auxotrophic genes which are required for biosynthesis of certain metabolites or genes for resistance to a toxin. Auxotrophic genes are only required when the specific metabolite they are required for is not present in the culture media. Resistance genes are only required when the toxin which they provide protection from is present.

[207] Examples of genetically engineered yeast host cells include genetically engineered DHY super-host strains. In some cases, strains are based on the BY4741 / BY4742 background (C. B.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1057 of 3077

Brachmann, et al, *Yeast,* 14:1 15-32, 1998). Strains may also contain any of the following genetic changes from the BY background: sporulation repair *(MKT1(30G) RME1(INS-308A) TA03(1493Q)), and* mitochondrial genome stability and function repair *(CAT5(91M) MIP1(661T) SAL1+ HAP1+)* (see **Figure 12).** It should be noted, as would be understood in by persons having ordinary skill in the art, that any and all these genetic changes can be performed in isolation, in part, or in totality. For example, it is expected that the a single genetic change of either *MKT1(30G), RME1(INS-308A), or TA03(1493Q)* would result in at least some repair in sporulation activity. Likewise, a single genetic change of either *CAT5(91M)* or *MIP1(661T)* or restoration of function of *SAL1 (SAL1+) or HAP 1 (HAP1+)* would result in at least some increase in mitochondrial genome stability.

[208] In some cases, a strain may be a prototroph. For example, some strains may require methionine, arginine or lysine in the media. In some cases, a strain may be a full heterozygote for several markers from which any combination of markers can be made by tetrad dissection. For example, heterozygous for genes required for synthesis of histidine, leucine, uracil, lysine and methionine, or heterozygous for genes required for synthesis of histidine, leucine, uracil, lysine and arginine. Some examples of strains are listed in Table 3.

[209] In some cases, the use of a strain as described herein allows for greater expression of BGC proteins and/or greater production of compounds from the BGCs. In some cases expression of heterologous proteins in a strain described herein is accomplished with an efficiency of at least 70%, 80%, 90%, 100%, 110%, 120%, 130%, 140%, or 150% as compared to heterologous protein expression in BJ5464. In some cases production of heterologous compounds in a strain described herein is accomplished with an efficiency of at least 70%, 80%, 90%, 100%, 110%, 120%, 130%, 140%, or 150% as compared to heterologous compound production in BJ5464.

Table 3: Description of strain genotypes

Strain	Parent	Genotype	Reference
BY4741	S288C	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0	(C. B. Brachmann, et al., 1998, cited supra)

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1058 of 3077

BY4743	S288C	MATa/ α his3Al/his3Al leu2A0/leu2A0 LYS2/lys2A0 metl5A0/MET15 ura3A0/ura3A0	(€. B. Brachmann, et a̱l., 1998, cited supra)
BJ5464		MATa ura3-52 trpl Ieu2-Al his3-Δ200 pep4::HIS3 prbl-A1.6R canl GAL	(E. W. Jones, *Methods Enzymol.* 194:428-53, 199̱)
BY4743AD NL4	S288C	MATa/Mata dnl4A/dnl4A	(E. A. Winzeler, et a̱.. *Science* 285.901-06, 1999)
Y800		MATa ade2-1 Ieu2-A98 ura3-52 lys2-801 ïrpï-ï his3-A200 [cirO]	(N. Burns, et a̱., *Genes Dev.* 8:1087-105, 1994}
DHY213	BY4741	MATa his3Al leu2A0 ura3A0 metl5A0 SAL1+ HAP1+ CAT5(91M) MIP1(661T) MKT1(30G) RME1(INS-308A) TA03(1493Q)	n/a
JHY693	DHY213	MATa his3Al leu2A0 ura3A0 metl5A0 SAL1+ HAP1+ CAT5(91M) MIP1(661T) MKT1(30G) RME1(INS-308A) TA03(1493Q) prblA pep4A	n/a
JHY651	DHY213	MATa his3Al leu2A0 ura3A0 metl5A0 SAL1+ HAP1+ CAT5(91M) MIP1(661T) MKT1(30G) RME1(INS-308A) TA03(1493Q) prblA pep4A lys2A0	n/a
JHY692	DHY213	MATa his3Al leu2A0 ura3A0 metl5A0 SAL1+ HAP1+	n/a

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1059 of 3077

		CAT5(91M) MIP1(661T) MKT1(30G) RME1(INS-308A) TA03Q493Q) prblA pep4A ADH2p-npgA-ACSlt	
JHY705	DHY213	MAT a his3Al leu2A0 ura3A0 metl5A0 SAL1+ HAP1+ CAT5(91M) MIP1(661T) MKT1(30G) RME1(INS-308A) TA03(1493Q) prblA pep4A ADH2p-CPR-ACSlt lys2A0	n/a
JHY702	DHY213	MATa/MATα his3Al/his3Al leu2A0/leu2A0 ura3A0/ura3A0 metl5A0/metl5A0 SAL1+/SAL1+ HAP1+/HAP1+ CAT5(91M)/CAT5(91M) MIP1(661T)/MIP1(661T) MKT1(30G)/MKT1(30G) RME1(INS-308A)/RME1(INS-308A) TA03 (1493Q/TA03 (1493Q)) prb1Δ/prb1Δ pep4Δ/pep4Δ ADH2p-npgA-ACS lt/ADH2p-CPR-ACSlt metl5A0/+ lys2A0/+	n/a

Detection and characterization of novel molecules

[210] Once a host cell is expressing the coding sequences of the identified gene cluster, a secondary metabolite may be synthesized in the host cell. The secondary metabolite may be identified by any method known in the art. In some cases, the secondary metabolite is identified by comparing a host cell expressing the cluster with a host cell which does not express the cluster. This comparison may utilize chromatography methods to separate different small molecules produced in

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1060 of 3077

the cells. For example, column chromatography, planar chromatography, thin layer chromatography, gas chromatography, liquid chromatography, supercritical fluid chromatography, ion exchange chromatography, size exclusion chromatography be done by high performance liquid chromatography (HPLC), mass spectrometry (MS), or by mass spectrometry high performance liquid chromatography (MS-HPLC). Any peaks which appear for the cluster expressing host cell and not from the control host cell indicate the presence of a novel chemical. The comparison between the cluster expressing host cell and the control host cell may comprise a comparison of a cell extract, a culture media, or an extracted cell lysate.

Compounds identified

[211] This disclosure also provides sequences of 43 BGCs, and structures of novel products produced by a subset of these BGCs.

[212] In one embodiment, this disclosure provides sequences of cryptic BGCs which encode various products, **SEQ ID NOs: 67-483.** These BGCs may also be reengineered to provide the coding sequences without the endogenous regulatory sequences. In some examples, the coding sequences may be predicted using known bioinformatics methods, experimental data, or obtained from databases such as default predicted gene coordinates (start, stop, and introns) as deposited in GenBank. Once the coding sequences have been identified the sequences may be isolated and cloned into one or more expression vectors for expression in a model host system such as *S. cerevisiae.*

[213] The expression vectors may be plasmids, viruses, linear DNA, bacterial artificial chromosomes or yeast artificial chromosomes. Each of the one or more expression vectors may contain one or more promoters suitable for expression of a heterologous gene in a model host system. Each expression vector may contain a single coding sequence or multiple coding sequences. Multiple coding sequences may be functionally linked to a single promoter, for example via an internal ribosome entry site, or may be linked to multiple promoters. The expression vectors may also contain additional elements to regulate or increase the transcriptional activity, for example enhancers, polyA sequences, introns, and posttranscriptional stability elements. The expression vectors may also contain one or more selectable markers.

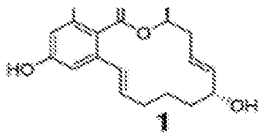
[214] The expression vectors may be transfected, or otherwise introduced, into host cells. Examples of host cells include but are not limited to yeast and bacterial cells. For example a host

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1061 of 3077

cell may be a *S. cerevisiae* cell or an *E. coli* cell. Incubating the expression vectors in the host cells allows for the transcription and translation of the coding sequences to recreate the proteins of the gene cluster. These proteins may then produce a secondary metabolite which can be isolated from the cells or the media in which the cells are grown.

[215] In another embodiment this disclosure provides host cell extracts containing non-host cell derived products. In some cases, these extracts may be produced by culturing host cells expressing one or more, or all, of the sequences from one of the following groups **SEQ ID NOs: 67-76, 77-81, 82-91, 92-97, 98-106, 107-111, 112-118, 119-127, 128-135, 136-153, 154-157, 158-162, 163-172, 173-181, 182-186, 187-191, 192-199, 200-206, 207-211, 212-224, 225-228, 229-235, 236-240, 241-244, 245-255, 256-267, 268-276, 277-285, 286-289, 290-293, 294-307, 308-313, 314-318, 319-324, 325-329, 330-334, 335-341, 342-350, 351-357, 358-367, 368-372, 373-380, 381-388, 389-395, 396-400, 401-406, 407-413, 414-423, 424-427, 428-439, 440-447, 448-453, 454-462, 463-471, 472-480, or 481-483.** In some cases, a host cell may express all the sequences from one of the following groups **SEQ ID NOs: 67-76, 77-81, 82-91, 92-97, 98-106, 107-111, 112-118, 119-127, 128-135, 136-153, 154-157, 158-162, 163-172, 173-181, 182-186, 187-191, 192-199, 200-206, 207-211, 212-224, 225-228, 229-235, 236-240, 241-244, 245-255, 256-267, 268-276, 277-285, 286-289, 290-293, 294-307, 308-313, 314-318, 319-324, 325-329, 330-334, 335-341, 342-350, 351-357, 358-367, 368-372, 373-380, 381-388, 389-395, 396-400, 401-406, 407-413, 414-423, 424-427, 428-439, 440-447, 448-453, 454-462, 463-471, 472-480, or 481-483.** In some cases the host cell(s) may express one or more sequences selected from **SEQ ID NOs: 67-483.** After culturing the cells, they may be collected, lysed, and the small molecules may be purified from nucleic acid, proteins, complex carbohydrates and lipid containing fractions. The secondary metabolites produced may also be secreted into the cell media. In this case this disclosure also provides a cell media containing secondary metabolites.

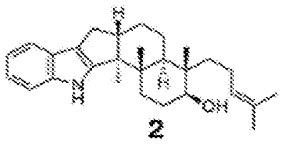
[216] This disclosure also provides compounds isolated from the host cell extracts or media. A compound of this disclosure may be Compound 1:



1

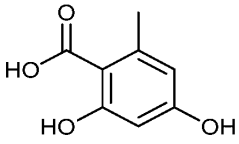
LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1062 of 3077

[217] Compound 2:



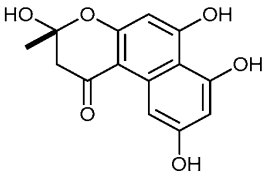
2

[218] Compound 3:



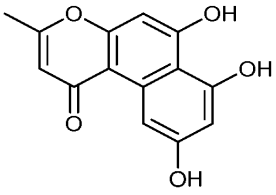
(orsellinic acid),

[219] Compound 4:



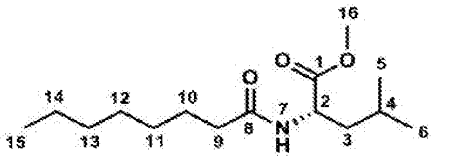
,

[220] Compound 5:



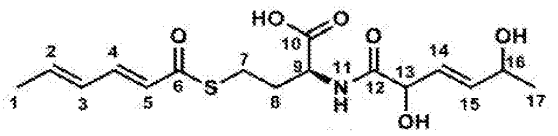
,

[221] Compound 6:



,

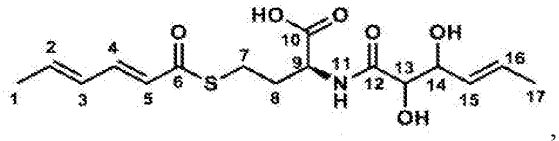
[222] Compound 7:



,

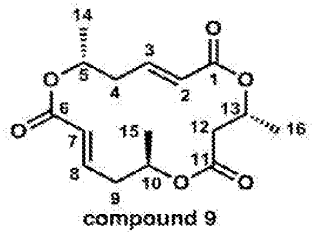
LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1063 of 3077

[223] Compound 8:



,

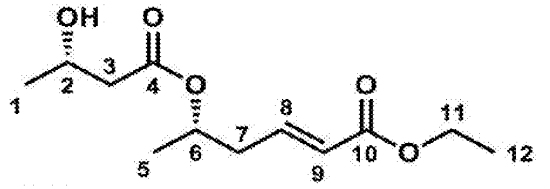
[224] Compound 9:



compound 9

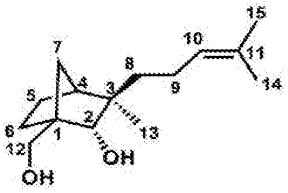
,

[225] Compound 10:



,

[226] Compound 11:

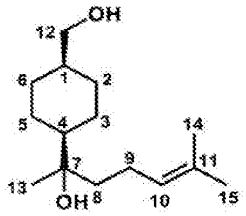


compound 11

,

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1064 of 3077

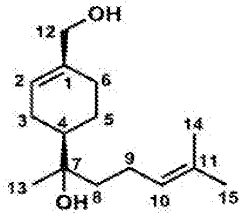
[227] Compound 12:



compound 12

,

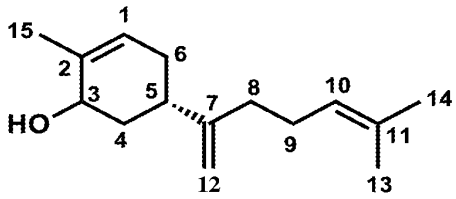
[228] Compound 13:



compound 13

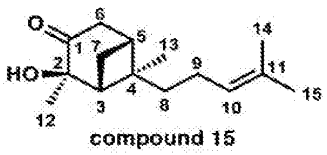
,

[229] Compound 14:



compound 14

[230] Compound 15:

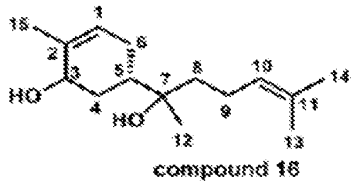


compound 15

, and

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1065 of 3077

[231] Compound 16:



compound 16

[232] Compounds of this disclosure may have useful therapeutic applications, for example in treating or preventing a disease or disorder. Compounds of this disclosure may be used to treat an infection, for example a bacterial, fungal or parasitic infection. Compounds of this disclosure may have antibiotic and/or antifungal activities. Compound 8 and Compound 9 may have antimicrobial, antifungal and/or antibacterial activities. Compounds of this disclosure may have non-medical applications.

[233] In some embodiments this disclosure provides pharmaceutical compositions comprising of a compound of this disclosure. In some cases a pharmaceutical composition contains at least one of Compounds: 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, and 16. Compositions as described herein may comprise a liquid formulation, a solid formulation or a combination thereof. Non-limiting examples of formulations may include a tablet, a capsule, a gel, a paste, a liquid solution and a cream. The compositions of the present disclosure may further comprise any number of excipients. Excipients may include any and all solvents, coatings, flavorings, colorings, lubricants, disintegrants, preservatives, sweeteners, binders, diluents, and vehicles (or carriers). Generally, the excipient is compatible with the therapeutic compositions of the present disclosure. Generally, the excipient is a pharmaceutically acceptable excipient. The pharmaceutical composition may also contain minor amounts of non-toxic auxiliary substances such as wetting or emulsifying agents, pH buffering agents, and other substances such as, for example, sodium acetate, and triethanolamine oleate.

[234] In some embodiments this disclosure provides a method of synthesizing a compound described herein. The method may include steps of providing one or more coding sequences of **SEQ ID NOs: 67-483** in a suitable vector, or vectors, together with regulatory sequences which will drive expression of the coding sequences in a host cell. The vector, or vectors, are then provided to a host cell, such as for example a yeast cell, and the cells are grown under conditions that allow for the expression of the coding sequences. In some cases, a host cell may be provided with 1, 2, 3, 4, 5, 6,

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1066 of 3077

or more than 6 different plasmids. The synthesized compound may be purified from the cell culture by centrifuging the cells to produce a cell pellet and a supernatant. The supernatant and cell pellet may be extracted using either ethyl acetate or acetone, or other suitable organic solvent. For compounds containing carboxylic acid groups, the pH of the supernatant may be adjusted to pH of 4 or less (e.g., 3) with an acid such as HC1 prior to extraction. After extraction, both organic phases are combined and evaporated to dryness. The compounds may then be dissolved in the desired solvent and further purified using standard purification methods.

[235] Although the present disclosure has been described in certain specific aspects, many additional modifications and variations would be apparent to those skilled in the art. In particular, any of the various processes described above can be performed in alternative sequences in order to achieve similar results in a manner that is more appropriate to the requirements of a specific application. It is therefore to be understood that embodiments of the present disclosure can be practiced otherwise than specifically described without departing from the scope and spirit of the present disclosure. Thus, embodiments of the present disclosure should be considered in ail respects as illustrative and not restrictive.

Example 1: Identification of production phase promoters

[236] Biological data supports the systems and constructs of production-phase promoter DNA vectors and applications thereof. Provided below are several examples of incorporating production-phase promoters into DNA vectors. Some of these vectors were used to produce biosynthetic products from multi-gene clusters derived from various fungal species. Compared to a constitutive promoter system, production-phase promoter systems in accordance with various embodiments produced several-fold greater product.

Production Phase Promoter Expression Analysis

[237] Because the ADH2 promoter **(SEQ ID NO. 1)** has properties of a production-phase promoter, a panel of promoter sequences was compared to the ADH2 promoter to identify other production-phase promoters. To begin, endogenous *S. cerevisiae* genes were identified that appeared co-regulated with ADH2 in a previous genome-wide transcription study (Z. Xu. *et al, Nature* 457: 1033-37, 2009, the disclosure of which is incorporated herein by reference). In this study, transcription of yeast genes was quantified during mid-exponential growth in several types of

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1067 of 3077

growth media. Of the 5171 ORFs examined, 35 appeared co-regulated with ADH2, with co-regulation defined as a greater than two-fold increase in expression with a non-fermentable carbon source (ethanol in a yeast-peptone-ethanol (YPE) media) as compared to a fermentable carbon source (dextrose in a yeast-peptone-dextrose (YPD) media). Because these data were collected at a single time point and assessed transcription of genes in their native context, their ability to co-regulate heterologous genes in a production-phase promoter system required further validation and characterization.

[238] A detailed characterization of the ability of 34 selected promoters to control expression of heterologous genes was performed. For this specific purpose, a promoter was defined as the shorter of (a) 500 bp upstream of the start codon, or (b) the entire 5' intergenic region. Each promoter was cloned upstream of the gene for monomeric enhanced GFP (eGFP) and integrated each of the resulting cassettes in a single copy at the *ho* locus of individual strains. Control strains were included in which strong constitutive FBA1 and TDH3 promoters were cloned upstream of eGFP in an identical manner. The 35 promoter sequences can be found in **SEQ ID NOs. 2-35.**

[239] In order to compare the 35 putative production-phase promoters, the expression of eGFP protein was assessed over 72 hours in each strain by flow cytometry in media with both fermentable (YPD) and non-fermentable (YPE) carbon sources (**Figures 16** and **17).** All cultures were started in YPD media and analysis of eGFP expression began when cells were in the midst of exponential fermentative growth ($\mathbf{OD_{600}}$ = 0.4, 0 hrs). At this point, cells were either left to continue growth in YPD or spun-down and resuspended in YPE. Consistent with previous work, pADH2 was entirely repressed at the point where the experiment commenced (during exponential fermentative growth, 0 hrs) unlike the constitutive promoters pTDH3 and pFBA1, which were expressed at near maximum levels regardless of phase. Moderate expression from pADH2 was observed after a further 6 hours in YPD culture or following a growth media switch to YPE. Within 24 hrs, expression reached levels exceeding those observed in the strong constitutive systems. Cytometry histograms and fluorescence microscopy demonstrated that within 48 hours, >95% of all cells with pADH2 and pPCK1 driven expression were fluorescing above background (**Figure 18).** Protein expression levels spanned 15-50 fold, with most showing little or no expression until 24 hours into the culture (**Figures 16 and 17).** Transgene expression driven by the PCK1, MLS1, and ICL1 promoters (**SEQ ID NOs. 2-4)** not only showed the same timing of expression as pADH2, but also expressed at an

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1068 of 3077

equivalently high level. The promoters of genes YLR307C-A, YGR067C, IDP2, ADY2, GAC1, ECM13 and FAT3 **(SEQ ID NOs. 5-11)** displayed semi-strong transgene expression **(Figure 16).** In addition, the promoters of genes PUT1, NQM1, SFC1, JEN1, SIP18, AT02, YIG1, and FBP1 **(SEQ ID NOs. 12-19)** displayed weak of transgene expression **(Figures 16 and 17).** The promoter PH089 **(SEQ ID NO. 20)** did not exhibit strong repression in during the growth phase **(Figure 16,** 0 and 6 hours). The results of the other sequences are also depicted in **Figure 16 (SEQ ID NOs. 22-36).** The constitutive promoters pTDH3 and pFBA1 **(SEQ ID NOs. 50 and 52)** were used as controls **(Figures 16-18).**

[240] The above analysis identified a large set of co-regulated promoters spanning a wide range of expression levels, three of which were as strong as pADH2. However, a more extensive set of strong production-phase promoters is desirable for assembly of constructs having multi-gene pathways, especially pathways having more than four genes. To identify other production-phase promoter candidates, the genomes of five closely related species within the *S. sensu stricto* complex were examined **(Figure 19).** The promoter region was identified for the closest ADH2 gene homolog in the genomes of *Saccharomyces bayanus, Saccharomyces paradoxus, Saccharomyces mikitae, Saccharomyces kudriavzevii,* and *Saccharomyces castellii.* Multiple sequence alignment of the upstream activation sequences (UAS) revealed that nearly all sequences (except that from *S. castellii)* are highly conserved across this region, suggesting a potential for regulation similar to that of *S. cerevisiae* ADH2 **(Figure 20, SEQ ID NOs. 36-40).** In order to be used for single-step pathway assembly, all promoter sequences must be sufficiently unique to prevent undesired recombination between each other. Therefore, the pairwise identities for each of the *Saccharomyces sensu stricto* ADH2 promoter pairs were analyzed **(Figure 21).** The most similar promoter to the *S. cerevisiae* ADH2 promoter is that from *S. paradoxus,* with 83% identity, including a single 40 bp stretch located near the center of the promoter. This homology is significantly less than the 50-100 bp typically used for assembly by yeast homologous recombination, and recombination events between sequences with this level of identity occur at very low frequency, suggesting that these promoters should be compatible with a multi-gene assembly technique utilizing yeast homologous recombination as described above.

[241] As with the endogenous yeast promoter candidates, these other putative *Saccharomyces* promoters required detailed characterization of induction profiles. DNA encoding each of these

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1069 of 3077

promoter sequences was obtained by commercial synthesis and characterized expression of eGFP from each promoter in the same manner as the endogenous yeast promoters **(Figures 22** and **23).** Of the five *Saccharomyces sensu stricto* pADH2s tested **(SEQ ID NOs. 36-40),** the promoters derived from *S. paradoxus, S. kudriavzevii,* and *S. bayanus* show timing and strength of expression equivalent to that of *S. cerevisiae* pADH2. In combination with the endogenous yeast promoters, these three additional *Saccharomyces* pADH2s expand the number of strong promoters with the desired induction profile.

Expression of Compound Product Pathways Using the Production-Phase Promoter System

[242] To study the utility of the new promoter set for heterologous expression of a biosynthetic system, production of fungal-derived dehydrozearalenol (1) and indole-diterpene (2) was examined **(Figure 24,** Compounds 1 & 2). The biosynthesis of the indole-diterpene compound resulted from the coordinated expression of four in *Aspergillus tubingensis* genes **(Figure 25, SEQ ID NOs. 59-62).** Two versions of each pathway were constructed: one having all production-phase promoters, and the other having all constitutive promoters **(Figure 24).** The production-phase promoter system utilized the pADH2 from *S. cerevisiae* **(SEQ ID NO. 1),** pADH2 from *S. bayanus* **(SEQ ID NO. 38),** and pPCK1 **(SEQ ID NO. 2)** and pMLS1 **(SEQ ID NO. 3)** from *S. cerevisiae.* In the constitutive system, transcription was driven by four frequently used strong constitutive promoters: pTEF1, pFBA1, pPCK1, and pTPI1 **(SEQ ID NOs. 51-54).** Each indole-diterpene system was constructed on a single plasmid harboring four expression cassettes: promoter: :GGPPS::tADH2; promoter: :PT::tPGI1; promoter: :FMO::tEN02; and promoter: :Cyc::tTEF1; wherein, the promoter sequences corresponded to either the production-phase or the constitutive promoters **(Figure 24).** Similar constructs were built for the dehydrozearalenol compound with the two genes HR-PKS and NR-PKS **(SEQ ID NOs. 63** and **64).** All plasmids were constructed using yeast homologous recombination. It should be noted that pADH2 sequences from *S. cerevisiae* and *S. bayanus* (61% identity) are sufficiently unique for this type of assembly. The production of compounds 1 and 2 produced by *S. cerevisiae* BJ5464/«pgA/pRS424 transformed with each of these plasmids were measured over seventy-two hours in YPD batch culture **(Figure 26).** An 80-fold and 4.5-fold increase in titer of compound 1 and 2 was observed for the system using the production-phase promoters as compared to the constitutive system.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1070 of 3077

Materials and Methods Supporting the Production-Phase Promoter Experiments

[243] General techniques, reagents, and strain information: Restriction enzymes were purchased from New England Biolabs (NEB, Ipswich, 25 MA). Cloning was performed in *E. coli* DH5a. PCR steps were performed using Q5® high-fidelity polymerase (NEB). Yeast dropout media was purchased from MP Biomedicals (Santa Ana, CA) and prepared according to manufacturer specifications. Promoter characterization experiments were performed in BY4741 *(MATa, his3Al leu2A.O metl5A.O ura3A0)* while all experiments involving the production of 1 were performed in *BJ5464-npgA* which is BJ5464 *(MATaura3-52 Ms3A200 leu2A.l trplpep4::HIS3 prbl.A.1.6R canl GAL)* with two copies of *pADH2-npgA* integrated at δ elements. All Gibson assemblies were performed as previously described using 30 bp assembly overhangs.

[244] Construction and characterization of promoter-eGFP reporter strains: All promoters were defined as the shorter of 500 base pairs upstream of a gene's start codon or the entire 5' intergenic region. All promoters from *S. cerevisiae* were amplified from genomic DNA, while ADH2 promoters from all *Saccharomyces sensu stricto* were ordered as gBlocks from Integrated DNA Technologies (IDT, Coralville, Iowa). Minimal alterations were made to promoters from *S. kudriavzevii* and *S. mikitae* in order to meet synthesis specifications. In all constructs, eGFP was cloned directly upstream of the terminator from the CYC1 gene (tCYCl). pRS415 was digested with Sacl and Sail and a Notl-eGFP-tCYCl cassette was inserted by Gibson assembly generating pCH600. Digestion of pCH600 with Accl and Pmll removed the CEN/ARS origin, which was replaced by 500 bp sequences flanking the ho locus using Gibson assembly to yield plasmid pCH600-HOint. Each of the promoters to be analyzed was amplified with appropriate assembly overhangs and inserted into pCH600-HOint digested with Notl to generate the pCH601 plasmid series. Digestion of the pCH601 plasmid series with Ascl generated linear integration cassettes which were transformed into *S. cerevisiae* BY4741 by the LiAc/PEG method. Correct integration was confirmed by PCR amplification of promoters and Sanger sequencing.

[245] For characterization, all strains were initially grown to saturation overnight in 100 μĭ of YPD media. These cells were then reinoculated at an $\mathbf{OD_{600}}$ of 0.1 into 1ml of fresh YPD and allowed to grow to $\mathbf{OD_{600}}$ = 0.4 to reach mid-log phase growth (approximately 6hrs). 500 μĭ of each culture was pelleted by centrifugation and resuspended in YPE broth for YPE data while the remaining 500 μĭ was used for YPD data. The 0 hour time point was collected immediately after resuspension. For

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1071 of 3077

each time point, 10 μΐ of culture was diluted in 2 ml of DI water and sonicated for three short pulses at 35% output on a Branson Sonifier. Expression data were collected for 10,000 cells using a FACSCalibur flow cytometer (BD Bioscience) with the FL1 detector. Data were analyzed in R using the flowCore package.

[246] Construction of plasmids to produce compounds in *S. cerevisiae:* The sequences for genes assembled on IDT producing plasmids are contained in the supporting information. Regulatory cassettes of promoters and terminators were fused using overlap extension PCR. All genes and regulatory cassettes were amplified by PCR, ensuring 60 bases of homology between all adjacent fragments. 500 ng of each purified fragment was combined with 100 ng of pRS425 linearized with *NotI* and transformed into *S. cerevisiae* BJ5464/*npg*A. Sixteen clones were picked from each assembly plate and grown to saturation in 5 ml CSM-Leu medium. Plasmids were isolated, transformed into *E. coli* and purified prior to sequence confirmation using the Illumina MiSeq platform. Detailed plasmid maps for pCHIDT-2. land pCHIDT-2c are shown in **Figure 27A** illustrates the primers used and the assembly strategy **(SEQ ID NOs. 65 and 66).**

[247] Examining the productivity of indole diterpene generating systems Plasmids pCHIDT-2. 1 and pCHIDT-2c were transformed into BJ*5464/npgA* with pRS424 as a source of tryptophan overproduction (see, e.g., **Figure 27B).** Triplicates of each strain were inoculated into CSM - Leu/-Trp medium and grown overnight (OD_{600} = 2.5-3.0). Each culture was used to inoculate 20 ml cultures in YPD medium at an OD_{600} = 0.2 and incubated with shaking at 30°C for 3 days. Every 24 hrs, 2 mis were sampled from each culture. Supernatants were clarified by centrifugation and extracted with 2 ml ethyl acetate (EtOAc). Cell pellets were extracted with 2 ml 50% EtOAc in acetone. 500 μΐ each of pellet and supernatant extracts were combined and dried *in vacuo.* Samples were resuspended in 100 μΐ HPLC grade methanol and LC-MS analysis was conducted on a Shimadzu LC-MS-2020 liquid chromatography mass spectrometer with a Phenomenex Kinetex C18 reverse-phase column (1.7 μπι, 100 **A,** 100 mm x 2.1 mm) with a linear gradient of 15% to 95% acetonitrile (v/v) in water (0.1% formic acid) over 10 min followed by 95% acetonitrile for 7 min at a flow rate of 0.3 mL/min.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1072 of 3077

Example 2: Identification of gene clusters which produce compounds that interact with a target protein

[248] Thanks to next-generation sequencing, thousands of bacterial and fungal genomes have been sequenced. These species are known to be rich sources of **secondary** metabolites, for example penicillin, rapamycin, and the statins. These secondary metabolites are small molecules, **enzymatically** synthesized **by** the products of one or more genes, often arrayed contiguously **in** a "biosynthetic gene cluster".

[249] This disclosure describes a method for identifying specific biosynthetic gene clusters, where the target of the **secondary** metabolite is a specific protein, and expressing that secondary metabolite in a host organism.

[250] **In** certain cases, for example, when a secondary metabolite is being used as a weapon against other organisms, the **secondary** metabolite may also be toxic to the organism that produces it. **In** these cases, the producing organism may defend itself against **self-harm** in a number of ways: by pumping the **secondary** metabolite out of the cell; by enzymatically negating the **secondary** metabolite; or **by** producing an additional version of **the** target protein that is less sensitive or insensitive to the secondary metabolite.

[251] In those cases where **the** organism produces an additional version of the target protein, this "protective" version of the gene is often **colocalized** with the biosynthetic gene cluster. Although different to the gene that produces the target protein, the protective version should maintain detectable homology to the target protein. This method takes advantage of this homology to **identify** those biosynthetic gene clusters that contain or are adjacent to a protective homolog of the target protein.

[252] The input data required for this method are a list of biosynthetic gene clusters (e.g., **polyketide** synthase clusters, **non-ribosomal** peptide synthetase clusters) and a **list of target** proteins (i.e., proteins whose activity are to be modulated with secondary metabolites). The biosynthetic clusters may be identified based on the presence of certain protein domains, for example by the software program antiSMASH. The target proteins may be chosen based on their quantitative likelihood of being drug targets.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1073 of 3077

[253] The output of this method is a score for each biosynthetic cluster, where clusters with higher scores represent those that are more likely to produce secondary metabolites targeting specific proteins of interest.

[254] A score was constructed for each biosynthetic cluster based on the following factors:

1. the presence of one or more homologs of a target protein within or adjacent to the cluster, as determined by a homology search (e.g., using the tblastn algorithm, with a maximum score granted when one homolog is found)
2. the confidence in homology of the target to genes in a cluster (e.g., according to the tblastn algorithm, with an increasing score for lower e-values, and an upper bound threshold of le-30)
3. the fraction of the homologous gene that meets a certain threshold of identity (e.g., with an increasing score for more identity, and a lower bound threshold of 25% identity)
4. the total number of genes homologous to the target protein present in the entire genome of the organism (e.g., with a maximum score granted to cases with 2-4 homologs per genome)
5. the homology of the gene in or adjacent to the cluster to the target protein (e.g., using the blastx algorithm, with a maximum score granted when the gene in the biosynthetic gene cluster's closest homolog in the target protein's genome is the target protein itself)
6. the phylogenetic relationship of the target protein to the gene in the cluster (e.g., with an increasing score for homologs in the gene cluster that clade with the target protein, with confidence assigned by a bootstrap test or Bayesian inference of phylogeny, and a lower bound threshold defined as homologs in a phylogenetic context that appear in a clade with bootstrap value of 0.7 or Bayesian posterior probability of 0.8)
7. the expected number of homologs of the target in or adjacent to the biosynthetic cluster (e.g., with a greater score the lower the probability of a homolog of the target being present in or adjacent to a biosynthetic cluster of a certain size, given the number of total homologs in the genome, as determined by a permutation test)

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1074 of 3077

8. the likelihood that the target protein is essential for viability, growth, or other cellular processes in the native environment, e.g., through evidence that deletion of homologs in related organisms (such as *S. cerevisiae)* render the organism inviable

9. synteny of the gene cluster with related species (e.g., with a maximum score if the entire cluster, including the target homolog, is conserved across several species)

10. the functional class of the target homolog (e.g., with a greater score if the gene is in a protein complex already known to be targeted by secondary metabolites)

11. the presence of specific promoters adjacent to the target homolog (e.g., with a greater score when there is a bidirectional promoter upstream of the target homolog and a biosynthetic gene)

12. the presence of specific regulatory elements in the biosynthetic gene cluster (e.g., with a greater score when there is a transcription factor binding site that is shared between target genes and/or biosynthetic genes in the cluster)

13. the presence of target homologs outside the cluster (including on other chromosomes) that are co-regulated with some or all of the genes in the biosynthetic cluster (e.g., with a greater score when biosynthetic gene clusters are co-regulated with putative target homologs)

14. the presence of protein- and DNA-sequence-derived features within the clusters that have successfully been shown to produce secondary metabolites (e.g., with a greater score when a gene in a cluster shares a domain — as determined by a Hidden Markov Model (HMM) — with a cluster that has produced a secondary metabolite in one of the host organisms)

[255] The above score was calibrated with reference to a set of "true positives" (i.e., cases where there are one or more known targets in or adjacent to a biosynthetic gene cluster that produces a small molecule known to target that protein).

[256] This algorithm has been programmed in the Python programming language and has been applied to a set of more than 1,000 fungal genomes (and more than 10,000 biosynthetic gene clusters) to produce a list of potentially relevant biosynthetic clusters.

Expressing biosynthetic clusters in a host organism

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1075 of 3077

[257] Given the highest scoring biosynthetic clusters as defined by the algorithm above, DNA was synthesized for each of the genes in those clusters. The DNA was cloned into a host organism (e.g., *S. cerevisiae,* also known as baker's yeast) for expression. The host organism synthesized the proteins from the gene cluster, which produces a secondary metabolite. Using HPLC and mass spectrometry, the secondary metabolite-expressing strain can be compared to an unmodified strain and affirm the presence of a new secondary metabolite.

[258] This method has been successfully applied to the production of several secondary metabolites where there is evidence, based on the above method, for what the target protein of the secondary metabolite should be:

* An secondary metabolite derived from a biosynthetic gene cluster containing a homolog of the human gene SOS1;
* An secondary metabolite derived from a biosynthetic gene cluster containing a homolog of the human gene BRSK1; and
* An secondary metabolite derived from a biosynthetic gene cluster containing a homolog of the human gene DDX41 .

[259] A further example of a gene cluster which produces a product, for which there is evidence suggesting the target, is shown in Figure ISA.

Example 3: Prioritization of novel biosynthetic gene clusters by phylogenetic analysis.

[260] Two classes of fungal BGCs; those with either a polyketide synthase (PKS), or an UbiA-type sesquiterpene cyclase (UTC) as their core enzyme were chosen for analysis.

[261] A computational pipeline was developed to prioritize PKS and UTC containing BGCs for heterologous expression. 581 sequenced fungal genomes were analyzed from the publicly available GenBank database of the National Center for Biotechnology Information (NCBI, as of July 2015). Each genome was analyzed for BGCs using antiSMASH2, identifying 3512 BGCs harboring an iterative type *I* PKS (iPKS) and 326 BGCs harboring a UTC homologue. Phylogenetic trees of each of these enzyme types were generated with identified characterized homologs from the MIBiG database20. BGCs were primarily selected from clades having few characterized members (**Figure** 28A, Figure 29A). The selected BGCs were found in the genomes of both ascomycetes and basidiomycetes. Basidiomycetes are, in general, more difficult to culture with fewer tools for genetic

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1076 of 3077

manipulation available as compared to ascomycetes. As a result, BGCs from basidiomycetes are under- studied, with few PKS-containing clusters deposited in MIBig, suggesting that these organisms represent a reservoir of BGCs capable of producing compounds with interesting new-structures.

[262] The coding sequences of all BGCs were ordered as synthetic constructs according to the default predicted gene coordinates (start, stop, and introns) as deposited in GenBank and clusters are described in **Table 4.**

[263] Shown in **Figure 28A** is a cladogram of the ketosynthase sequences of the 3512 iPKS sequences identified in this study. Of these, 28 were selected and the associated BGC containing the selected iPKS was analyzed using heterologous expression. Selected BGCs met the following criteria: (a) genetic structure was conserved across 3 or more species, (b) exhibited canonical domain architecture, and (c) contained an in- cis or proximal in- trans protein capable of releasing the polyketide from the carrier protein of the PKS **(Figure 30A).** Seven of these clusters were derived from distinct clades comprised entirely of sequences from basidiomycetes **(Figure 28A).**

[264] The 28 selected PKS clusters were edited according to the methods described here to form expression vectors suitable for expression of the cluster coding sequences in yeast cells. The host cells were incubated and analyzed for the presence of novel chemical compounds by HPLC, as described in the methods section below. Of the PKS clusters selected from ascomycetes, 13 produce compounds. The most notable is the PKS1 cluster, which only contains an iPKS, a hydrolase, and the genes for three tailoring enzymes: a Cytochrome p450 (P450), a Flavin-dependent monooxygenase (FMO), and a Short-chain dehydrogenase/reductase (SDR).

[265] For the study of fungal UTCs, the phylogenetic tree shown in **Figure 29A** was constructed based on the UbiA-type sesquiterpene cyclase, Fma-TC, from the fumagillin biosynthetic pathway. Moreover, the P450, Fma-P450, from the same pathway was shown to be a powerful enzyme catalyzing the 8 e oxidation of bergamotene to generate a highly oxygenated product. UTC BGCs spanning the entirety of the cladogram were selected in **Figure 29A** where a cytochrome P450 was proximal to the UTC gene **(Figure 30B).** Ultimately, 13 UTC BGCs from both ascomycetes and basidiomycetes were selected for analysis.

[266] Screening of strains expressing these clusters by LC/HRMS revealed novel spectral features consistent with oxidized sesquiterpenoids being produced by five clusters **(Figure 29A).** These

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1077 of 3077

results demonstrate that the membrane-bound UTCs represent a general class of terpene cyclase encoded by the genomes of diverse fungi. Several clusters and compounds produced are shown in **Figures 28B, 28C, 28D, 28E** and **28F**.

[267] Including both PKS and UTC BGCs, 24 of the 41 clusters produced measurable compounds, see **Table 4** for a summary of the type, species of origin, and productivity of the clusters. Gene annotation errors introduced by incorrect intron prediction may have contributed to this failure rate. Manual inspection of one UTC (TC5) that initially had yielded no products suggested an incorrect intron prediction at the 5' terminus of the gene. Correction of this intron led to a C-terminal protein sequence that aligned well with known functional UTCs. When tested by heterologous expression in a host cell, the version with the corrected intron produced a compound confirming that incorrect intron prediction is a failure mode in approaches that rely on publicly available gene annotations, **(Figure 29B).** These results illustrate the importance of careful gene curation and the need for improved eukaryotic gene prediction, particularly with sequences from taxa with few well-studied members.

[268] The results summarized in **Table 4** demonstrate the utility of the methods herein for the selection of cryptic fungal BGCs. With the tools developed here, strains were built expressing 41 such clusters with 22 (54%) producing detectable levels of products not native to *S. cerevisiae.* While both basidiomycetes and ascomycetes are known to be prolific producers of bioactive compounds, to date, the bulk of research on the biosynthesis of fungal natural products has been undertaken in ascomycetes. In this study, heterologous expression allowed a large-scale survey of cryptic fungal BGCs from both ascomycetes and basidiomycetes, a less studied and more difficult to culture division of fungi with fewer tools for genetic manipulation. Using this platform, a panel of new products produced by the selected PKS and UTC clusters was identified.

Methods

[269] antiSMASH2 software was applied to 581 public fungal genomes deposited in the Genbank database of the National Center for Biotechnology Information (NCBI), to search for type 1 PKS and UbiA-iike terpene cyclase gene clusters. This analysis identified 3,512 type 1 PKS gene clusters and 326 UbiA-like terpene gene clusters in 538 fungal genomes.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1078 of 3077

[270] Phylogenetic analysis of both sequence sets was performed by building multiple sequence alignments of ail protein sequences using MAFFT and building phylogenetic trees as shown in Figure 28A and Figure 29A using FastTree 2.

[271] 28 of the 3,512 sequenced type 1 PKS gene clusters and 13 of the 326 terpene gene clusters were selected for expression in yeast as described above.

Construction and culture of production strains:

[272] Production strains were constructed by transforming piasmid DNA isolated out of *E. coli* (Qiagen miniprep 27106) into the appropriate expression host (JHY692 for PKS containing plasmids, JHY705 for all others) using the Frozen-EZ Yeast Transformation Iï kit (Zymo Research T2001) followed by plating on the appropriate SDC dropout media (CSM -Leu for PKS containing plasmids, CSM -Ura for all others). For BGCs encoded on at least two plasmids, three biological replicates for each haploid transformant were mated on YPD plates and incubated at 30°C for 4-16 hrs prior to streaking for single colonies on CSM -Ura/-Leu and incubated at 30°C.

[273] Small-scale cultures for analysis were begun by picking three biological replicates of each production strain along with empty vector controls into 500 µL of the appropriate SDC dropout medium in a 1 ml deep-well block and grown for approximately 24 hrs at 30°C. 50 µL of overnight culture was used to inoculate 500 µL of each of the production media to be tested in the experiment (generally both YPD and YPEG) in 1 ml deep well blocks. All blocks were covered with gas-permeable plate seals (Thermo Scientific AB-0718) and incubated at 30°C for 72 hrs with shaking at 1000 rpm. Supernatants were clarified by centrifugation for 20 mins at 2800 g and a minimum of 100 µl of clarified supernatant was stored for future analysis. The remainder of the supernatant was discarded and the cell pellets extracted by mixing with 400 µL of 1:1 ethyl acetate:acetone. Cell debris was precipitated by centrifugation for 20 mins at 2800 g and 200 µL of the extraction solvent pipetted to a fresh block and evaporated in a speedvac.

[274] Prior to analysis, all supernatants were passed through a 0.2 µm filter plate while ail cell pellet extracts were resuspended in 200 µl of HPLC grade methanol prior to filtering.

Analysis of small scale cultures:

[275] LC-MS analysis was conducted on an Agilent 6545 quantitative time-of-flight mass spectrometer interfaced to an Agilent 1290 HPLC system. The ion source for most analyses was an 73eiectrospray ionization source (dual-inlet Agilent Jet Stream or "dual AJS"). In some analyses, an

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1079 of 3077

Agilent Multimode Ion Source was also used for atmospheric pressure chemical ionization. The parameters used for both ionization sources are outlined in Table 5.

[276] The HPLC column for ail analyses was a 50 mm x 2.1 mm Zorbax RRHD Eclipse C18 column with 1.8 μη beads (Agilent, 959757-902). No guard column was used.

[277] Gradient conditions were isocratic at 95% A from 0 to 0.2 min, with a gradient from 95% A to 5% A from 0.2 to 4.2 minutes, followed by isocratic conditions at 5% A from 4.2 to 5.2 minutes, followed by a gradient from 5% A to 95% A from 5.2 to 5.2 minutes, followed by isocratic re-equilibration at 95% A from 5.2 to 6 minutes. For electrospray analyses, A was 0.1% v/v formic acid in water and B was 0.1 % v/v formic acid in acetonitrile. For APCI analyses, B was substituted by 0.1% v/v formic acid in methanol.

[278] Data analysis by untargeted metabolomics was performed with xcms, using optimal parameters determined by IP025. For PKS containing clusters, automated analyses were set to generate extracted ion chromatograms (EICs) for the top 50 spectral features as defined by both fold-change and p-value. These EICs were then manually inspected to identify the subset of automatically identified features that appear specific to the expressed BGC as defined by presence in each of three biological replicates of the production strain and absence from three biological replicates of a negative control strain **(FIG. 31).** EICs of all BGC specific features are illustrated in FIGS. 32-49.

Example 4: Construction of Yeast strains

[279] In the current example, yeast strains are based on the BY474 1/BY4742 background, which is in turn based on S288c (C. B. Brachmann, et al., 1998, cited supra). The strains were made in two stages: 1) creation of a core DHY set with restored sporulation and mitochondrial genome stability and 2) creation of JHY derivatives modified for other benefits, which may include protein production. All changes introduced in this study were confirmed by diagnostic PCR and sequencing.

[280] A sporulation-restored strain set was built by crossing BY4710 (C. B. Brachmann, et al, 1998, cited supra) to a haploid derivative of YAD373 (A. M. Deutschbauer and R. W. Davis, *Nat. Genet.* 37: 133-40, 2005), a BY-based diploid that contains three QTLs that restore sporulation: *MKT1 (30G),* i?ME7(INS-308A), and *TA03(\493Q).* A spore clone from the resulting diploid was repaired for *HAP1,* which encodes a zinc-finger transcription factor localized to mitochondria and the nucleus. *HAP1* is important for mitochondrial genome stability (see J. R. Matoon, E. Caravajal, and D. Gurthrie *Curr. Genet.* 17:179-83, 1990) and likely also important for sporulation. S288c and derivatives contain a Tyl insertion in the 3' end of HAPI that inactivates function. The transposon was excised using the Delitto Perfetto method (F. Storici and M. A.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1080 of 3077

Resnick, *Methods Enzymol.* 409:329-45, 2006) and confirmed repaired *HAP1* function based on transcription of a *CYClp-lacZ* reporter (M. Gaisne, et al., *Curr. Genet.* 36: 195-200, 1999). The sporulation-restored, *HAP1*-repaired strain and its auxotrophic and prototrophic derivatives were then used to create the DHY set of strains that were additionally restored for mitochondrial genome stability.

[281] The above sporulation-restored strains were used to repair the poor mitochondrial genome stability known to be a problem with S288c and BY derivatives. Mitochondrial genome stability is likely to improve growth and *ADH2p-like* gene expression under conditions of respiration, and for reducing the frequency of petite cells (slow-growing, respiration-defective cells that cannot grow on non-fermentable carbon sources). For a detailed description of the "mito-repair" method, see construction of JHY650 (J.D. Smith, 2017, cited supra). Briefly, the 50:50 genome editing method was used to introduce the wild-type alleles of three genes shown to be important for mitochondrial genome stability by QTL analysis[31]. The repaired QTLs are: *SAL1*$^{+}$ (repair of a frameshift), *CAT5(9\M)* and *MIP1* (66 IT). Crosses with prototrophic and auxotrophic strains completed the DHY core set of about a dozen sporulation and mitochondrial genome stability restored strains that can be further modified as needed. DHY213 (see **Table 3**) is one such strain: it contains the seven desired changes described above, is otherwise congenic with BY4741, and was used in this study to create derivatives for the HEx platform (see **Table 3).**

[282] Marker-free, seamless deletion of the complete *PRBl* and *PEP4* ORFs was performed using the 50:50 method (J. Horecka and R. W. Davis, 20 14, cited supra). Integration of a 1609 bp *ADH2p-npgA-ACSlt* expression cassette on the chromosome was performed using a similar method used to integrate DNA segments with the REDI method (J. D. Smith, et al, 2017, cited supra), except that *URA3,* not *FCY1,* was used as the counter-selectable marker. For an integration site, an 1166 bp cluster of three transposon LTRs located centromere-distal to YBR209W on chromosome II was replaced (deletion of chrll 643438 to 644603). Two DNA segments were simultaneously inserted via homologous recombination at the integration site that had been cut with *Seel* to create double strand breaks. One inserted segment was *ADH2p-npgA* (1448 bp) PCR amplified from a *BJ5464/npgA* expression strain *(npgA* from *A. nidulans)* (K. K. M. Lee, N. A. Da Silva, and J. T. Kealey, *Anal. Biochem.,* 394:75-80, 2009*)*. The *npgA* 3' end was repaired to wildtype using a reverse PCR primer that replaced the *npgA* intron included previously with the wildtype *npgA* 3' sequence. To preclude recombination of the expression cassette with the native *ADH2* locus, the 161 bp *ACS1* terminator was used as the second DNA segment (not *ADH2X)* and PCR amplified from BY474 1. The resulting strain (JHY692) was used in a similar fashion to replace only *npgA* with the *CPR* ORF (cytochrome P450 reductase, ATEG_05064 from *A. ierreus).* Finally, a strain with both *npgA* and *CPR* expression cassettes (JHY702) was created by mating JHY692 and JHY705.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1081 of 3077

Example 5: Determination of Chemical Structures

[283] For compound isolation, large-scale fermentation was carried out with the strains and clusters of Example 3. The yeast strains were first struck out onto the appropriate SDC dropout agar plates and incubated for 48 hrs at 30°C. A colony was then inoculated into 40 mL SDC dropout medium and incubated at 28°C for two days with shaking at 250 rpm. This seed culture was used to inoculate 4 L of YPD medium (1.5% Glucose) and cultured for 3 days at 28°C and 250 rpm. Supernatants were then clarified by centrifugation and extracted with equal volume of ethyl acetate. Cell pellets were extracted with 1 L of acetone. For compounds containing carboxylic acid groups, the pH value of the supernatant was adjusted to 3 by adding HC1 prior to extraction. The organic phases were combined and evaporated to dryness. The residue was purified by ISCO-CombiFlash® Rf 200 (Teledyne Isco, Inc) with a gradient of hexane and acetone. After analysis by LC-MS, the fractions containing the target compounds were combined and further purified by semi-preparative FIPLC using C18 reverse-phase column. The purity of each compound was confirmed by LC-MS, and the structure was solved by NMR **(Figure 49-59).**

[284] All NMR spectra including ¾ , ^{13}C, COSY, HSQC, HMBC and NOESY spectra were obtained on Bruker AV500 spectrometer with a 5 mm dual cryoprobe at the UCLA Molecular Instrumentation Center. The NMR solvents used for these experiments were purchased from Cambridge Isotope Laboratories, Inc.

Table 4: Summary of control and cryptic fungal BGCs examined in this study.

Cluster ID	Type	Native Locus				Species of origin	Division	Productive?
		Genbank ID	Start	End	Length			
IDT	Ctl					*Aspergillus tubingensis*	Ascomycota	Y
DHZ	Ctl					*Hypomyces subiculosus*	Ascomycota	Y

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1082 of 3077

PKS1	PKS	KV441552	530394	546723	16329	*Coniothyrium sporulosum*	Ascomycota	Y
PKS2	PKS	KV441551	60641	83767	23126	*Coniothyrium sporulosum*	Ascomycota	Y
PKS3	PKS	Deposition pending			9892	*Acremonium Sp. KY4917*	Ascomycota	N
PKS4	PKS	AM270992	578654	603367	24713	*Aspergillus niger*	Ascomycota	Y
PKS5	PKS	CP003009	8730831	8753560	22729	*Thielavia terrestris*	Ascomycota	N
PKS6	PKS	ABDF02000052	9685	28459	18774	*Trichoderma virens*	Ascomycota	Y
PKS7	PKS	JPJY01000093	2671	30026	27355	*Pseudogymnoascus pannorum*	Ascomycota	N
PKS8	PKS	JOWA01000110	1607857	1643205	35348	*Scedosporium apiospermum*	Ascomycota	Y
PKS9	PKS	KE384750	270098	292755	22657	*Metarhizium anisopliae*	Ascomycota	N
PKS10	PKS	KB445572	91380	115199	23819	*Cochliobolus heterostrophus*	Ascomycota	Y
PKS11	PKS	JPKB01001000	12749	37136	24387	*Pseudogymnoascus pannorum*	Ascomycota	N
PKS12	PKS	JPJU01000852	23274	38823	15549	*Pseudogymnoascus pannorum*	Ascomycota	N

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1083 of 3077

PKS13	PKS	JPJR01000396	848	19416	18568	*Pseudogymnoascus pannorum*	Ascomycota	Y
PKS14	PKS	KN847553	300827	326489	25662	*Verruconis gallopava*	Ascomycota	Y
PKS15	PKS	AWSO01000045	179446	203577	24131	*Moniliophthora roreri*	Basidiomycota	Y
PKS16	PKS	JH687542	431482	465332	33850	*Punctularia strigosozonata*	Basidiomycota	Y
PKS17	PKS	KN839868	143119	188013	44894	*Hydnomerulius pinastri*	Basidiomycota	Y
PKS18	PKS	DS989828	1389449	1407499	18050	*Arthroderma gypseum*	Ascomycota	Y
PKS19	PKS	KB908593	409171	443069	33898	*Setosphaeria turcica*	Ascomycota	N
PKS20	PKS	GL532685	4724	17845	13121	*Pyrenophora teres*	Ascomycota	Y
PKS21	PKS	AMGW01000002	1294251	1322017	27766	*Cladophialophora yegresit*	Ascomycota	N
PKS22	PKS	DF933843	225991	254307	28316	*Talaromyces cellulolyticus*	Ascomycota	Y
PKS23	PKS	KE720645	48853	80314	31461	*Endocarpon pusillum*	Ascomycota	Y
PKS24	PKS	DF933834	523551	558018	34467	*Talaromyces cellulolyticus*	Ascomycota	Y

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1084 of 3077

PKS25	PKS	AWSO01000633	5071	30440	25369	*Moniliophthora roreri*	Basidiomycota	N
PKS26	PKS	KN8 17529	130968	152240	21272	*Hypholoma sublateritium*	Basidiomycota	N
PKS27	PKS	KB445800	1362248	1403385	41137	*Ceriporiopsis subvermispora*	Basidiomycota	N
PKS28	PKS	AWSO01000632	8804	37857	29053	*Moniliophthora roreri*	Basidiomycota	Y
TCI	UTC	ABBF02000086	384012	368868	15144	*Trichoderma Virens*	Ascomycota	Y
TC2	UTC	ABDF02000083	37740	49247	11507	*Trichoderma Virens*	Ascomycota	N
TC3	UTC	FQ790293	97750	131296	33546	*Botryotonia cinerea*	Ascomycota	Y
TC4	UTC	JH7 17969	858521	869226	10705	*Formitiporia mediterranea*	Basidiomycota	Y
TC5	UTC	KI925459	2412992	2432365	19373	*Heterobasidion annosum*	Basidiomycota	Y
TC6	UTC	KB445798	581049	618823	37774	*Ge/atoporia subvermispora*	Basidiomycota	N
TC7	UTC	JH7 19450	83620	114270	30650	*Dichomitus squalens*	Basidiomycota	N
TC8	UTC	KL198014	1718438	1746540	28102	*Pleurotus ostreatus*	Basidiomycota	N
TC9	UTC	GL377319	205808	212935	7127	*Schizophyllum commune*	Basidiomycota	Y

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1085 of 3077

TC10	UTC	JH687394	99692	114281	14589	*Stereum hirsutum*	Basidiomycota	N
TC11	UTC	JH687396	33983	48346	14363	*Sternum hirsutum*	Basidiomycota	N
TC12	UTC	JH719415	280135	300470	20335	*Dichomitus squalens*	Basidiomycota	N
TC13	UTC	JH795868	73132	97338	24206	*Dacryopinax primogenitus*	Basidiomycota	N
							Total	43
							Productive	24

Table 5: Ion source parameters used in this study

Ion source parameter	Dual AJS	MMI
gas temperature	250 °C	350 °C
drying gas	12 L / min	7.5 L / min
nebulizer	10 psig	20 psig
sheath gas temp.	400 °C	-
sheath gas flow	12 L / min	-
vaporizer	-	250 °C
capillary voltage (Vcap)	3500 V	1500 V
nozzle voltage	1400 V	-
corona discharge	-	4 μA
fragmentor	100 V	120 V
skimmer	50 V	50 V
octopole 1 RF Vpp	750 V	750 V
charging voltage	-	1000 V

Table 6: Description of promoter sequences

SEQ ID NO.	Description
1	S. cerevisiae pADH2
2	S. cerevisiae pPCK1

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1086 of 3077

3	S. cerevisiae pMLS1
4	S. cerevisiae pICL1
5	S. cerevisiae pYLR307C-A
6	S. cerevisiae pYGR067C
7	S. cerevisiae pIDP2
8	S. cerevisiae pADY2
9	S. cerevisiae pGAC1
10	S. cerevisiae pECM13
11	S. cerevisiae pFAT3
12	S. cerevisiae pPUT1
13	S. cerevisiae pNQM1
14	S. cerevisiae pSFC1
15	S. cerevisiae pJEN1
16	S. cerevisiae pSIP18
17	S. cerevisiae pAT02
18	S. cerevisiae pYIG1
19	S. cerevisiae pFBP1
20	S. cerevisiae PH089
21	S. cerevisiae CAT2
22	S. cerevisiae CTA1
23	S. cerevisiae ICL2
24	S. cerevisiae ACS1
25	S. cerevisiae PDH1
26	S. cerevisiae REG2
27	S. cerevisiae CIT3
28	S. cerevisiae CFRC1
29	S. cerevisiae RGI2

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1087 of 3077

30	S. cerevisiae PUT4
31	S. cerevisiae NCA3
32	S. cerevisiae STL1
33	S. cerevisiae ALP1
34	S. cerevisiae NDE2
35	S. cerevisiae QNQ1
36	S. paradoxus pADH2
37	S. kudriavzevii pADH2
38	S. bayanus pADH2
39	S. mikitae pADH2
40	S. castellii pADH2
41	S. paradoxus pPCKl
42	S. kudriavzevii pPCKl
43	S. bayanus pPCKl
44	S. paradoxus pMLSl
45	S. kudriavzevii pMLSl
46	S. bayanus pMLSl
47	S. paradoxus pICLl
48	S. kudriavzevii pICLl
49	S. bayanus pICLl
50	S. cerevisiae pTDH3
51	S. cerevisiae pTEFl
52	S. cerevisiae pFBAl
53	S. cerevisiae pPDCl
54	S. cerevisiae pTPIl
55	S. cerevisiae tADH2

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1088 of 3077

56	S. cerevisiae tPGIl
57	S. cerevisiae tEN02
58	S. cerevisiae tTEFl
59	A. tubingensis GGPPS
60	A. tubingensis PT
61	A. tubingensis FMO
62	A. tubingensis Cyc
63	H. subiculosis hpm8
64	H. subiculosis hpm3
65	pCHIDT-2.1
66	pCHIDT-2c

Table 7: Description of gene sequences

Cluster ID NO.	**SEQ ID NO.**	**Description**
AFU3G	**67**	AFOC1
	68	AFOC9
	69	AFOC6
	70	AFOC5
	71	AFOC8
	72	AFOC4
	73	AFOC7
	74	AFOC5N
	75	AFOC2_PKS
	76	AFOC3
Afu1g17740	77	A1OC1_TF
	78	A1OC2_serine_hydrolase
	79	A1OC3_aldose_epimerase

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1089 of 3077

	80	AlOC4_P450
	81	Afulgl7740_2_PKS
Cal57	**82**	Cal57_1_SDR
	83	Cal 57_2_Acyl_CoA_oxidase
	84	Cal57_3_P450
	85	Cal57_4_FMO
	86	Cal57_5_PKS
	87	Cal 57_6_transferase
	88	Cal57_7_PfpI
	89	Cal57_8_hyp
	90	Cal 57_9_AB_hydrolase
	91	Cal57_10_hyp
Ca2032	**92**	Ca2032_1_3_GHMP_kinase
	93	Ca2032_1_PKS
	94	Ca2032_3_MT
	95	Ca2032_4_P450
	96	Ca2032_5_Ca_uniporter
	97	Ca2032_6_GNAT_acetyltransferase
KU14	**98**	KU14_SC3_4774_Cyclase
	99	KU14_SC3_4776_P450
	100	KU14_SC3_4773_polyprenyl_synthetase
	101	KU14_SC3_4771_AK_reductase
	102	KU14_SC3_4770_polyprenyl_synthetase
	103	KU14_SC3_4768_AK_reductase
	104	KU14_SC3_4772_DNA_repair
	105	KU14_SC3_47775_QacA_daig_transporter
	106	KU14_SC3_4769_Hyp
KU18	**107**	KU1 8_SH9_7287_Cyclase
	108	KU18_SH9_7288_P450

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1090 of 3077

	109	KU18_SH9_7289_P450
	110	KU18_SH9_7286_P450
	111	KU18_SH9_7285_DH
KU26	112	KU26_ETS81063.1_metallo_hydrolase
	113	KU26_ETS81064.1_hyp
	114	KU26_ETS81065.1_esterase
	115	KU26_ETS81066.1_PKS
	116	KU26_ETS81067.1_hyp
	117	KU26_ETS81068.1_SDR
	118	KU26_ETS81069.1_SDR
KU29	119	KU29_B023_6166_ADH
	120	KU29_B023_6167_aminotransferase_3
	121	KU29_B023_6168_fasciclin
	122	KU29_B023_6169_hyp
	123	KU29_B023_6170_esterase
	124	KU29_B023_6171_glycoside_hydrolase
	125	KU29_B023_6_172_hyp
	126	KU29_B023_6173_P450
	127	KU29_B023_6174_PKS
KU40	128	KU40_KFA56048.1_NmrA_like
	129	KU40_KFA56160.1_hyp
	130	KU40_KFA56046.1_phytanoly-CoA_dioxygenase
	131	KU40_KFA56190.1_NmrA_like
	132	KU40_KFA56035.1_PKS
	133	KU40_KFA56227.1_hyp
	134	KU40_KFA56040.1_hyp
	135	KU40_KFA56229.1_alkaline_serine_protease
KU41	136	KU41_KIL85236.1jihenylalanine_specificjiermease
	137	KU41_KIL85237.1_aldehyde_DH

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1091 of 3077

	138	KU41_KIL85238.1_DH
	139	KU41_KIL85239.1_hyp
	140	KU41_KIL85240.1_hyp
	141	KU41_KIL85241.1_hyp
	142	KU41_KIL85242.1_1-amino_cyclopropane-1-carboxylate_oxidase
	143	KU41_KIL85243.1_hyp
	144	KU41_KIL85244.1_PKS
	145	KU41_KIL85245.1_hyp
	146	KU41_KIL85246.1_aromatic_dioxygenase
	147	KU41_KIL85247.1_peptidase
	148	KU41_KIL85248.1_kinase
	149	KU41_KIL85249.1_metalloprotease
	150	KU41_KIL85250.1_BLA
	151	KU41_KIL85251.1_SDR
	152	KU41_KIL85252.1_AK_reductase
	153	KU41_KIL85253.1_metallopeptidase
KU44	154	KU44_SL06_4460_TPR_repeat
	155	KU44_SL06_4461_PKS
	156	KU44_SL06_4462_aminotransferase_V
	157	KU44_SL06_4463_MFS_transporter
PKS1	158	CS100GC1_CDS1_PKS
	159	CS100GC1_CDS2_serine_hydrolase
	160	CS100GC1_CDS3_p450
	161	CS100GC1_CDS4_short_chain_dehydrogenase
	162	CS100GC1_CDS5_FAD_dehydrogenase
PKS10	163	KU34_CH10_1770_epimerase_DH
	164	KU34_CH10_1802_P450
	165	KU34_CH10_1821_PKS
	166	KU34_CH10_1888_metallo_bla

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1092 of 3077

	167	KU34_CH10_1909_hyp
	168	KU34_CH10_1922_DUF 1772
	169	KU34_CH10_1937_NTF2_like
	170	KU34_CH10_1951_NmrA_like
	171	KU34_CH10_1975_SDR
	172	KU34_CH10_2002_ABC_transporter
PKS11	173	KU35_KFY73936.1_SDR
	174	KU35_KFY73937.1_DUF_3425
	175	KU35_KFY73938.1_Zn_finger
	176	KU35_KFY73939.1_0MT
	177	KU35_KFY73940.1_metallo_lactamase
	178	KU35_KFY73941.1_PKS
	179	KU35_KFY73942.1_acetyl_transferase
	180	KU35_KFY73943.1_NmrA_like
	181	KU35_KFY73944.1_FAD_linked_oxygenase
PKS12	182	KU36_KFY14209.1_hyp
	183	KU36_KFY14210.1_OMT
	184	KU36_KFY14211.1_Metallo_BLA
	185	KU36_KFY14212.1_PKS
	186	KU36_KFY14213.1_FAD_linked_oxidase
PKS13	187	KU37_KFY01907.1_DH
	188	KU37_KFY01908.1_ABC_tranporter
	189	KU37_KFY01909.1_PKS
	190	KU37_KFY01910.1_MFS
	191	KU37_KFY01911.1_PKc_like
PKS14	192	KU38_KIW01747.1_metallo_hydrolase
	193	KU38_KIW01748.1_halogenase
	194	KU38_KIW01749.1_P450
	195	KU38_KIW01750.1_cupin_like

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1093 of 3077

	196	KU3 8_KIW0 1751. 1_0MT
	197	KU3 8_KIW0 1752. 1_MHR_TF
	198	KU3 8_KIW0 1753. 1_monoxygenase
	199	KU38_KIW01754.1 _PKS
PKS15	200	KU39_ESK96608.1 _monocarboxylate _permease
	201	KU39_ESK96609. I_NAD _epimerase_DH
	202	KU39_ESK96610. 1_hyp
	203	KU39_ESK9661 1.1_carbonyl_reductase
	204	KU39_ESK96612.1_phenol_2_monoxygenase
	205	KU3 9_ESK966 13.1_PKS
	206	KU39_ESK96614.1 _SDR
PKS16	207	NW_006767437_-_cystathionine_beta-synthase_CDS
	208	NW_006767437_-_hypothetical_protein_l_CDS
	209	NW_006767437_-_hypothetical_protein_2_CDS
	210	NW_006767437_-_hypothetical_protein_3_CDS
	211	NW_006767437_-_Pkinase-domain-containing_protein _CDS
PKS17	212	KU43_KIJ60838.1
	213	KU43_KIJ60843 .1_P450
	214	KU43_KIJ60845.1
	215	KU43_KIJ60839.1 _isoprenylcysteine _carboxyl_methyltransferase
	216	KU43_KIJ60847.1 _isoprenylcysteine _carboxyl_methyltransferase
	217	KU43_KIJ60848. 1_ABC_transporter
	218	KU43_KIJ60844.1
	219	KU43_KIJ60846. 1_P450
	220	KU43_KIJ60842.1
	221	KU43_KIJ60840. 1_ABC_transporter
	222	KU43_KIJ60837. 1_P450
	223	KU43_KIJ60841.1 _P450
	224	KU43_KIJ60886. 1_PKS

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1094 of 3077

PKS18	**225**	SU62_EFR04826.1_hypothetical_protein
	226	SU62_EFR04827.1_fatty-acid-CoA_ligase
	227	SU62_EFR04828.1_pks
	228	SU62_EFR04829. 1_esterase
PKS19	**229**	SU64_EOA86426. 1_YfhR_like
	230	SU64_EOA86427. 1_ER
	231	SU64_EOA86425. 1_FAD-hydroxylase
	232	SU64_EOA86421 .1_Drug _resistance _transporter
	233	SU64_EOA86423 .1_0MT
	234	SU64_EOA86422. 1_pks
	235	SU64_EOA86424. 1_esterase
PKS2	**236**	CS163GCl_CDSl_serine_hydrolase
	237	CS163GCl_CDS2_p450
	238	CS163GC1 _CDS3 _PKS
	239	CS163GC1_CDS4_short_chain_dehydrogenase
	240	CS163GCl_CDS_FAD_dehydrogenase
PKS20	**241**	SU65 _EFQ95559. 1_BLA
	242	SU65 _EFQ95560. 1_PKS
	243	SU65_EFQ95561.1_hyp
	244	SU65 _EFQ95562.1 _KR
PKS21	**245**	SU67_EXJ6 1964.1 _drug _resistance _transporter
	246	SU67_EXJ6 1965.1 _scytalone _dehydratase
	247	SU67_EXJ6 1966.1 _versicolorin _reductase
	248	SU67_EXJ6 1967.1_AMP_ligase
	249	SU67_EXJ6 1968.1_hypothetical_protein
	250	SU67_EXJ6 1969.1 _FAD _monooxygenase
	251	SU67_EXJ61970.1_pks
	252	SU67_EXJ6 197 1.1 _metallo _BLA
	253	SU67_EXJ61972.1_hyp

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1095 of 3077

	254	SU67_EXJ6 1973.1_SDR
	255	SU67_EXJ6 1974.1_AifR_reg
PKS22	**256**	SU68_GAM43 180.1_beta-lactamase_family_protein
	257	SU68_GAM43 181.1_NAD-dependent_epimerase_dehydratase
	258	SU68_GAM43 183.1_oxidoreductase
	259	SU68_GAM43 185.1_benzoate_4-monooxygenase_cytochrome_P450
	260	SU68_GAM43 187.1_scytalone_dehydratase
	261	SU68_GAM43 176.1_riboflavin_biosynthesis_protein
	262	SU68_GAM43 177.1_sugar_transport_protein
	263	SU68_GAM43 178.1_short-chain_dehydrogenase
	264	SU68_GAM43 179.1_pks
	265	SU68_GAM43 182. ljialogenase
	266	SU68_GAM43 184.1_NAD-dependent_epimerase_dehydratase
	267	SU68_GAM43 186. 1_SDR
PKS23	**268**	SU70_ERF77218. 1_SDR
	269	SU70_ERF772 19.1_Fungal_TF
	270	SU70_ERF77220. 1_p450
	271	SU70_ERF77221.1_pks
	272	SU70_ERF77222. 1_DABB_superfamily
	273	SU70_ERF77223 .1_FAD_monooxygenase
	274	SU70_ERF77224. 1_alcohol_DH
	275	SU70_ERF77225. 1_OMT
	276	SU70_ERF77226. 1_alcohol_DH
PKS24	**277**	SU7 1_GAM40295 .1_sulfhydrolase
	278	SU71_GAM40298. 1_AMP_ligase
	279	SU71_GAM40299. 1_MT
	280	SU71_GAM40301 .1_carnosine_synthase
	281	SU71_GAM40303.1_carnitine_acetyl-CoA_transferase
	282	SU71_GAM40296. ^aminotransferase

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1096 of 3077

	283	SU71_GAM40297. 1_ammonia_lyase
	284	SU71_GAM40300. 1_pks
	285	SU71_GAM40302. 1_ABC_tranporter
PKS25	**286**	SU72_ESK88623.1_pks
	287	SU72_ESK8 8624.1 _carotenoi d_eleavage_dioxygenase_ 1
	288	SU72_ESK88625.1_long-chain-fatty-acid-ligase
	289	SU72_ESK88626.1_amino_acid_permease
PKS26	**290**	SU73_KN817529.1_pks
	291	SU73_KN817529.1_hyp
	292	SU73_KN817529.1_AMP_ligase
	293	SU73_KN8 17529. 1_8_amino_7_oxanoate_synthase
PKS27	**294**	SU74_EMD3 5673 .1_NAD _DH
	295	SU74_EMD35676. 1_sulfhydrylase
	296	SU74_EMD35664. 1_hyp
	297	SU74_EMD35669. 1_halogenase
	298	SU74_EMD3 5663 .1_AB_hydrolase
	299	SU74_EMD35666. 1 _alcohol _DH
	300	SU74_EMD3 5667. 1 _Drug _resistance _transporter
	301	SU74_EMD35668. 1_hyp
	302	SU74_EMD35670. 1_P450
	303	SU74_EMD3 5665 .1_hyp
	304	SU74_EMD3 567 1.1_pks
	305	SU74_EMD35672. 1_hyp
	306	SU74_EMD35674. 1_hyp
	307	SU74_EMD3 5675 .1_hyp
PKS28	**308**	SU75_ESK88629.1_pks
	309	SU75_ESK88630.1_dmg_resistance_subfamily
	310	SU75 _ESK8863 1.1_hypothetical_protein
	311	SU75_ESK8 863 2. 1_dead-box_protein_ab strakt

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1097 of 3077

	312	SU75_ESK88633.1_hyp
	313	SU75_ESK88634.1_nadh-ubiquinone _oxidoreductase
PKS3	314	AK_C2470 1GC76_CDS 1_class_11_aminotransferase
	315	AK_C24701GC76_CDS2_p450
	316	AK_C2470 1GC76_CDS3_PKS
	317	AK_C2470 1GC76_CDS4_ferric_chelate_reductase
	318	AK_C2470 1GC76_CDS5_DUF4243
PKS4	319	KU27_AN22_1 464_esterase
	320	KU27_AN22_1465_PKS
	321	KU27_AN22_1 466_hyp
	322	KU27_AN22_1 467_A_TD
	323	KU27_AN22_1 468_hyp
	324	KU27_AN22_1 469_amino_oxidase
PKS5	325	KU28_TT08_272 1_AK_reductase
	326	KU28_TT08_2722_ABC_transporter
	327	KU28_TT08_2723_esterase
	328	KU28_TT08_2724_PKS
	329	KU28_TT08_2725_sugar_transport
PKS6	330	KU30_TV43_5580_esterase
	331	KU30_TV43_5581_P450
	332	KU30_TV43_5582_PKS
	333	KU30_TV43_5583_OMT
	334	KU30_TV43_5584_P450
PKS7	335	KU3 1_KFY69032. 1_P450
	336	KU3 1_KFY69033 .1_hyp
	337	KU3 1_KFY69034. 1_esterase
	338	KU31_KFY69035.1 _P450
	339	KU31_KFY69036.1 _PKS
	340	KU3 1_KFY69037. 1_glycoside _hydrolase

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1098 of 3077

	341	KU3 1_KFY69038. 1_thymine_dioxygenase
PKS8	**342**	KU32_KEZ41287. ^nucleotidyltransferase
	343	KU32_KEZ41288. 1_0MT
	344	KU32_KEZ41289. 1_AB_hydrolase
	345	KU32_KEZ4 1290.1_mito_phos_carrier
	346	KU32_KEZ41291 .1_hyp
	347	KU32_KEZ4 1292.1_crotonyl_CoA_reductase
	348	KU32_KEZ41293. 1_PKS
	349	KU32_KEZ41294.1 _Drug_resistance_transporter
	350	KU32_KEZ4 1295.1_NADB_monoxy genase
PKS9	**351**	KU33_KJK75348.1_alkaline_serine_hydrolase
	352	KU33_KJK75349. 1_LysM_containing_protein
	353	KU33_KJK75350.1 _PKS
	354	KU33_KJK7535 1.1_drug_resistance_transporter
	355	KU33_KJK75352.1_phytanoyl_CoA_dioxygenase
	356	KU33_KJK75353.1 _SDR
	357	KU33_KJK75354.1 _amino-7_oxonolate_synthase
SU61	**358**	SU61_ENH82084. 1_choline_oxidase
	359	SU6 1_ENH82085 .1_fungal_specific_transcription _factor
	360	SU6 1_ENH82086 .1_short-chain_dehydrogenase
	361	SU6 1_ENH82087. 1_hypothetical_protein
	362	SU6 1_ENH 82088. 1_hypothetical_protein
	363	SU61_ENH82089.1_pks
	364	SU61_ENH82090. 1_esterase
	365	SU6 1_ENH8209 1.1_ABC_multidrug_transporter _mdrl
	366	SU6 1_ENH82092. 1_cytochrome_b5_type_b
	367	SU6 1_ENH82093 .1_sulfite_reductase_subunit_alpha
SU63	**368**	SU63_KJZ74253 .1_hyp_signalling_protein
	369	SU63_KJZ74254.1_mtRNA_formyl_transferase

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1099 of 3077

	370	SU63_KJZ74255. 1_esterase
	371	SU63_KJZ74256. 1_PKS
	372	SU63_KJZ74257.1_choline_DH_halogenase
SU66	**373**	SU66_EMF 173 84.1 _ras-domain-containing _protein
	374	SU66_EMF 17385. 1_phosphoinositide _phosphatase
	375	SU66_EMF 173 87. 1_hypothetical_protein
	376	SU66_EMF17389.1 _versicolorin _reductase
	377	SU66_EMF 17390. 1_scytalone_dehydratase
	378	SU66_EMF17386.1_pks
	379	SU66_EMF 173 88. 1_Metallo-hydrolase _oxidoreductase
	380	SU66_EMF 17391. 1_StcQ-like_protein
SU69	**381**	SU69_KFY04761 .1_aldehyde _oxidase
	382	SU69_KFY04762. 1_SDR
	383	SU69_KFY04763 .1_hyp
	384	SU69_KFY04764. 1_reg_protein
	385	SU69_KFY04765. 1_OMT
	386	SU69_KFY04766. 1_metallo_BLA
	387	SU69_KFY04767. 1_pks
	388	SU69_KFY04768.1_FAD-linked_oxidase
TCI	**389**	Tv86_130_CDS
	390	Tv86_132_CDS
	391	Tv86_133_CDS
	392	Tv86_134_CDS
	393	Tv86_135_ORF
	394	Tv86_136_CDS
	395	Tv86_137_CDS
TC10	**396**	KU19_SH16_10821_Cyclase
	397	KU19_SH16_10820_P450
	398	KU19_SH16_10822_P450

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1100 of 3077

	399	KU19_SH16_10823_AK_reductase
	400	KU19_SH16_10819_QacA_MFS
TC11	401	KU_20_SH18_1 1663_Cyclase
	402	KU_20_SH18_1 1664_P450
	403	KU_20_SH18_11665_P450
	404	KU_20_SH18_11662_P450
	405	KU_20_SH18_11660_NMT
	406	KU_20_SH18_11661_MFS
TC12	407	KU21_DS19_71 12_Cyclase
	408	KU21_DS19_71 13_P450
	409	KU21_DS19_7108_P450
	410	KU21_DS19_7109_MT
	411	KU21_DS19_7111_GMC_oxido
	412	KU21_DS19_7114_AK_reductase
	413	KU21_DS19_71 10_Hyp
TC13	414	KU22_DDJM14_6568_Cyclase
	415	KU22_DDJM14_6570_P450
	416	KU22_DDJM14_6567_0MT
	417	KU22_DDJM14_6565_MT
	418	KU22_DDJM14_6564_MT
	419	KU22_DDJM14_6562_P450
	420	KU22_DDJM14_6563_MT
	421	KU22_DDJM14_6561_GST
	422	KU22_DDJM14_6569_V-type_ATPase
	423	KU22_DDJM14_6566_TF
TC2	424	Tv83_13_CDS
	425	Tv83_14_CDS
	426	Tv83_15_CDS
	427	Tv83_16_CDS

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1101 of 3077

TC3	428	BFT4_1_FAD_binding_domain_protein
	429	BFT4_2_Aldo_keto_reductase_oxidoreductase
	430	BFT4_3_cytochrome_P450
	431	BFT4_4_UbiA_cyclase
	432	BFT4_6_hyp_protein_4
	433	BFT4_7_hyp_protein_5
	434	BFT4_8_hyp_protein_6
	435	BFT4_10_FAD_FMN_isoamyl_alcohol_oxidase
	436	BFT4_11_hyp_protein_8
	437	BFT4_14_hyp_protein_1_1
	438	BFT4_15_glutathione_S_transferase
	439	BFT4_16_D_isomer_specific_2_hydroxyacid_dehydrogenase
TC4	440	KU11_FM3_3034_Cyclase
	441	KU1 1_FM3_3033_P450
	442	KU1 1_FM3_3032_P450
	443	KU1 1_FM3_3031_FAD
	444	KU11_FM3_3037_Hydrox
	445	KU1 1_FM3_3027_AMP_SDR
	446	KU1 1_FM3_3030_PHOS
	447	KUl 1_FM3_3035_Hyp
TC5	448	KU12_HI6_11661_Cyclase
	449	KU12_HI6_11655_P450
	450	KU12_HI6_11638_P450
	451	KU12_HI6_11667_AMP_ligase
	452	KU12_HI6_11646_monocarbox_MFS
	453	KU12_HI6_11632_MFS
TC6	454	KU13_CeS8_5906_Cyclase
	455	KU13_CeS8_5905_P450
	456	KU13_CeS8_591 1_P450

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1102 of 3077

	457	KU13_CeS8_5904_choline_DH
	458	KU13_CeS8_5910_halogenase
	459	KU13_CeS8_5907_AMP_ligase
	460	KU13_CeS8_5909_superoxide_dismutase
	461	KU13_CeS8_5908_MFS
	462	KU13_CeS8_5912_MFS
TC7	**463**	KU15_DS54_10337_Cyclase
	464	KU15_DS54_10336_P450
	465	KU15_DS54_10340_P450
	466	KU15_DS54_10341_P450
	467	KU15_DS54_10338_polyprenyl_synthetase
	468	KU15_DS54_10333_GMC_oxido
	469	KU15_DS54_10334_GMC_oxido
	470	KU15_DS54_10339_Hyp
	471	KU15_DS54_10335_Hyp_transmembrane
TC8	**472**	KU16_P01_1_11845_Cyclase
	473	KU16_P01_1_11844_FAD_oxidase
	474	KU16_P01_1_11843_P450
	475	KU16_P01_1_11841_2_P450
	476	KU16_P01_1_11840_FAD_oxidase
	477	KU16_P01_1_11847_FAD_oxidase
	478	KU16_P01_1_11846_SDR
	479	KU16_P01_1_11848_AMP_ligase
	480	KU16_P01_1_11839_AMP_ligase
TC9	**481**	KU17_SC18_16687_Cyclase
	482	KU17_SC18_16686_P450
	483	KU17_SC18_16688_SDR

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1103 of 3077

[285] While preferred embodiments of the present disclosure have been shown and described herein, those skilled in the art will understand that such embodiments are provided by way of example only. Numerous variations, changes, and substitutions will now occur to those skilled in the art without departing from the disclosure.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1104 of 3077

CLAIMS

WHAT IS CLAIMED IS:

1. A method for screening a plurality of compounds, the method comprising:

identifying a gene cluster that includes or is within 20 kilobases of a region that encodes for a protein that is identical with or homologous to a first target protein, wherein the gene cluster comprises a region that encodes for a protein selected from the group consisting of (1) polyketide synthases and (2) non-ribosomal peptide synthetases, (3) terpene synthetases, (4) UbiA-type terpene cyclases, and (5) dimethylallyl transferases;

introducing a plurality of genes from the gene cluster into a vector;

introducing the vector into a host cell;

expressing the proteins encoded by the plurality of genes in the host cell;

determining whether a compound that is formed or modified by the expressed proteins modulates the first target protein.

2. The method of claim 1, wherein the host cell is a yeast cell.

3. The method of clam 2, wherein the yeast cell is a yeast cell that has been modified to increased sporulation frequency and increased mitochondrial stability.

4. The method of any one of claims 1-3, wherein each gene of the plurality of genes is under the control of a different promoter.

5. The method of claim 4, wherein the promoters are designed to increase expression when the host cell is in the presence a nonfermentable carbon source.

6. The method of any one of claims 1-5, wherein the plurality of genes are introduced into the vector via homologous recombination.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1105 of 3077

7. The method of claim 6, wherein introducing the plurality of genes into the vector via homologous recombination comprises combining a first plurality of nucleotides with a second plurality of nucleotides, wherein:

each polynucleotide of the first plurality of DNA polynucleotides encodes for a promoter and terminator, wherein each promoter and terminator is distinct from the promoter and terminator of other polynucleotides of the first plurality of nucleotides; and

each polynucleotide of the second plurality of nucleotides includes a coding sequence, a first flanking region on the 5' side of the polynucleotide and a second flanking region on the 3' side of the polynucleotide; and

introducing the polynucleotides into a host cell that includes machinery for homologous recombination, wherein the host cell assembles the expression vector via homologous recombination that occurs in the flanking regions of the second plurality of polynucleotides;

wherein the expression vector is configured to facilitate simultaneously production of a plurality of proteins encoded by the second plurality of nucleotides.

8. The method of claim 7, wherein the first flanking region and the second flanking region are each between 15 and 75 base pairs in length.

9. The method of claim 8, wherein the first flanking region and the second flanking region are each between 40 and 60 base pairs in length.

10. A method for identifying a gene cluster capable of producing a small molecule for modulating a first target protein, the method comprising:

selecting, from a database comprising a list of biosynthetic gene clusters, one or more gene clusters that include or are positioned proximal to a region that encodes a protein that is identical with or homologous to the first target protein.

11. The method of claim 10, wherein the one or more gene clusters are selected from the group consisting of (1) clusters that comprise one or more polyketide synthases and (2) clusters that comprise one or more non-ribosomal peptide synthetases, (3) clusters that comprise one or more

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1106 of 3077

terpene synthases, (4) clusters that comprises one or more UbiA-type terpene cyclases, and (5) clusters that comprise one or more dimethylallyl transferases.

12. The method of any one of claims 10-1 1, wherein the protein that is encoded by the region that is included in or positioned proximal to the biosynthetic gene cluster is identical to or has greater than 30% homology to the first target protein.

13. The method of any one of claims 10-12, wherein the region that encodes the protein that is identical with or homologous to the first target protein is within 20,000 base pairs of a region of a portion of the gene cluster that encodes a polyketide synthase, a non-ribosomal peptide synthetase, a terpene synthetase, a UbiA-type terpene cyclase, or a dimethylallyl transferase.

14. The method of any one of claims 10-13, wherein the first target protein is BRSK1.

15. The method of any one of claims 10-14, wherein selecting the one or more gene clusters comprises operating a computer, wherein operation of the computer comprises running an algorithm that takes into account both an input sequence for the first target protein and sequence information from a database that includes sequence information from a plurality of species such that the computer returns information corresponding to one or more gene clusters.

16. The method of claim 15, wherein the algorithm takes into account the phylogenetic relationship between gene clusters in the database.

17. The method of any one of claims 10-16, wherein the one or more gene clusters include a coding sequence for a protein that is an extracellular protein, a membrane tethered protein, a protein involved in a transport or secretion pathway, a protein homologous to a protein involved in a transport or secretion pathway, a protein with a peptide targeting signal, a protein with a terminal sequence with homology to a targeting signal, an enzyme that degrades small molecules, or a protein with homology to an enzyme that degrades small molecules.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1107 of 3077

18. A method for producing a compound that modulates the first target protein, the method comprising:

identifying a gene cluster via the method of any one of claims 10-17;

expressing the gene cluster or a plurality of genes from the gene cluster in a host cell; and

isolating a compound produced by the gene cluster.

19. The method of claim 18, further comprising screening the isolated compound for modulation of an activity of the first target protein.

20. The method of claim 18 or claim 19, wherein the cluster-encoded protein that is homologous to the first target protein is resistant to modulation by the isolated compound when compared to modulation of the first target protein.

21. The method of claim 20, wherein the compound is not toxic to the species from which the cluster originates due to one or more of (1) sequence differences between target protein and the cluster-encoded protein, (2) spatial separation of the compound from the cluster-encoded protein and (3) high expression levels for the cluster-encoded protein.

22. A method for making a DNA vector, the method comprising:

identifying a gene cluster comprising a plurality of genes that are capable of producing a small molecule for modulating a first target protein;

introducing two or more genes of the plurality of genes into a vector, wherein the vector is configured to facilitate expression of the two or more genes in a host organism;

wherein (1) the gene cluster encodes one or more proteins selected from the group consisting of a polyketide synthase, a non-ribosomal peptide synthetase, a terpene synthetase, a UbiA-type terpene cyclase, and a dimethylallyl transferase and (2) the gene cluster includes or is positioned proximal to a region that encodes a protein that is identical with or homologous to the first target protein.

23. The method of claim 22, wherein the DNA vector is a circular plasmid.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1108 of 3077

24. The method of any one of claims 22-23, wherein the DNA vector comprises a plurality of promoters, wherein each promoter of the plurality of promoters is configured to, when the vector is introduced into a *Saccharomyces cerevisiae* cell, promote a lower level of heterologous expression when the cell exhibits predominantly anaerobic energy metabolism than when the cell exhibits aerobic energy metabolism.

25. The method of claim 24, wherein each promoter of the plurality of promoters differs in sequence from one another.

26. The method of any one of claims 24-25, wherein each promoter of the plurality of promoters has a sequence selected from the group consisting of: SEQ ID NOs: 1-66.

27. The method of claim 26, wherein each promoter of the plurality of promoters has a sequence selected from the group consisting of: SEQ ID NOs: 20-35, and SEQ ID NOs: 41-50.

28. The method of any one of claims 24-27, wherein, when the *Saccharomyces cerevisiae* cell is exhibiting anaerobic energy metabolism, the cell is catabolizing a fermentable carbon source selected from glucose or dextrose; and when the *Saccharomyces cerevisiae* cell is exhibiting aerobic energy metabolism, the cell is catabolizing a non-fermentable carbon source selected from ethanol or glycerol.

29. A method for the heterologous expression of a plurality of genes in a yeast strain, the method comprising:

obtaining a yeast strain that includes a vector for expressing a plurality of genes from a single gene cluster of a non-yeast organism; and

inducing expression of the plurality of genes;

wherein the gene cluster in the non-yeast organism includes or is positioned proximal to a region that encodes a protein that is at least 30% homologous to a target protein.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1109 of 3077

30. The method of claim 29, further comprising introducing the plurality of genes from the single gene cluster into the vector.

31. The method of any one of claims 29-30, further comprising introducing the vector into the yeast strain.

32. The method of any one of claims 29-31, wherein expression of the plurality of genes results in the formation of small molecule, wherein the small molecule modulates the activity of the target protein.

33. The method of any one of claims 29-32, wherein the gene cluster is a gene cluster of a non-yeast fungus.

34. The method of any one of claims 29-33, wherein the yeast strain is from *Saccharomyces cerevisiae.*

35. The method of any one of claims 29-34, wherein the target protein is a human protein.

36. A system for identifying a gene cluster for introduction of a plurality of genes from the gene cluster into a host organism, the system comprising:

a processor;

a non-transitory computer-readable medium comprising instructions that, when executed by the processor, cause the processor to perform operations, the operations comprising:

loading the identity or sequence of a first target protein into memory;

loading the identity or sequence of a plurality of biosynthetic gene clusters into memory;

identifying, from the plurality of biosynthetic gene clusters, one or more gene clusters that encode or are positioned proximal to a region that encodes a protein that is identical with or homologous to the first target protein; and

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1110 of 3077

scoring the one or more gene clusters based on the likelihood of each gene cluster being capable of use to produce a small molecule that modulates the first target protein.

37. The system of claim 36, wherein scoring the one or more gene clusters comprises comparing the sequence of the first target protein (or a DNA sequence encoding the first target protein) to a sequence of a protein encoded in or proximal to the gene cluster (or to a DNA sequence encoding the protein that is in or proximal to the gene cluster).

38. A system for identifying one or more biosynthetic gene clusters for introduction into a host organism to produce one or more compounds that modulate a specific target protein, the system comprising:

a processor;

a memory containing a gene cluster identification application;

wherein the gene cluster identification application directs the processor to:

load data describing at least one target protein into the memory;

load data describing a plurality of biosynthetic gene clusters into the memory;

score each of the plurality of biosynthetic gene clusters based upon:

performing a homolog search for each biosynthetic gene cluster to determine a presence of at least one homolog of a target protein within or adjacent the biosynthetic gene cluster;

confidence of homology of the at least one target protein to at least one gene in a biosynthetic gene cluster;

a fraction of a homologous gene that meets an identity threshold;

a total number of genes homologous to the at least one target protein present in the entire genome of an organism;

homology of the at least one homolog of at least one target protein within or adjacent the biosynthetic gene cluster to genes in the target protein's genome;

phylogenetic relationship of the at least one target protein to a gene in a cluster;

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1111 of 3077

expected number of homologs of the at least one target protein in or adjacent to a biosynthetic cluster; and

a likelihood that at least one target protein is essential for cellular process in the natural environment; and

output a report identifying one or more biosynthetic gene clusters that are most likely to produce a compound that modulates the at least one target protein.

39. A method for selecting a biosynthetic gene cluster that produces a secondary metabolite, the method comprising:

obtaining a list of gene clusters;

performing a phylogenetic analysis of the genes within the clusters compared to known genes from known biosynthetic gene clusters; and

selecting the biosynthetic gene cluster based on its phylogenetic relationship with the known genes.

40. The method of claim 39, wherein the biosynthetic gene cluster with the most distant phylogenetic relationship from the known genes is selected.

41. A method for identifying a gene cluster that produces a compound that binds a protein of interest, the method comprising:

obtaining sequence information for a plurality of contiguous sequences, wherein each contiguous sequence includes a biosynthetic gene cluster and flanking genomic sequences;

analyzing the contiguous sequences for the presence of a gene that encodes a protein with homology to the protein of interest, and

selecting a biosynthetic gene cluster which includes, or is proximal to, a gene that encodes a protein that is homologous to the protein of interest.

42. The method of claim 41, wherein the contiguous nucleotide sequence is less than 40,000 base pairs in length.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1112 of 3077

43. A modified yeast cell having a BY background, wherein relative to unmodified BY4741 and BY4742, the modified yeast cell has both (1) increased sporulation frequency and (2) increased mitochondrial stability.

44. The modified yeast cell of claim 43, wherein the modified yeast cell grows faster on non-fermentable carbon sources than unmodified BY4741 and BY4742.

45. The modified yeast cell of any one of claims 43-44, wherein the yeast cell comprises one or more of the following genotypes: *MKT1(30G), RME1(INS-308A),* and *TA03(1493Q).*

46. The modified yeast cell of any one of claims 43-45, wherein the yeast cell comprises one or more of the following genotypes: *CAT5(91M), M P1(66IT), SALJ+,* and *HAPJ+.*

47. A method of forming an expression vector, the method comprising:

combining a first plurality of DNA polynucleotides with a second plurality of polynucleotides, wherein:

each polynucleotide of the first plurality of DNA polynucleotides encodes for a promoter and terminator, wherein each promoter and terminator is distinct from the promoter and terminator of other polynucleotides of the first plurality of nucleotides; and

each polynucleotide of the second plurality of nucleotides includes a coding sequence, a first flanking region on the 5' side of the polynucleotide and a second flanking region on the 3' side of the polynucleotide, wherein each flanking region is between 15 and 75 base pairs in length; and

introducing the polynucleotides into a host cell that includes machinery for homologous recombination, wherein the host cell assembles the expression vector via homologous recombination that occurs in the flanking regions of the second plurality of polynucleotides;

wherein the expression vector is configured to facilitate simultaneously production of a plurality of proteins encoded by the second plurality of nucleotides.

48. The method of claim 47, wherein the host cell is a yeast cell.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1113 of 3077

49. The method of any one of claims 47-48, wherein each flanking region is between 40 and 60 base pairs in length.

50. The method of any one of claims 47-49, wherein at least one polynucleotide of the first plurality of nucleotides encodes a selection marker.

51. A system for generating a synthetic gene cluster via homologous recombination, the system comprising 1 though N unique promoter sequences, 1 through N unique terminator sequences, and 1 through N unique coding sequences, wherein:

coding sequence 1 is attached to an additional 30-70 base pair sequence on each end such that a first end portion is identical to the last 30-70 base pairs of promoter 1 and a second end portion is identical to the first 30-70 base pairs of terminator 1;

coding sequence 2 is attached to an additional 30-70 base pair sequence on each end such that a first end portion is identical to the last 30-70 base pairs of promoter 2 and a second end portion is identical to the first 30-70 base pairs of terminator 2;

and coding sequence N is attached to an additional 30-70 base pair sequence on each end such that a first end portion is identical to the last 30-70 base pairs of promoter N and a second end portion is identical to the first 30-70 base pairs of terminator N.

52. The system of claim 51, wherein terminator 1 and promoter 2 are portions of the same double-stranded oligonucleotide.

53. A method for assembling a synthetic gene cluster, the method comprising:

a. obtaining 1 through N unique promoters, 1 through N unique terminators, and 1 through N unique coding sequences, wherein:

i. coding sequence 1 is attached to an additional 30-70 base pair sequence on each end such that a first end portion is identical to the last 30-70 base pairs of promoter 1 and a second end portion is identical to the first 30-70 base pairs of terminator 1;

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1114 of 3077

ii. coding sequence 2 is attached to an additional 30-70 base pair sequence on each end such that a first end portion is identical to the last 30-70 base pairs of promoter 2 and a second end portion is identical to the first 30-70 base pairs of terminator 2; and

iii. and coding sequence N is attached to an additional 30-70 base pair sequence on each end such that a first end portion is identical to the last 30-70 base pairs of promoter N and a second end portion is identical to the first 30-70 base pairs of terminator N;

b. transforming the 1 through N promoters, terminators and coding sequences into a yeast cell;

c. isolating a plasmid containing the 1 through N promoters, terminators and coding sequences from the yeast cell.

54. The method of claim 53, further comprising a coding sequence for a selection marker.

55. The method of claim 54, wherein the selection marker is an auxotrophic marker.

56. The method of any one of claims 53-55, wherein the yeast cell has a deficiency in a DNA ligase gene.

57. A yeast strain which allows for both (1) homologous DNA assembly and (2) production of heterologous genes in the same strain.

58. The yeast strain of claim 57, wherein the yeast strain is a DHY strain.

59. The yeast strain of and one of claims 57-58, wherein the strain allows DNA assembly via homologous recombination with an efficiency of at least 80% as compared to DNA assembly in BY.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1115 of 3077

60. The yeast strain of any one of claims 57-59, wherein production of heterologous compounds in the strain is accomplished with an efficiency of at least 80% as compared to heterologous compound production in BJ5464.

61. The yeast strain of any one of claims 57-60, wherein the strain allows production of heterologous proteins with an efficiency of at least 80% as compared to heterologous protein production in BJ5464.

62. A method for isolating a plasmid from a yeast cell, the method comprising:

a. isolating total DNA from a yeast cell that includes a plasmid;
b. incubating the DNA with an exonuclease such that the exonuclease degrades substantially all of the linear DNA in the isolated total DNA from the yeast cell;
c. optionally inactivating the exonuclease; and
d. recovering the plasmid DNA.

63. The method of claim 62, wherein the isolated plasmid DNA is of sufficient purity for use in a sequencing reaction.

64. The method of any one of claims 62-63, wherein the plasmid DNA is further prepared for a sequencing reaction.

65. A pharmaceutical composition comprising Compound 6 and a pharmaceutically acceptable excipient.

66. A pharmaceutical composition comprising Compound 7 and a pharmaceutically acceptable excipient.

67. A pharmaceutical composition comprising Compound 8 and a pharmaceutically acceptable excipient.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1116 of 3077

68. A pharmaceutical composition comprising Compound 9 and a pharmaceutically acceptable excipient.

69. A pharmaceutical composition comprising Compound 10 and a pharmaceutically acceptable excipient.

70. A pharmaceutical composition comprising Compound 11 and a pharmaceutically acceptable excipient.

71. A pharmaceutical composition comprising Compound 12 and a pharmaceutically acceptable excipient.

72. A pharmaceutical composition comprising Compound 13 and a pharmaceutically acceptable excipient.

73. A pharmaceutical composition comprising Compound 14 and a pharmaceutically acceptable excipient.

74. A pharmaceutical composition comprising Compound 15 and a pharmaceutically acceptable excipient.

75. A pharmaceutical composition comprising Compound 16 and a pharmaceutically acceptable excipient.

76. A method of producing Compound 3, the method comprising:

providing a vector or vectors comprising the coding sequences of **SEQ ID NOs: 200-206 ;**

transforming a host cell with the vector or vectors;

incubating the host cell in culture media under conditions suitable for the expression of the coding sequences; and

isolating the compound produced by the host cell.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1117 of 3077

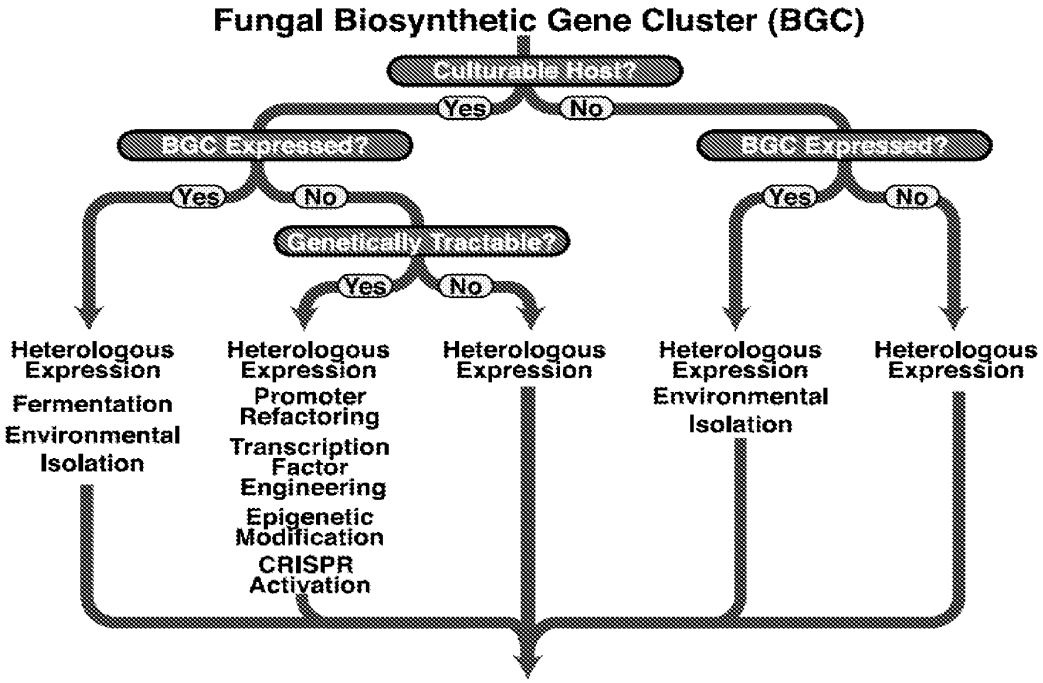

Fungal Biosynthetic Gene Cluster (BGC)
Culturable Host?
Yes
No
BGC Expressed?
Yes
No
Genetically Tractable?
Yes
No
BGC Expressed?
Yes
No
Heterologous Expression
Fermentation
Environmental Isolation
Heterologous Expression
Promoter Refactoring
Transcription Factor Engineering
Epigenetic Modification
CRISPR Activation
Heterologous Expression
Heterologous Expression
Environmental Isolation
Heterologous Expression

FIG. 1A

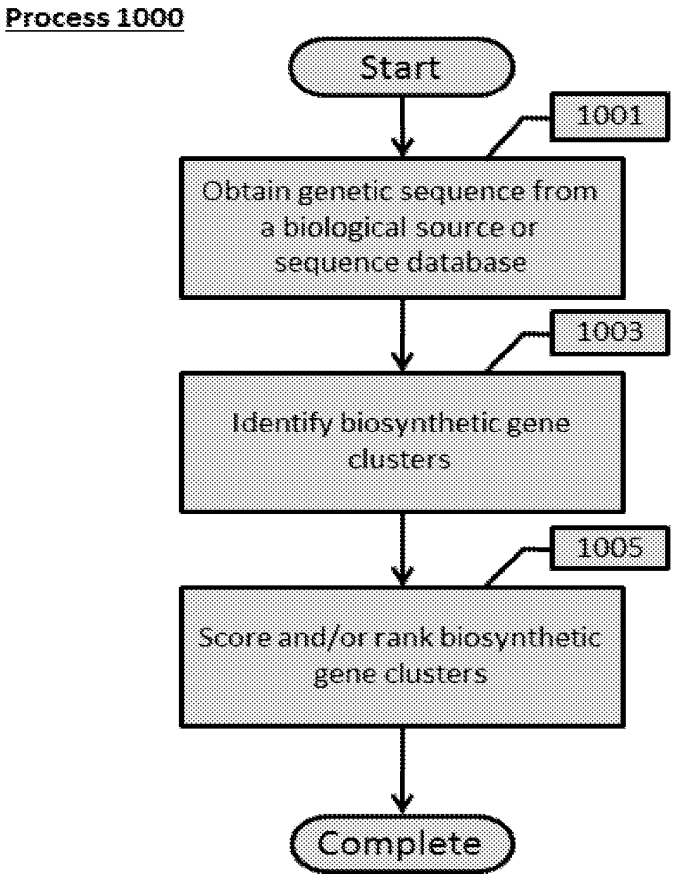

Process 1000
Start
1001
Obtain genetic sequence from a biological source or sequence database
1003
Identify biosynthetic gene clusters
1005
Score and/or rank biosynthetic gene clusters
Complete

FIG. 1B

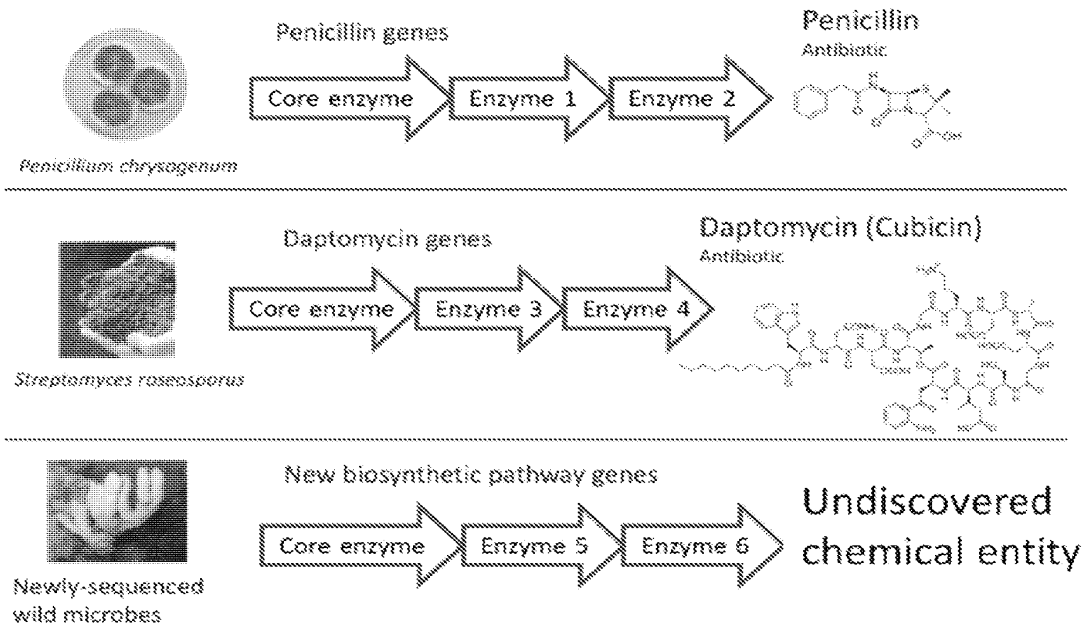

Penicillin genes
Core enzyme
Enzyme 1
Enzyme 2
Penicillin
Antibiotic
Penicillium chrysogenum
Daptomycin genes
Core enzyme
Enzyme 3
Enzyme 4
Daptomycin (Cubicin)
Antibiotic
Streptomyces roseosporus
New biosynthetic pathway genes
Core enzyme
Enzyme 5
Enzyme 6
Undiscovered chemical entity
Newly-sequenced wild microbes

FIG. 1C

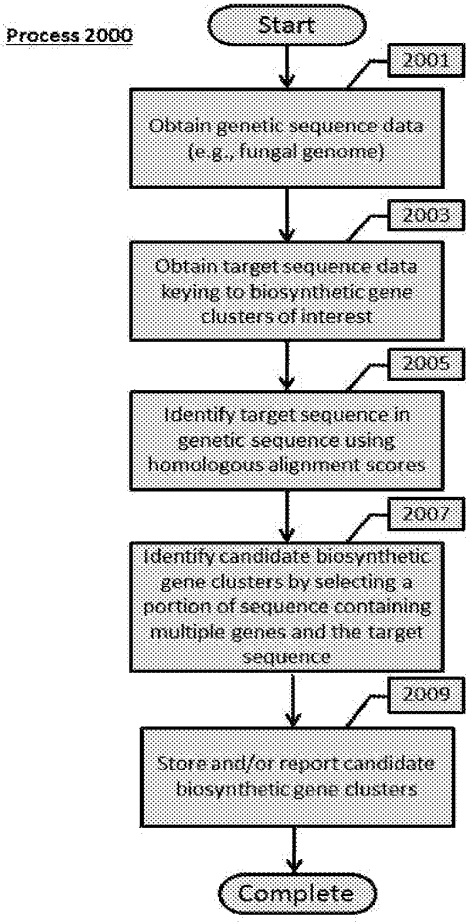

Process 2000
Start
2001
Obtain genetic sequence data (e.g., fungal genome)
2003
Obtain target sequence data keying to biosynthetic gene clusters of interest
2005
Identify target sequence in genetic sequence using homologous alignment scores
2007
Identify candidate biosynthetic gene clusters by selecting a portion of sequence containing multiple genes and the target sequence
2009
Store and/or report candidate biosynthetic gene clusters
Complete

FIG. 1D

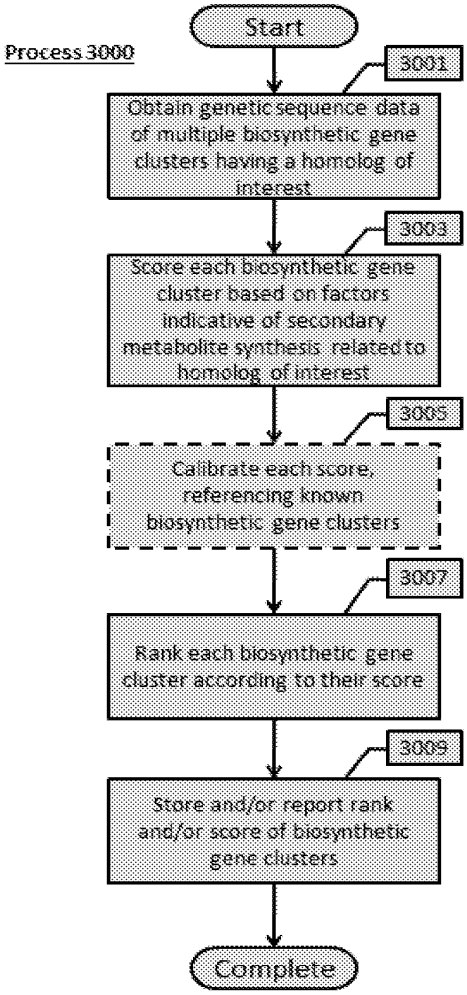

Process 3000
Start
3001
Obtain genetic sequence data of multiple biosynthetic gene clusters having a homolog of interest
3003
Score each biosynthetic gene cluster based on factors indicative of secondary metabolite synthesis related to homolog of interest
3005
Calibrate each score, referencing known biosynthetic gene clusters
3007
Rank each biosynthetic gene cluster according to their score
3009
Store and/or report rank and/or score of biosynthetic gene clusters
Complete

FIG. 1E

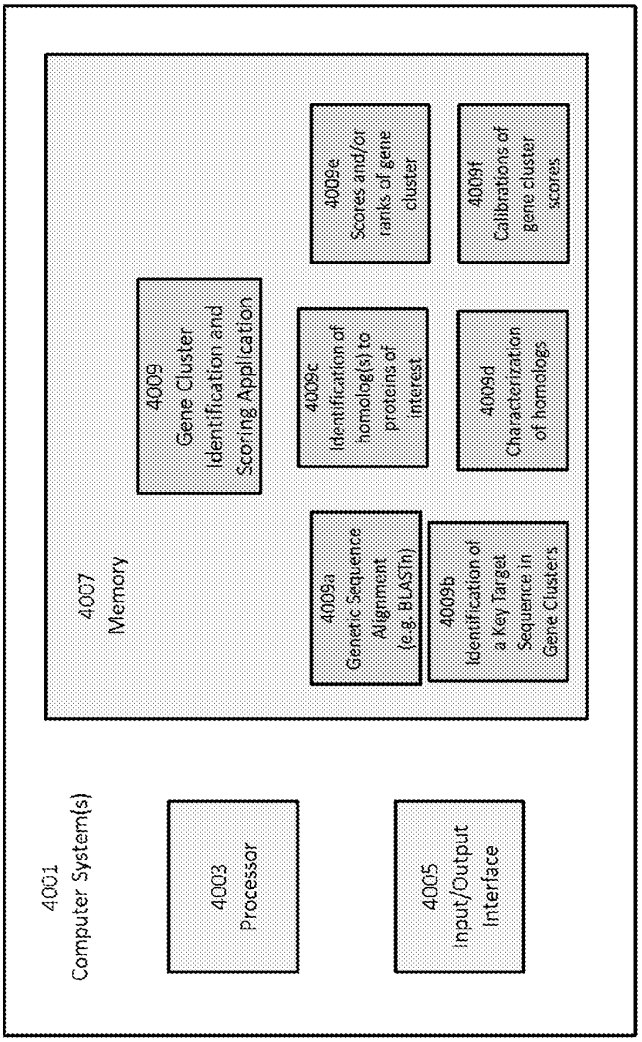

4001
Computer System(s)
4003
Processor
4005
Input/Output
Interface
4007
Memory
4009
Gene Cluster
Identification and
Scoring Application
4009a
Genetic Sequence
Alignment
(e.g. BLASTn)
4009b
Identification of
a Key Target
Sequence in
Gene Clusters
4009c
Identification of
homolog(s) to
proteins of
interest
4009d
Characterization
of homologs
4009e
Scores and/or
ranks of gene
cluster
4009f
Calibrations of
gene cluster
scores

FIG. 1F

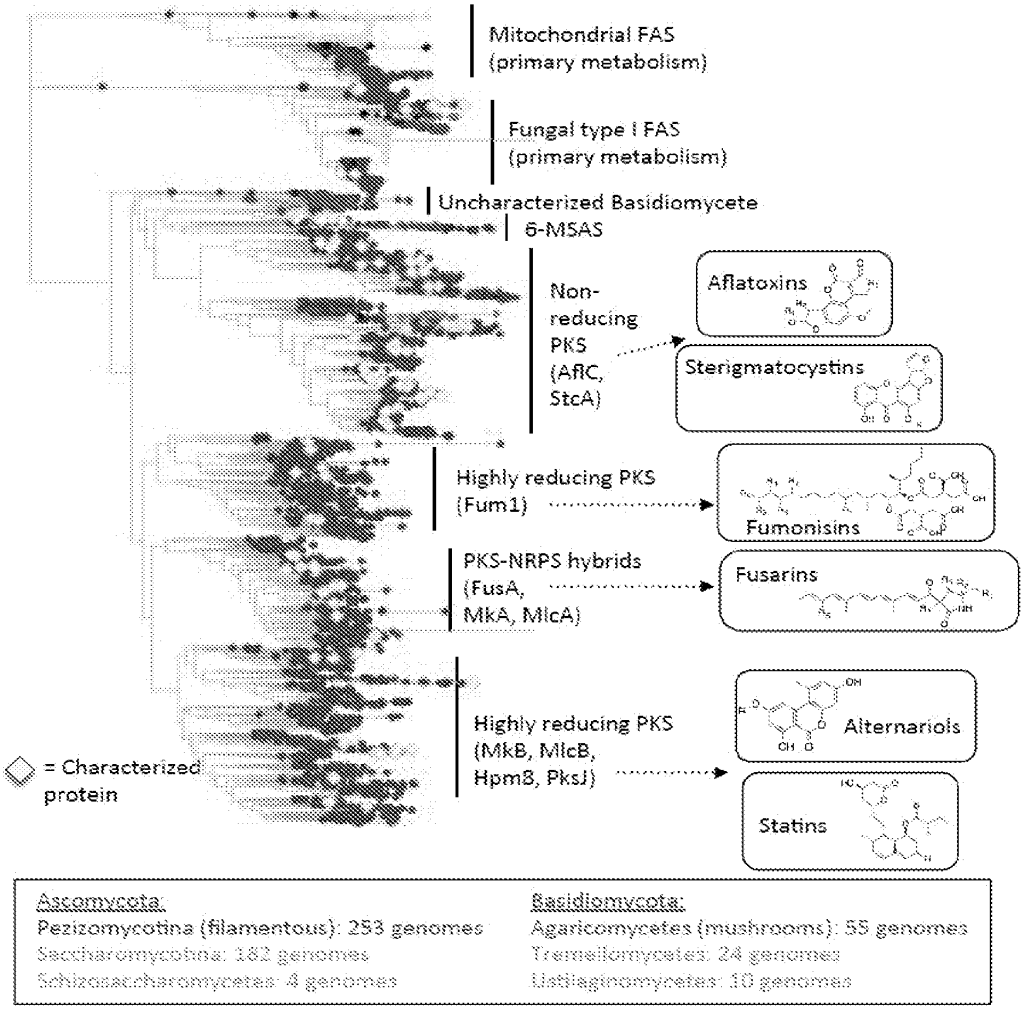
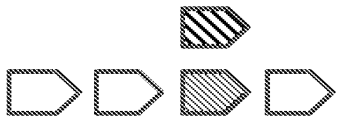

Mitochondrial FAS
(primary metabolism)
Fungal type I FAS
(primary metabolism)
Uncharacterized Basidiomycete
6-MSAS
Non-reducing PKS (AflC, StcA)
Aflatoxins
Sterigmatocystins
Highly reducing PKS (Fum1)
Fumonisins
PKS-NRPS hybrids (FusA, MkA, MlcA)
Fusarins
Highly reducing PKS (MkB, MlcB, Hpm8, PksJ)
Alternariols
Statins
= Characterized protein
Ascomycota:
Pezizomycotina (filamentous): 253 genomes
Saccharomycotina: 182 genomes
Schizosaccharomycetes: 4 genomes
Basidiomycota:
Agaricomycetes (mushrooms): 55 genomes
Tremellomycetes: 24 genomes
Ustilaginomycetes: 10 genomes

FIG. 2

Self-resistance mechanisms to toxic natural products



The hatched gene encodes mutated target protein, which is resistant to natural product

Striped gene encodes a protein that performs enzymatic modification of toxic small molecule into a nontoxic form

Dotted gene encodes a transporter protein that pumps the natural product out of the host cell

FIG. 3

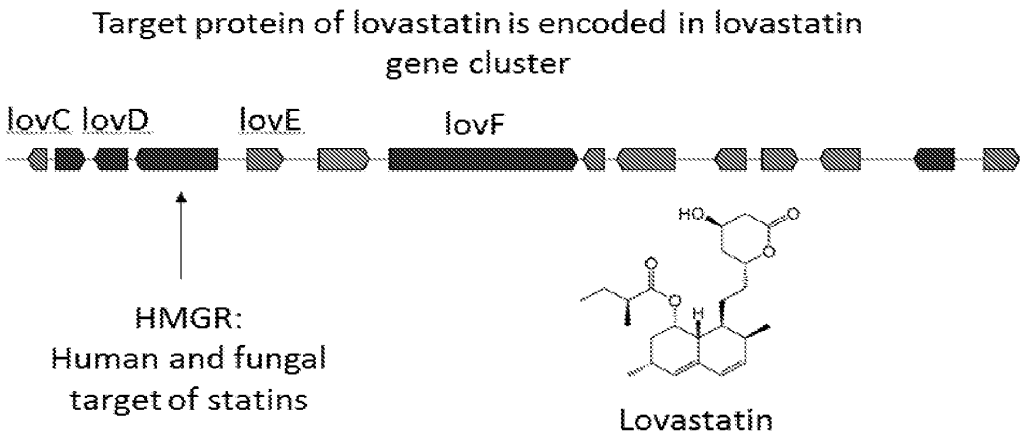

Target protein of lovastatin is encoded in lovastatin gene cluster
lovC lovD
lovE
lovF
HMGR:
Human and fungal
target of statins
HO
O
O
O
O
H
Lovastatin

FIG. 4

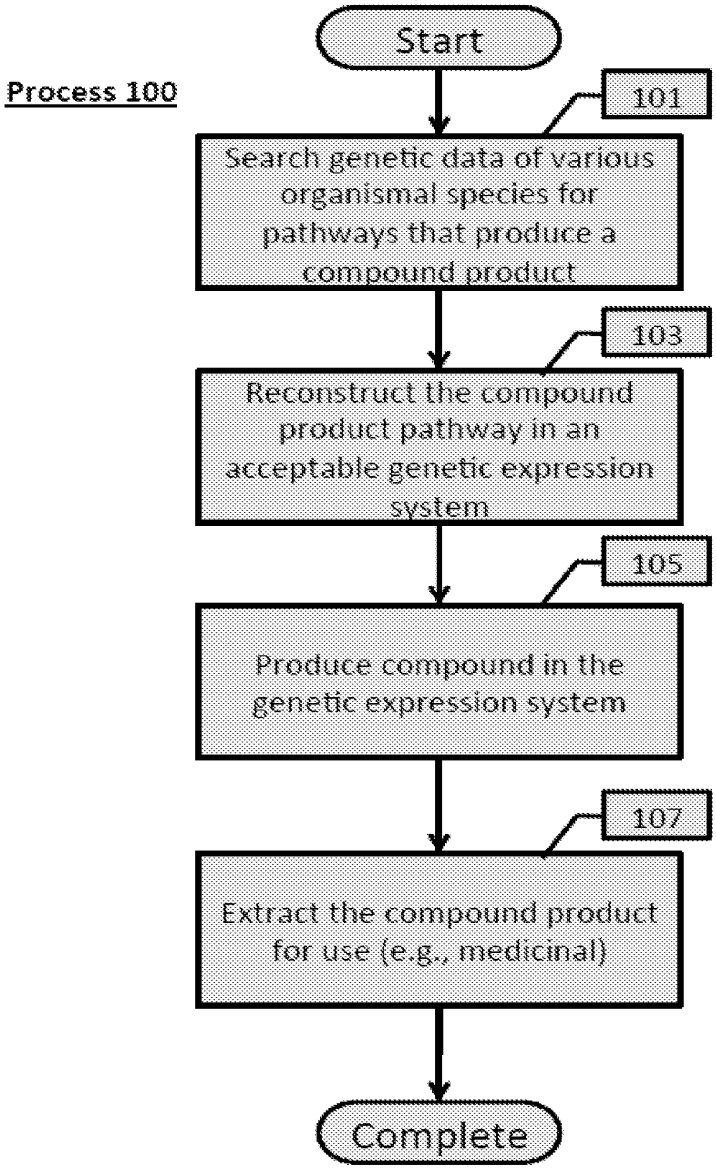

Process 100
Start
101
Search genetic data of various organismal species for pathways that produce a compound product
103
Reconstruct the compound product pathway in an acceptable genetic expression system
105
Produce compound in the genetic expression system
107
Extract the compound product for use (e.g., medicinal)
Complete

FIG. 5A

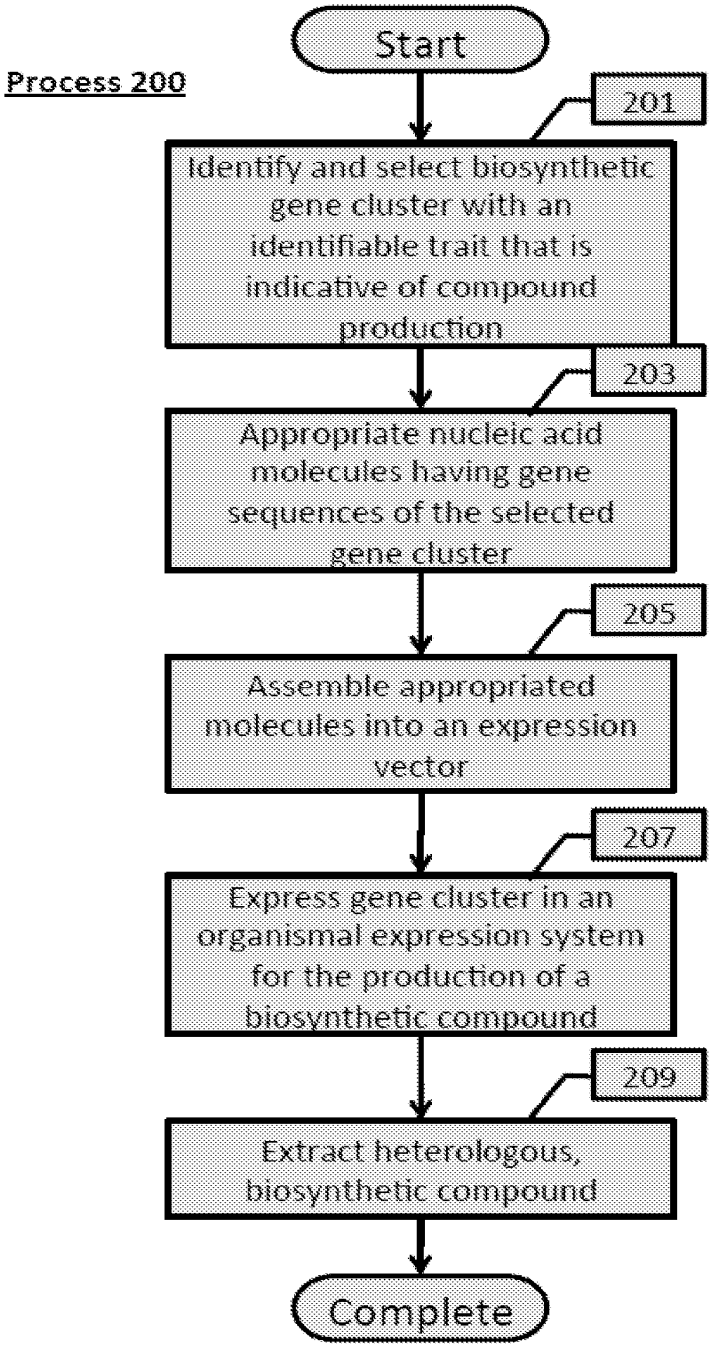

Process 200
Start
201
Identify and select biosynthetic gene cluster with an identifiable trait that is indicative of compound production
203
Appropriate nucleic acid molecules having gene sequences of the selected gene cluster
205
Assemble appropriated molecules into an expression vector
207
Express gene cluster in an organismal expression system for the production of a biosynthetic compound
209
Extract heterologous, biosynthetic compound
Complete

FIG. 5B

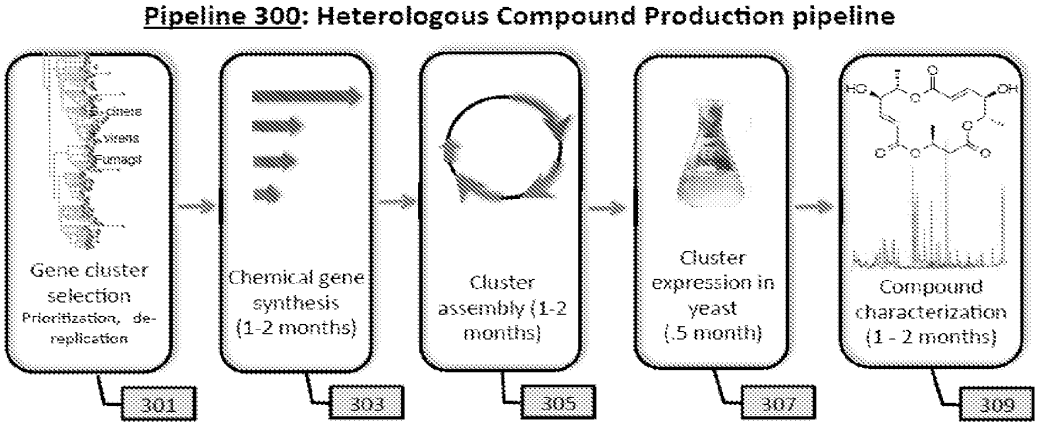

Pipeline 300: Heterologous Compound Production pipeline
Gene cluster selection
Prioritization, de-replication
301
Chemical gene synthesis (1-2 months)
303
Cluster assembly (1-2 months)
305
Cluster expression in yeast (.5 month)
307
Compound characterization (1 - 2 months)
309

FIG. 5C

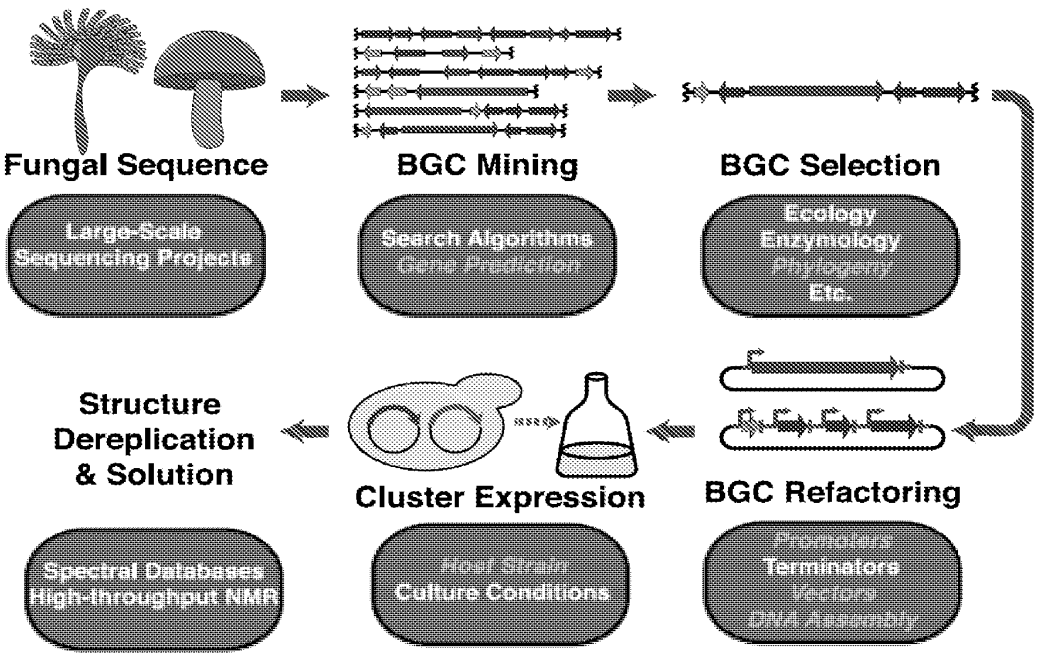

Fungal Sequence
Large-Scale Sequencing Projects
BGC Mining
Search Algorithms
Gene Prediction
BGC Selection
Ecology
Enzymology
Phylogeny
Etc.
BGC Refactoring
Promoters
Terminators
Vectors
DNA Assembly
Cluster Expression
Host Strain
Culture Conditions
Structure Dereplication & Solution
Spectral Databases
High-throughput NMR

FIG. 5D

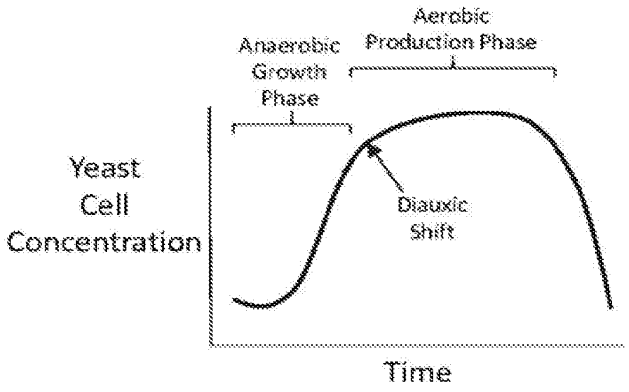


FIG. 6A

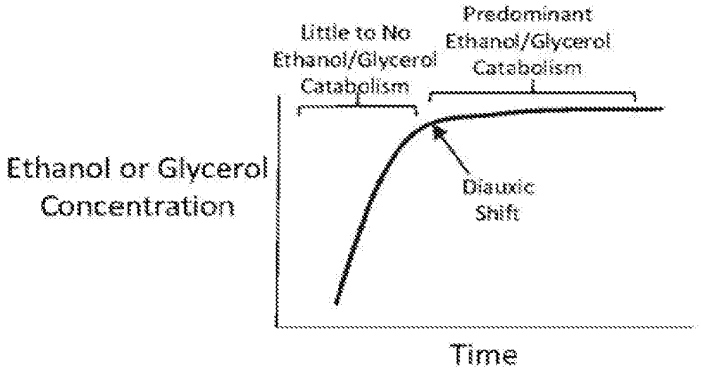


FIG. 6B

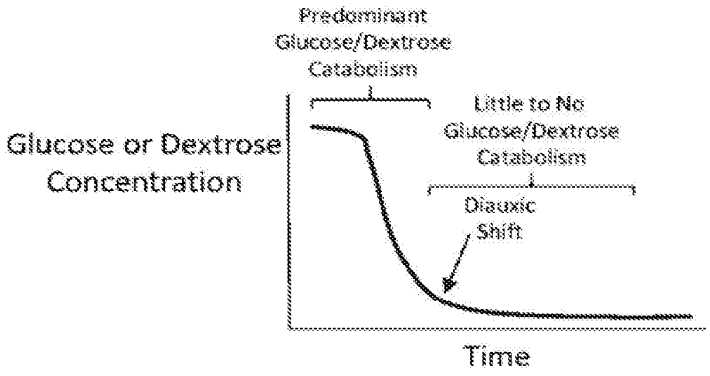


FIG. 6C

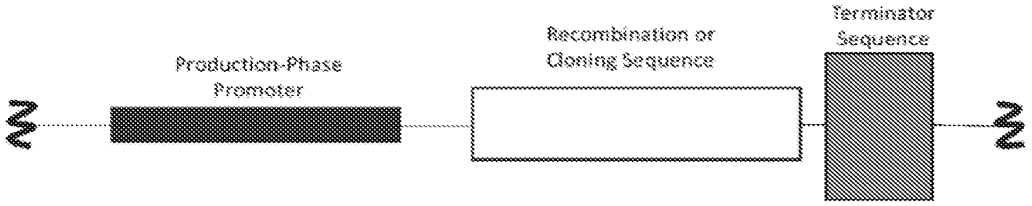

Production-Phase Promoter
Recombination or Cloning Sequence
Terminator Sequence

FIG. 7A

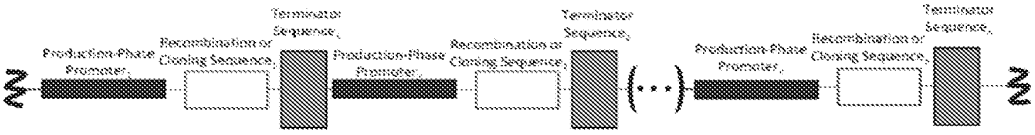

Production-Phase Promoter
Recombination or Cloning Sequence
Terminator Sequence
Production-Phase Promoter
Recombination or Cloning Sequence
Terminator Sequence
Production-Phase Promoter
Recombination or Cloning Sequence
Terminator Sequence

FIG. 7B

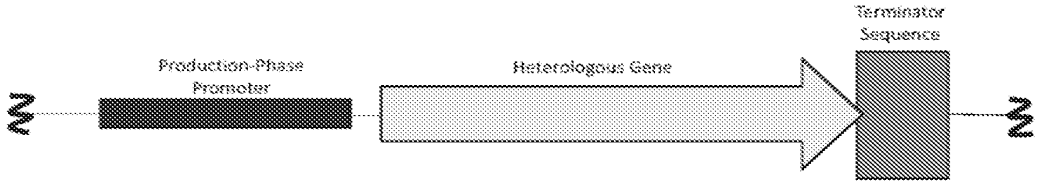

Terminator Sequence
Production-Phase Promoter
Heterologous Gene

FIG. 8A

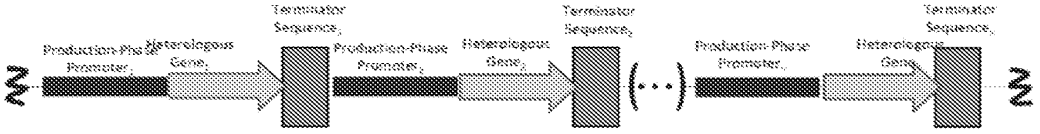

Terminator Sequence1
Terminator Sequence2
Terminator Sequencen
Production-Phase Promoter1
Heterologous Gene1
Production-Phase Promoter2
Heterologous Gene2
Production-Phase Promotern
Heterologous Genen

FIG. 8B

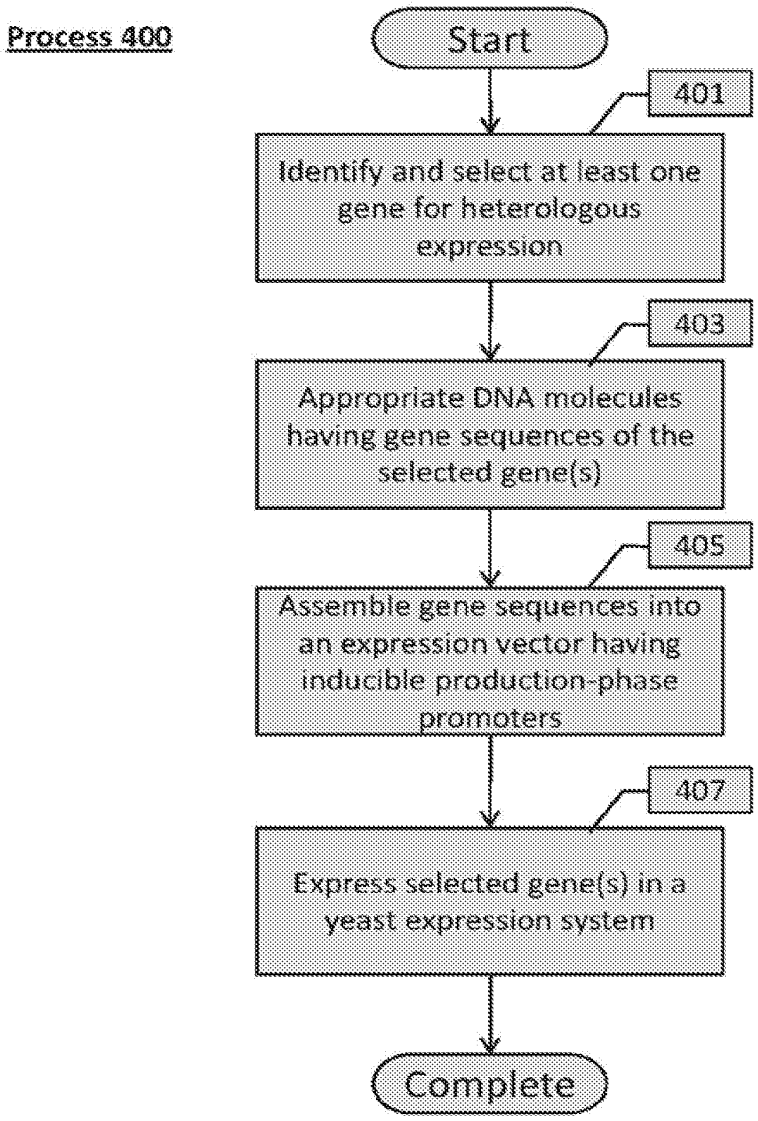
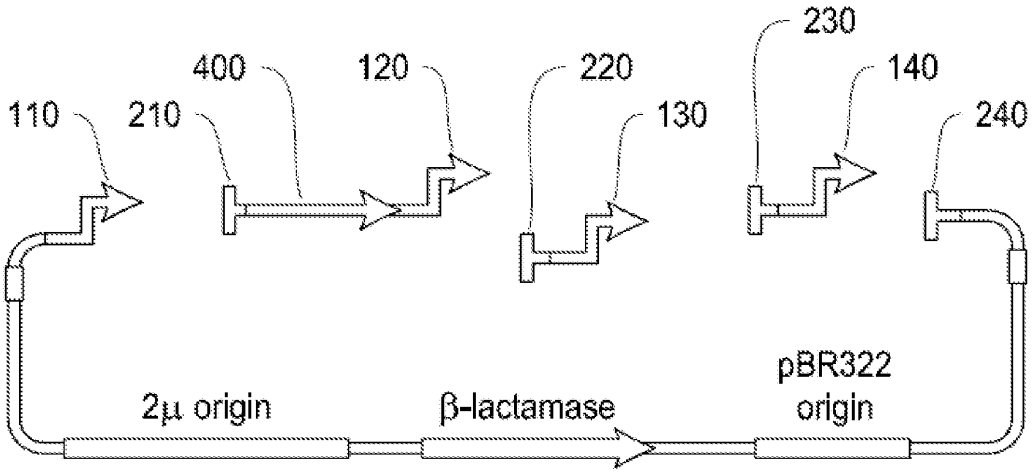

Process 400
Start
401
Identify and select at least one gene for heterologous expression
403
Appropriate DNA molecules having gene sequences of the selected gene(s)
405
Assemble gene sequences into an expression vector having inducible production-phase promoters
407
Express selected gene(s) in a yeast expression system
Complete

FIG. 9


110
210
400
120
220
130
230
140
240
2μ origin
β-lactamase
pBR322
origin

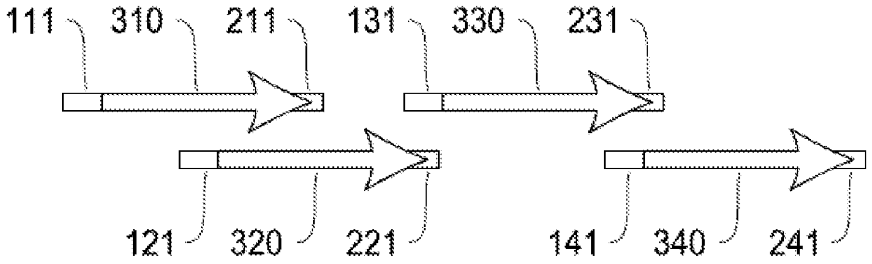

111
310
211
131
330
231
121
320
221
141
340
241

FIG. 10A

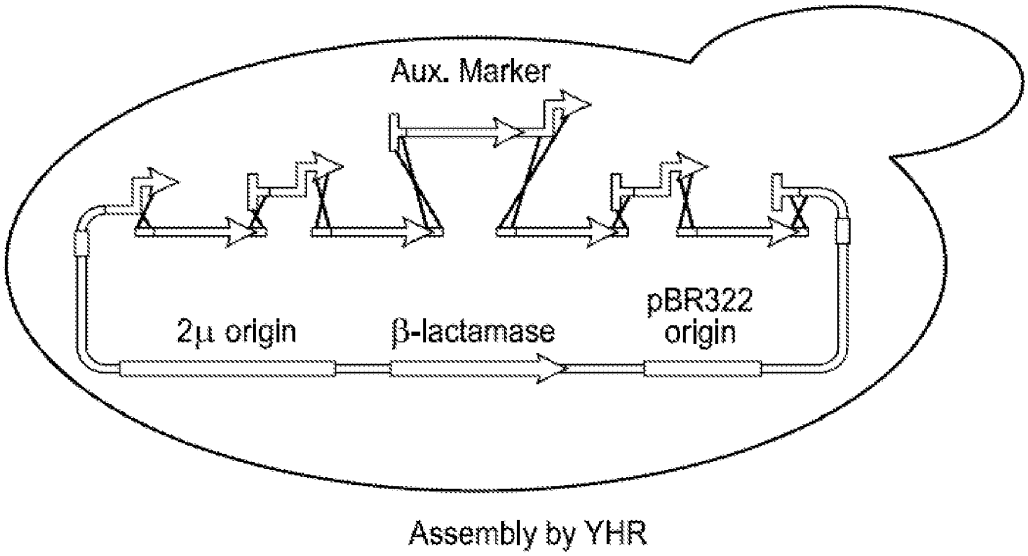


FIG. 10B

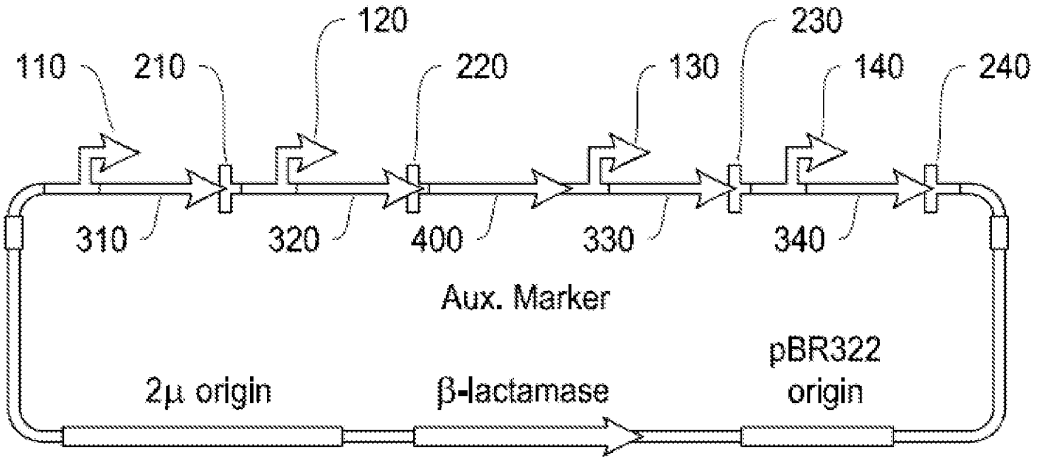


FIG. 10C

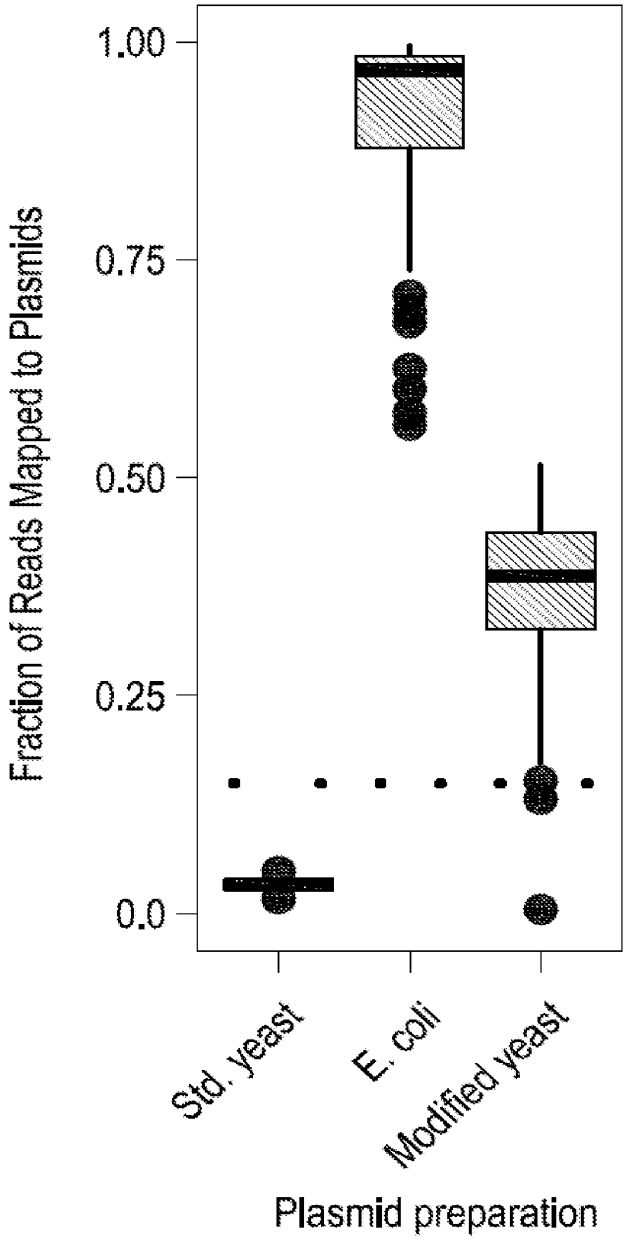

1.00
0.75
0.50
0.25
0.0
Fraction of Reads Mapped to Plasmids
Std. yeast
E. coli
Modified yeast
Plasmid preparation

FIG. 10D

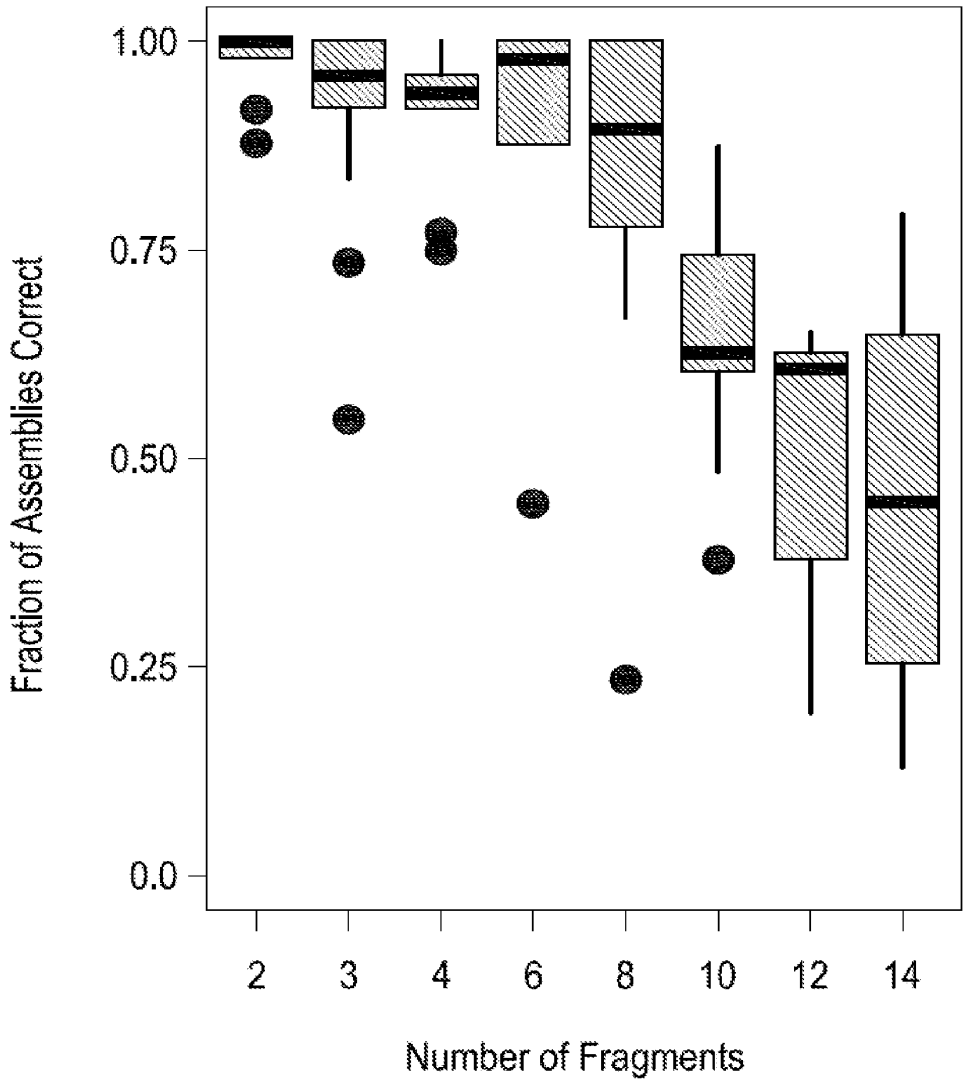

1.00
0.75
0.50
0.25
0.0
Fraction of Assemblies Correct
2
3
4
6
8
10
12
14
Number of Fragments

FIG. 10E

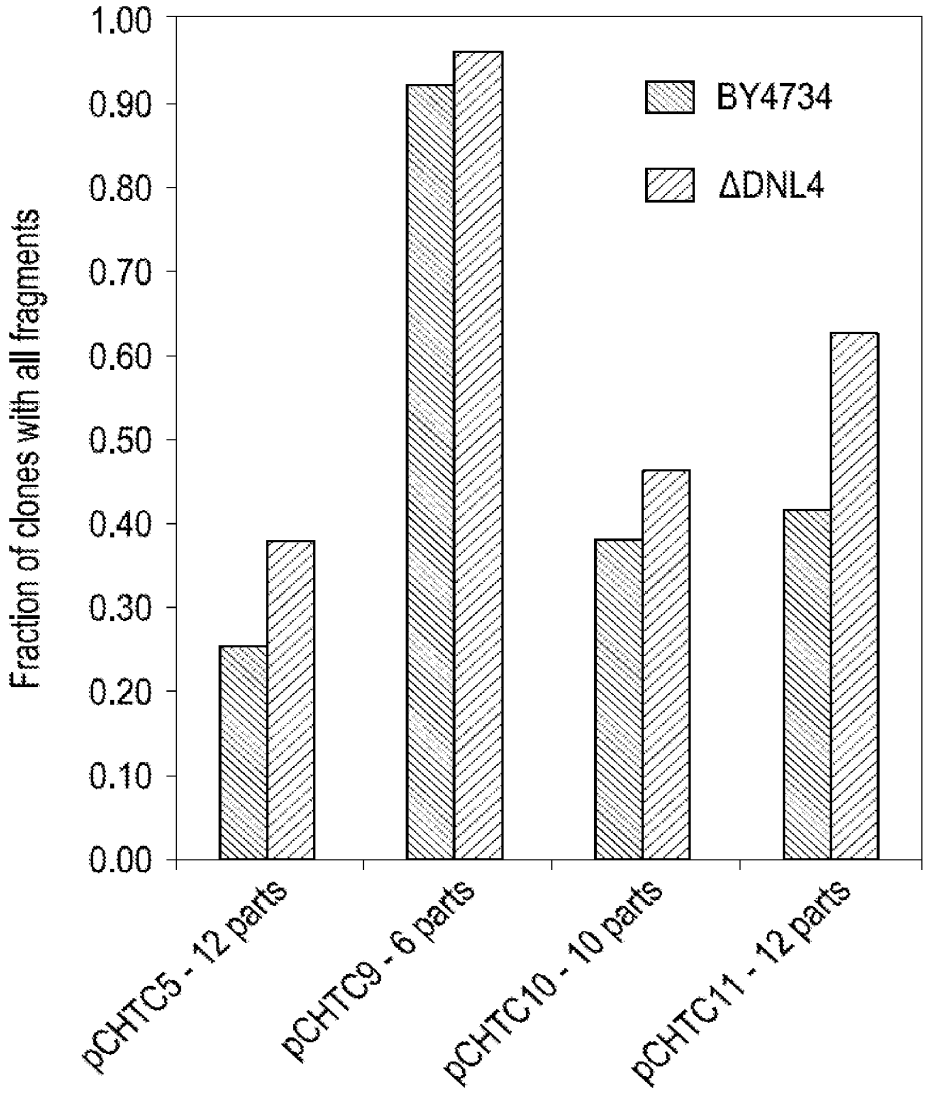

1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
Fraction of clones with all fragments
BY4734
ΔDNL4
pCHTC5 - 12 parts
pCHTC9 - 6 parts
pCHTC10 - 10 parts
pCHTC11 - 12 parts

FIG. 11A

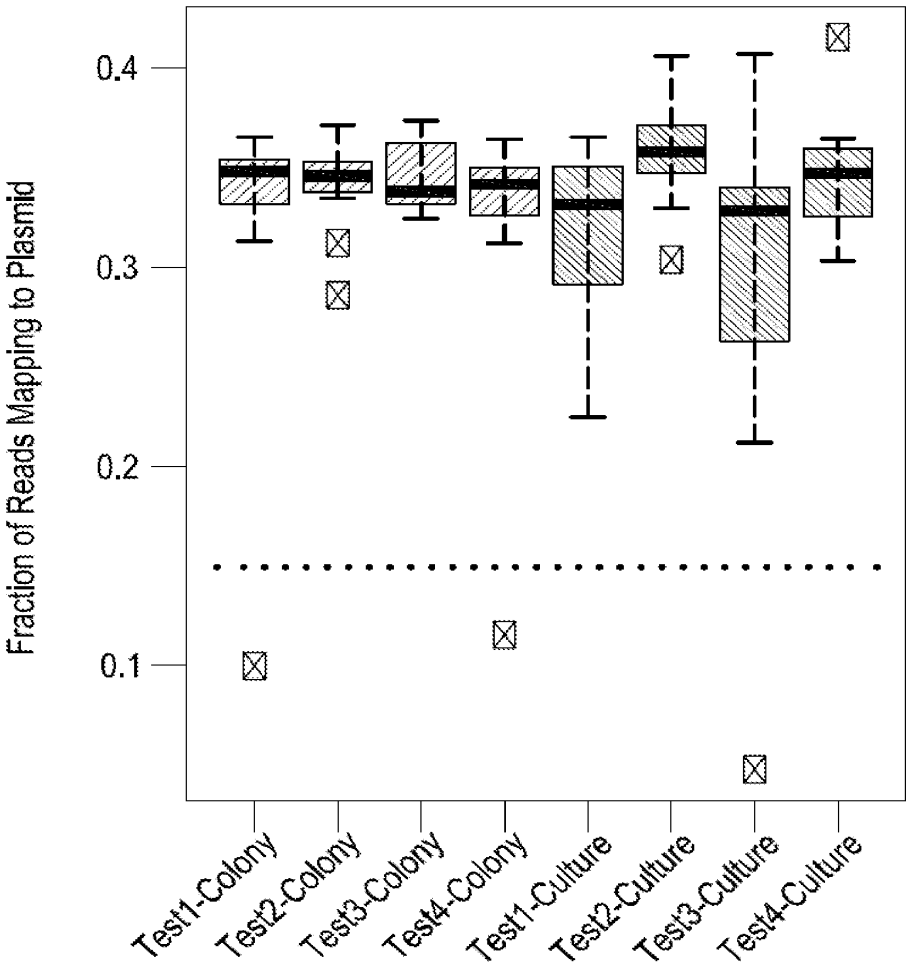

Fraction of Reads Mapping to Plasmid
0.4
0.3
0.2
0.1
Test1-Colony
Test2-Colony
Test3-Colony
Test4-Colony
Test1-Culture
Test2-Culture
Test3-Culture
Test4-Culture

FIG. 11B

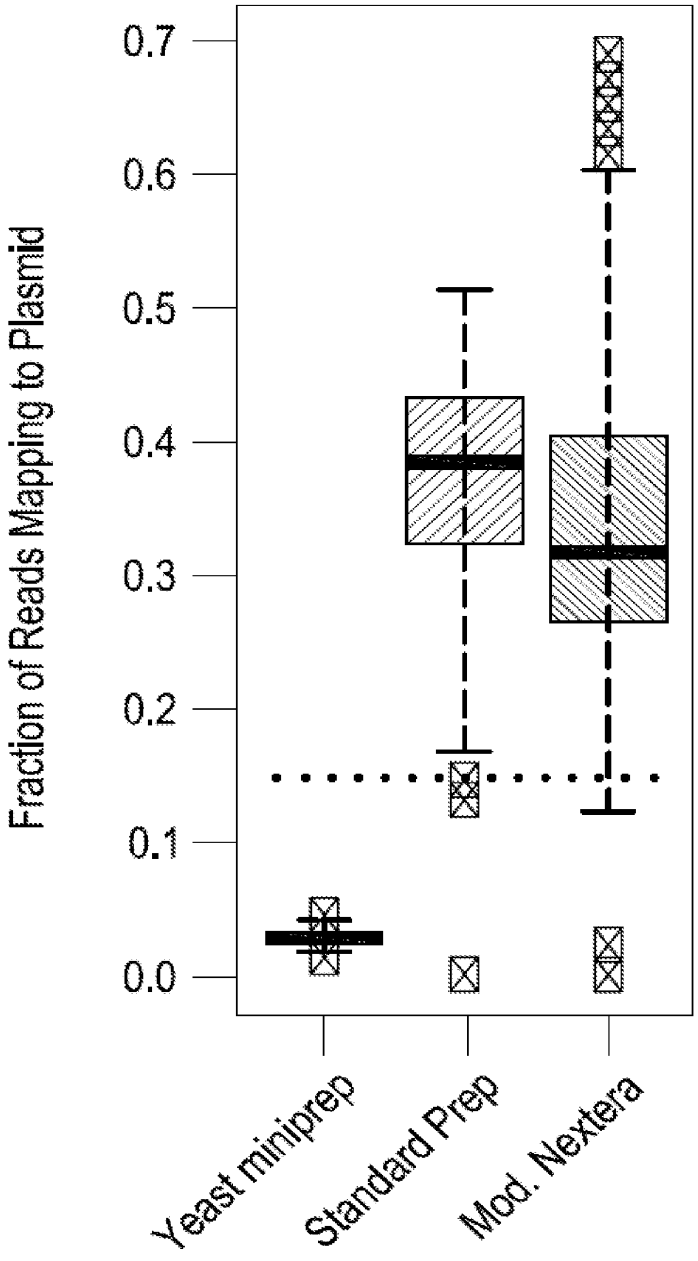

Fraction of Reads Mapping to Plasmid
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Yeast miniprep
Standard Prep
Mod. Nextera

FIG. 11C

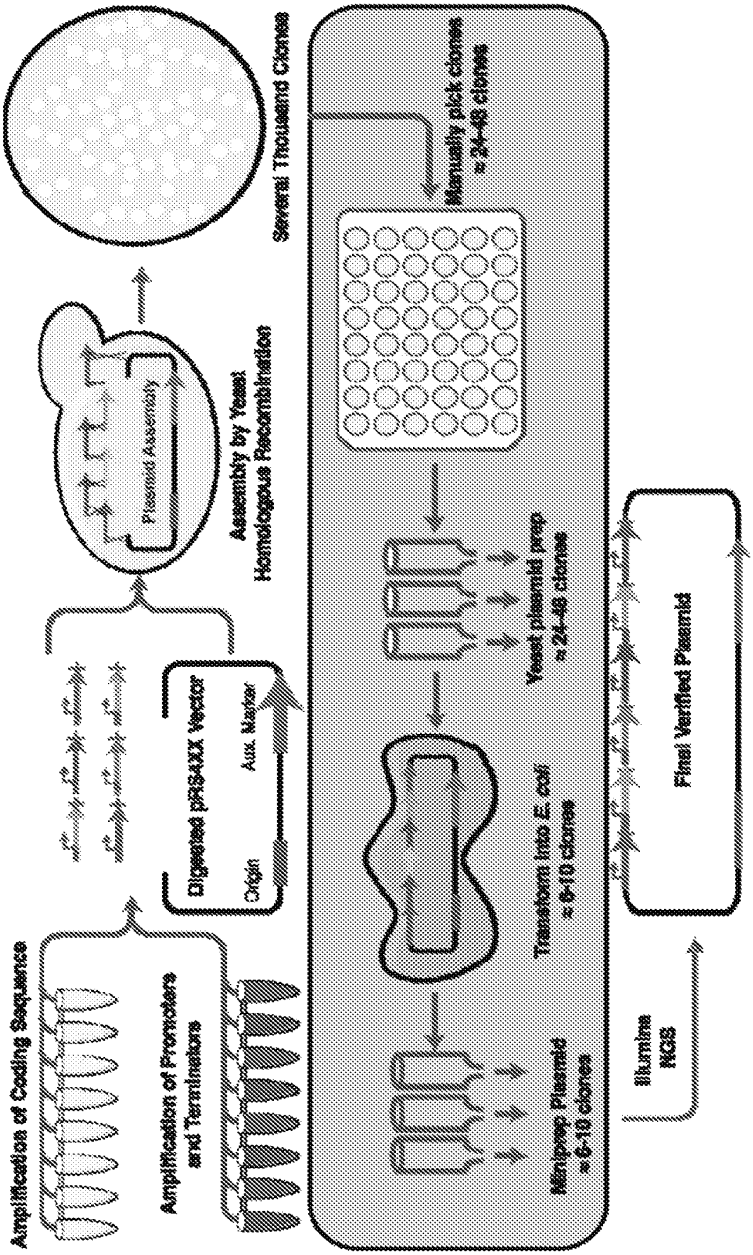

Several Thousand Clones
Manually pick clones
~ 24-48 clones
Assembly by Yeast
Homologous Recombination
Plasmid Assembly
Digested pRS4XX Vector
Origin
Aux. Marker
Amplification of Coding Sequence
Amplification of Promoters
and Terminators
Yeast plasmid prep
~ 24-48 clones
Transform into E. coli
~ 6-10 clones
Miniprep Plasmid
~ 6-10 clones
Illumina
NGS
Final Verified Plasmid

FIG. 11D

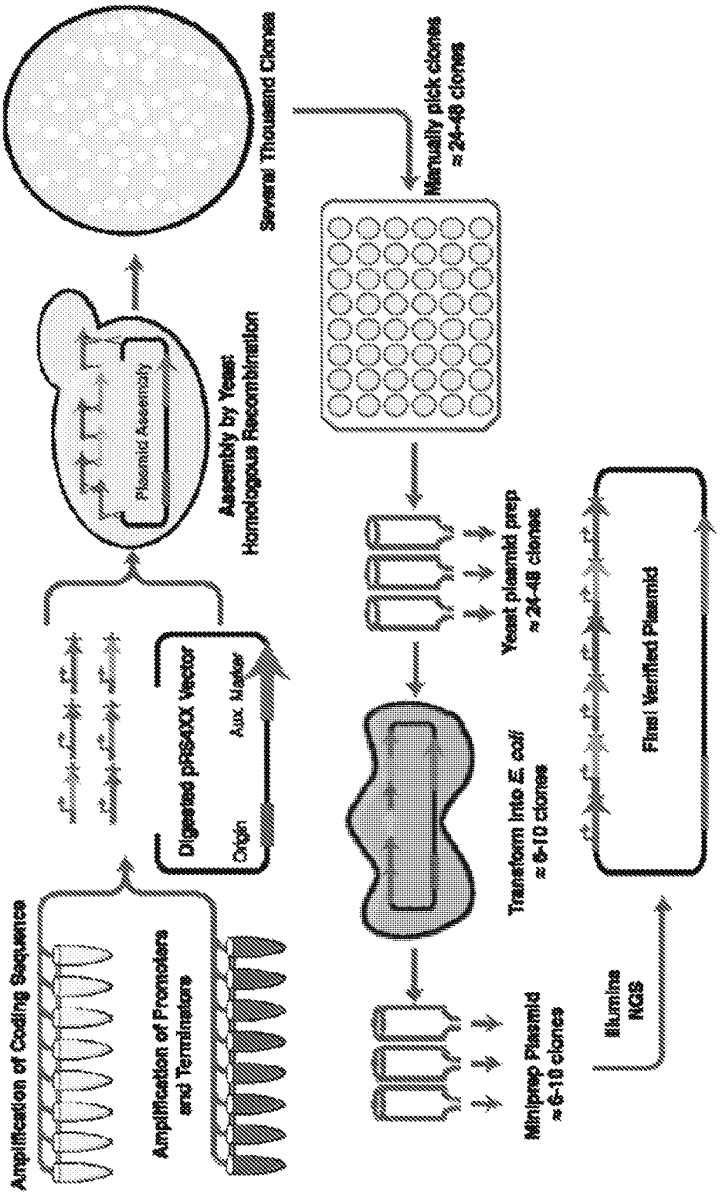

Amplification of Coding Sequence
Amplification of Promoters and Terminators
Digested pRS40X Vector
Origin
Aux. Marker
Assembly by Yeast Homologous Recombination
Plasmid Assembly
Several Thousand Clones
Manually pick clones ≈ 24-48 clones
Yeast plasmid prep ≈ 24-48 clones
Transform into E. coli ≈ 6-10 clones
Miniprep Plasmid ≈ 6-10 clones
Illumina NGS
Final Verified Plasmid

FIG. 11E

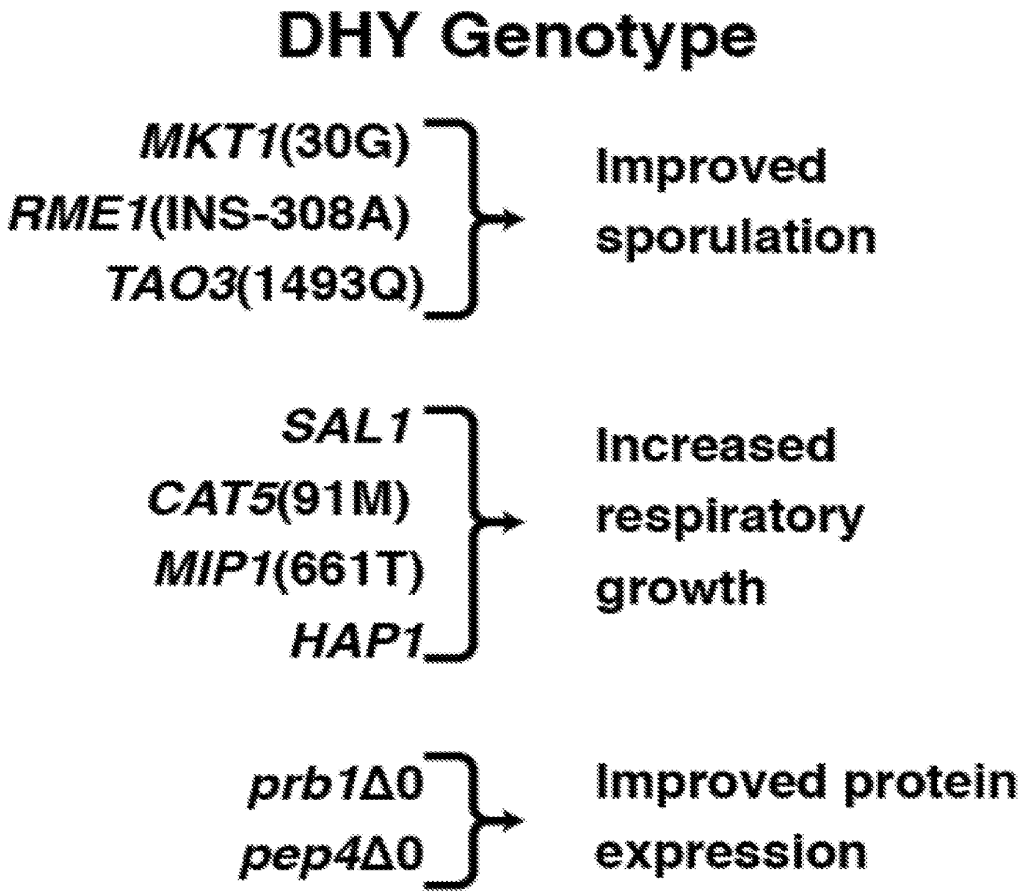

DHY Genotype
MKT1(30G)
RME1(INS-308A)
TAO3(1493Q)
Improved sporulation
SAL1
CAT5(91M)
MIP1(661T)
HAP1
Increased respiratory growth
prb1Δ0
pep4Δ0
Improved protein expression

FIG. 12

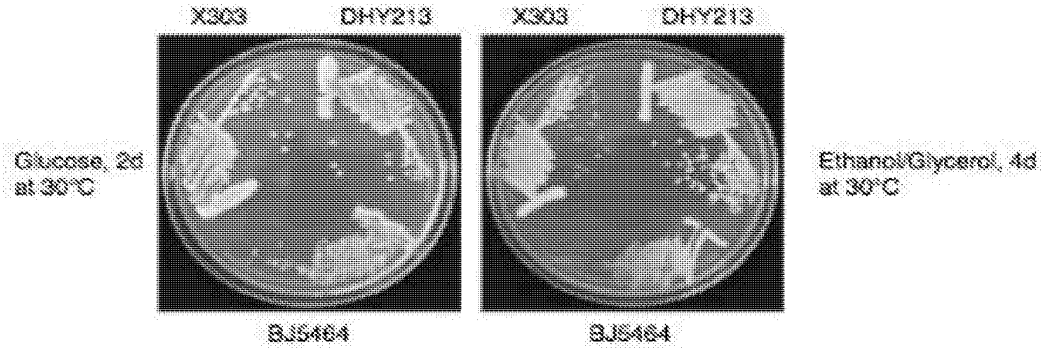

X303
DHY213
X303
DHY213
Glucose, 2d at 30°C
Ethanol/Glycerol, 4d at 30°C
BJ5464
BJ5464

FIG. 13

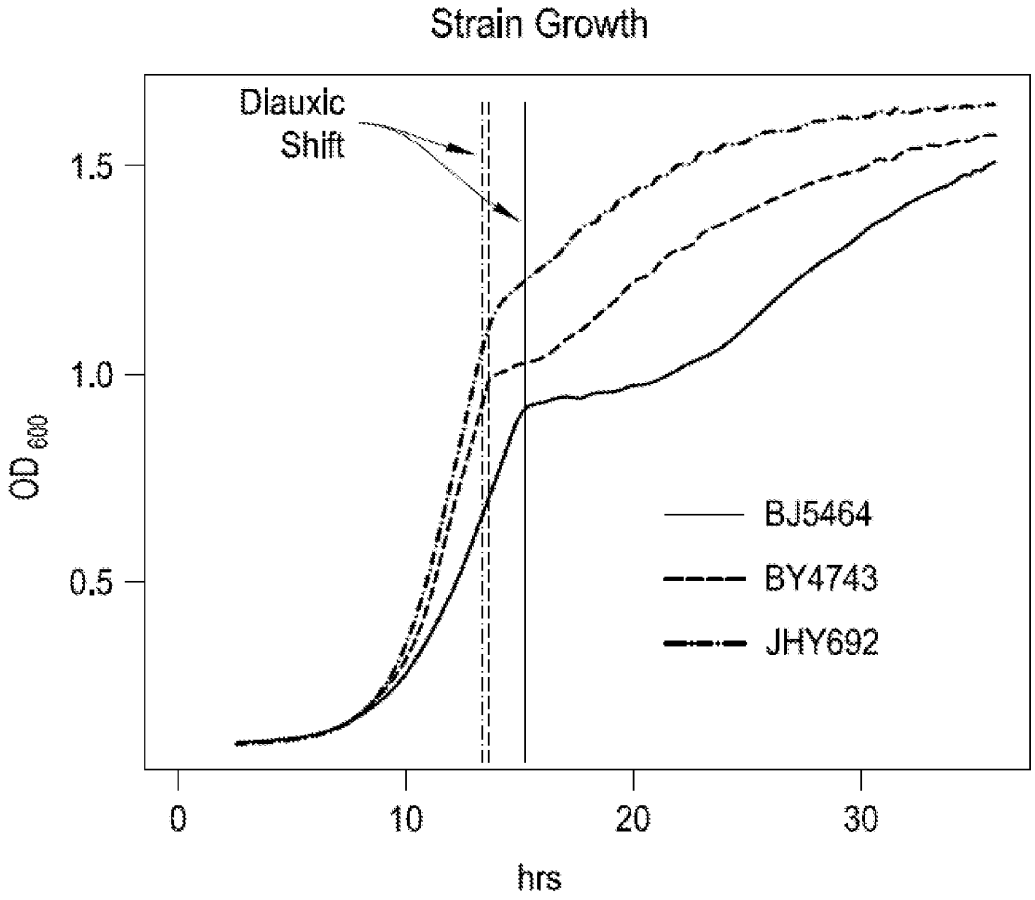

Strain Growth
Diauxic Shift
1.5
1.0
0.5
OD600
BJ5464
BY4743
JHY692
0
10
20
30
hrs

FIG. 14A

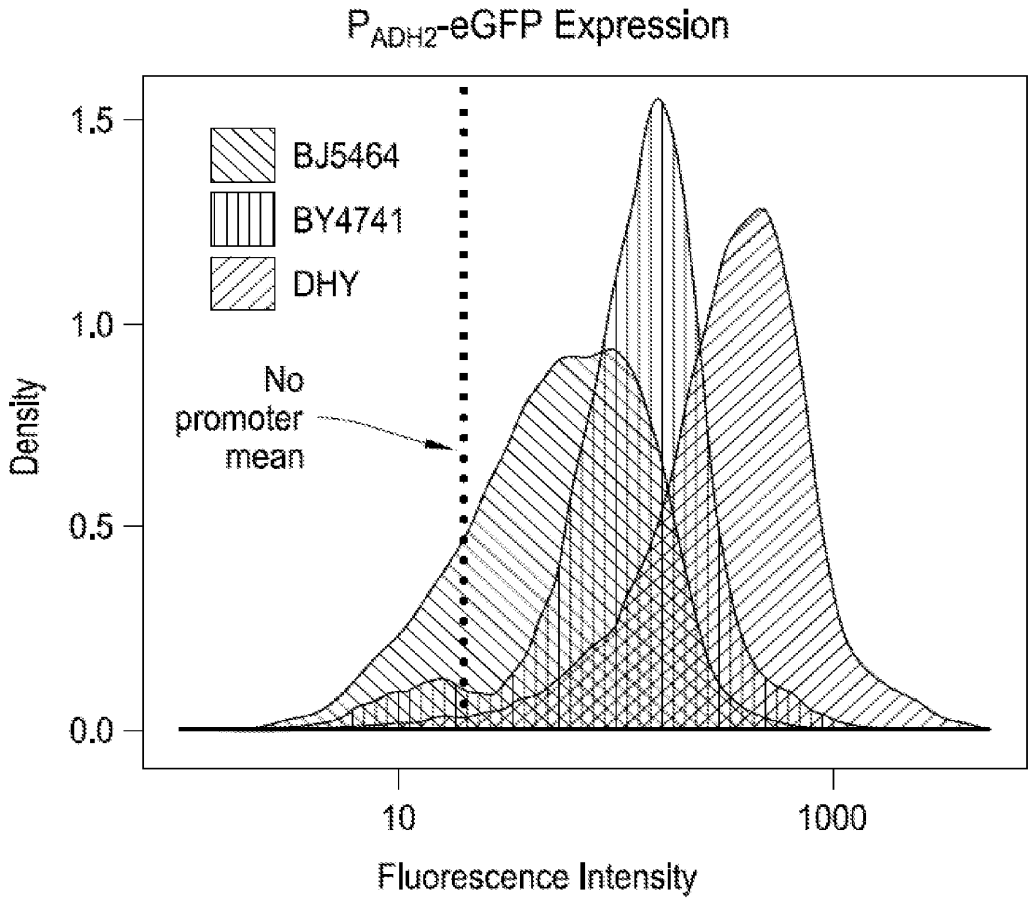

P_{ADH2}-eGFP Expression
BJ5464
BY4741
DHY
No promoter mean
Density
1.5
1.0
0.5
0.0
10
1000
Fluorescence Intensity

FIG. 14B

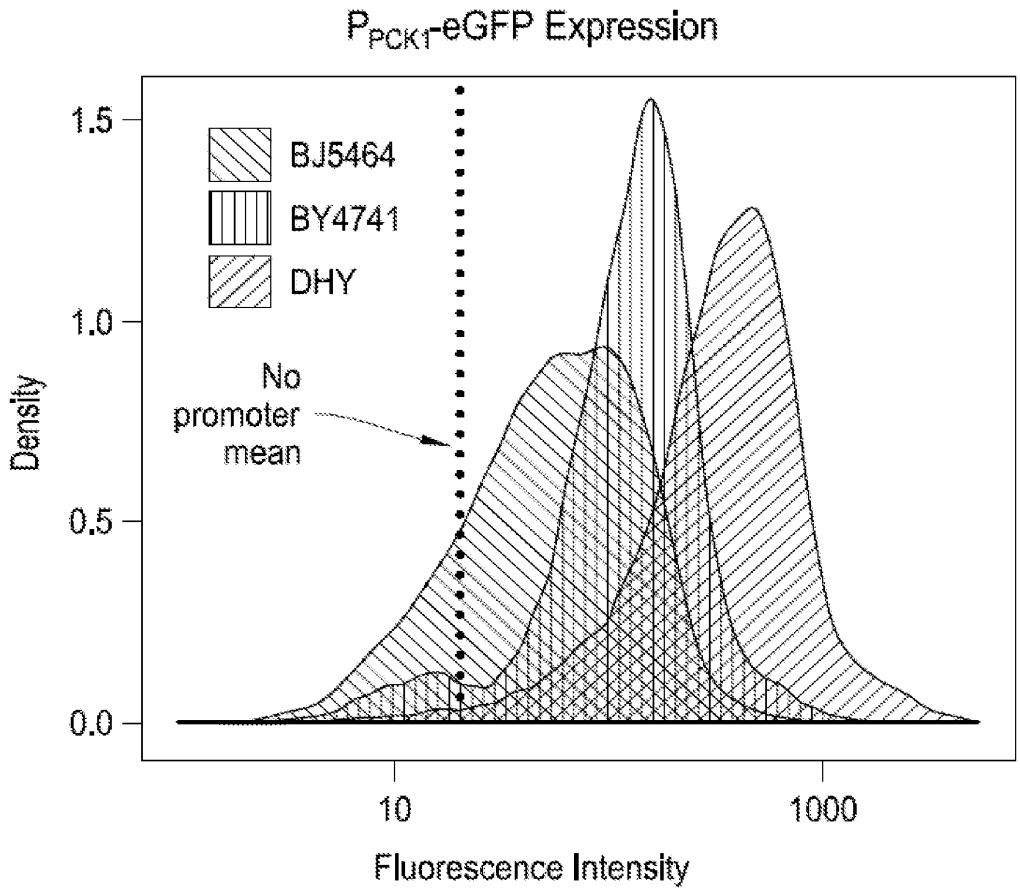

P_{PCK1}-eGFP Expression
BJ5464
BY4741
DHY
No promoter mean
Density
1.5
1.0
0.5
0.0
10
1000
Fluorescence Intensity

FIG. 14C

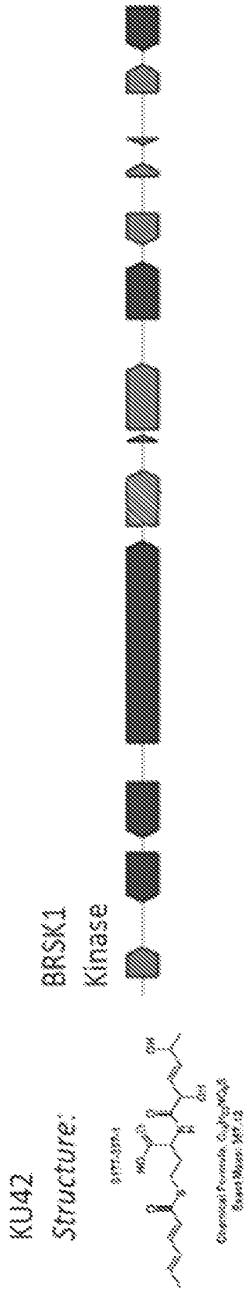


FIG. 15A

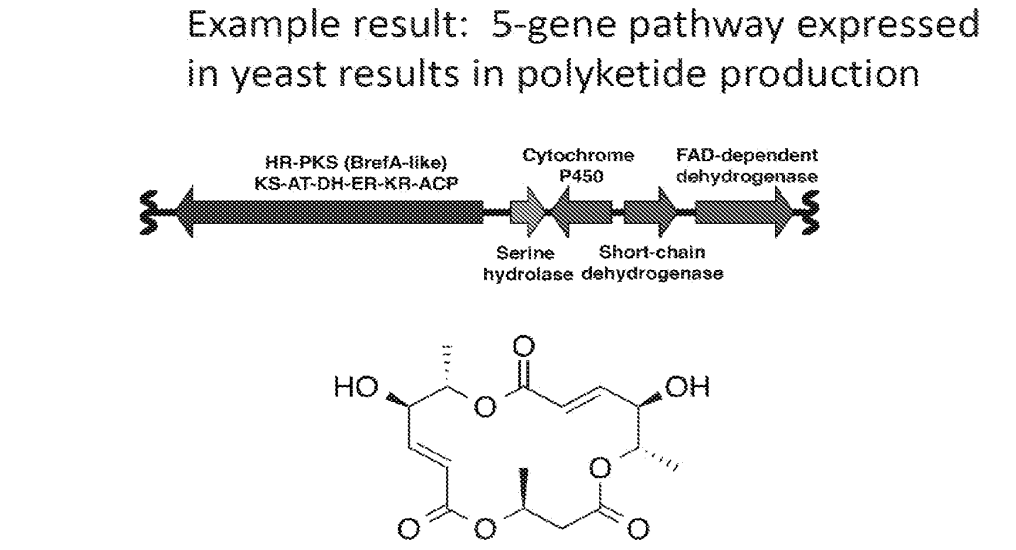

Example result: 5-gene pathway expressed in yeast results in polyketide production
HR-PKS (BrefA-like)
KS-AT-DH-ER-KR-ACP
Cytochrome P450
FAD-dependent dehydrogenase
Serine hydrolase
Short-chain dehydrogenase
HO
OH
O
Ascomycota; Pezizomycotina; Dothideomycetes; Pleosporomycetidae; Pleosporales; Massarineae; Didymosphaeriaceae; Paraphaeosphaeria
Microsphaeropis
Ascomycota; Pezizomycotina; Dothideomycetes; Pleosporomycetidae; Pleosporales; Pleosporineae; Pleosporineae incertae sedis; Microsphaeropsis.

FIG. 15B

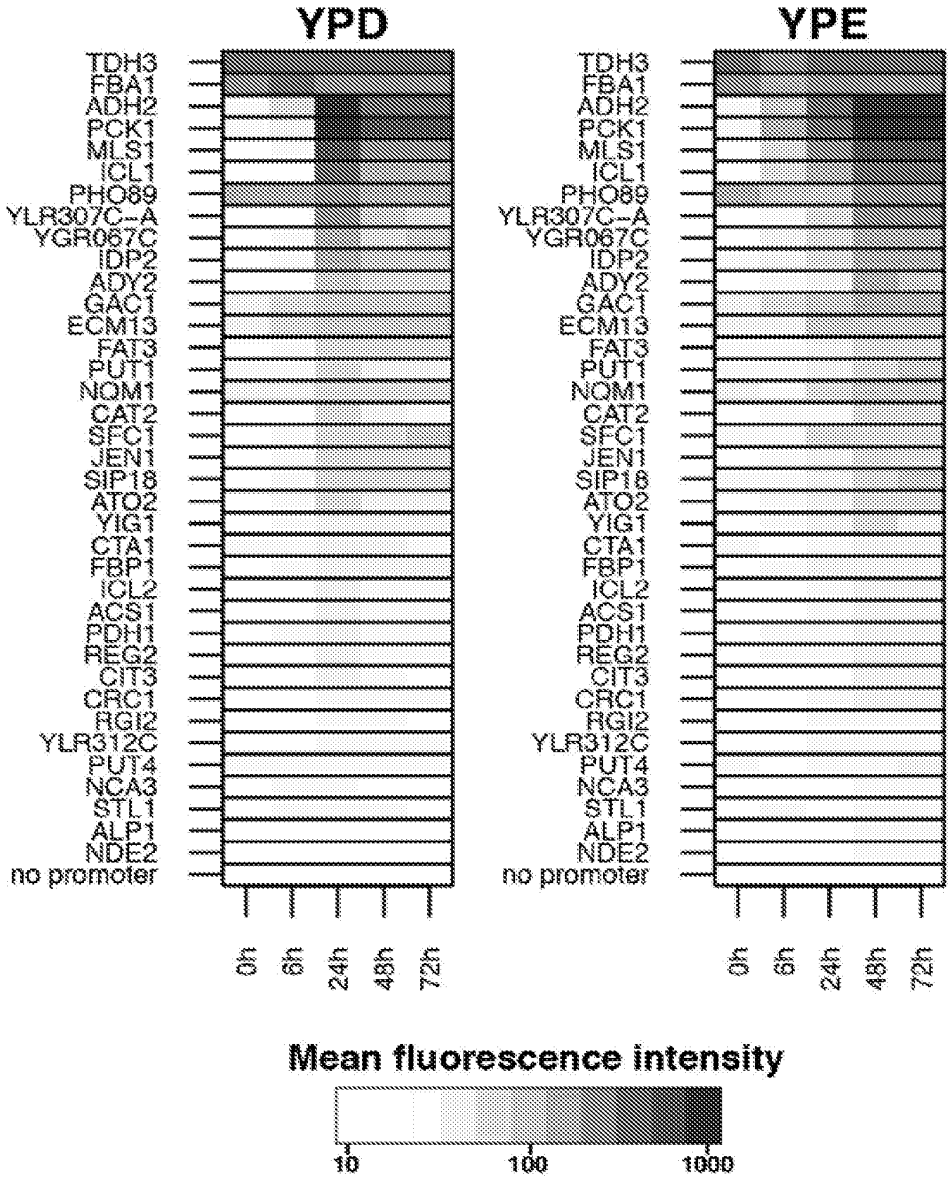

YPD
YPE
TDH3
FBA1
ADH2
PCK1
MLS1
ICL1
PHO89
YLR307C-A
YGR067C
IDP2
ADY2
GAC1
ECM13
FAT3
PUT1
NQM1
CAT2
SFC1
JEN1
SIP18
ATO2
YIG1
CTA1
FBP1
ICL2
ACS1
PDH1
REG2
CIT3
CRC1
RGI2
YLR312C
PUT4
NCA3
STL1
ALP1
NDE2
no promoter
0h
6h
24h
48h
72h
Mean fluorescence intensity
10
100
1000

FIG. 16

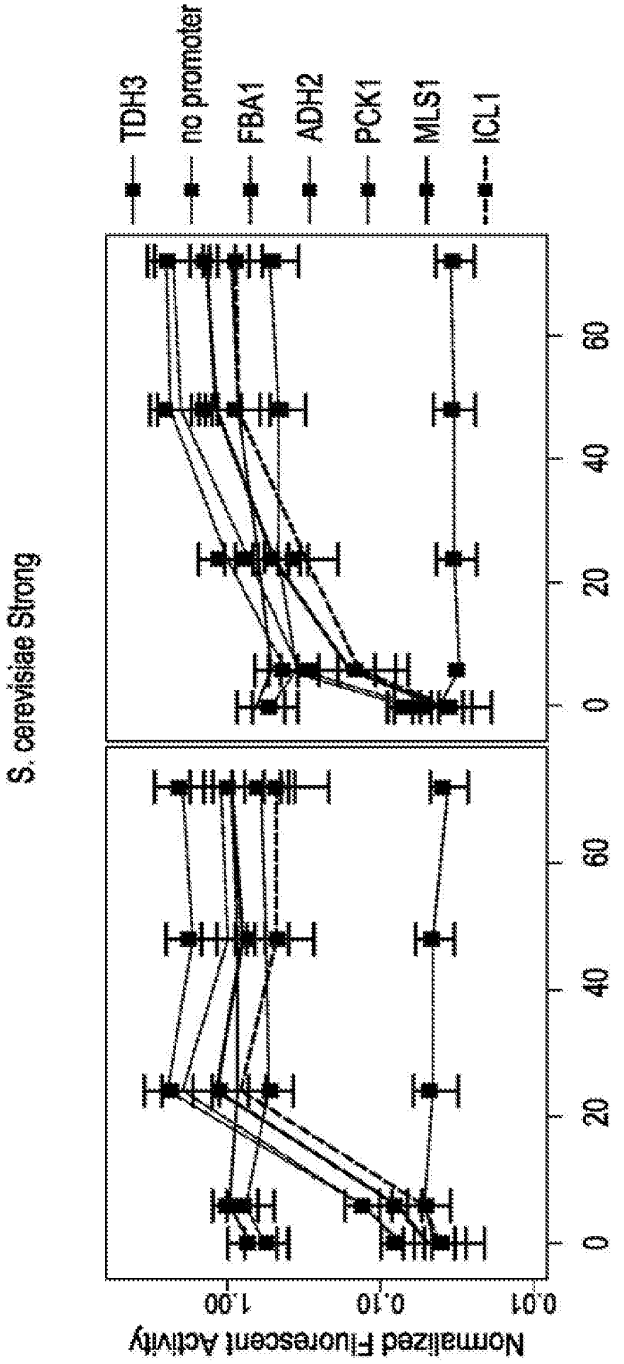


FIG. 17 (part 1/3)

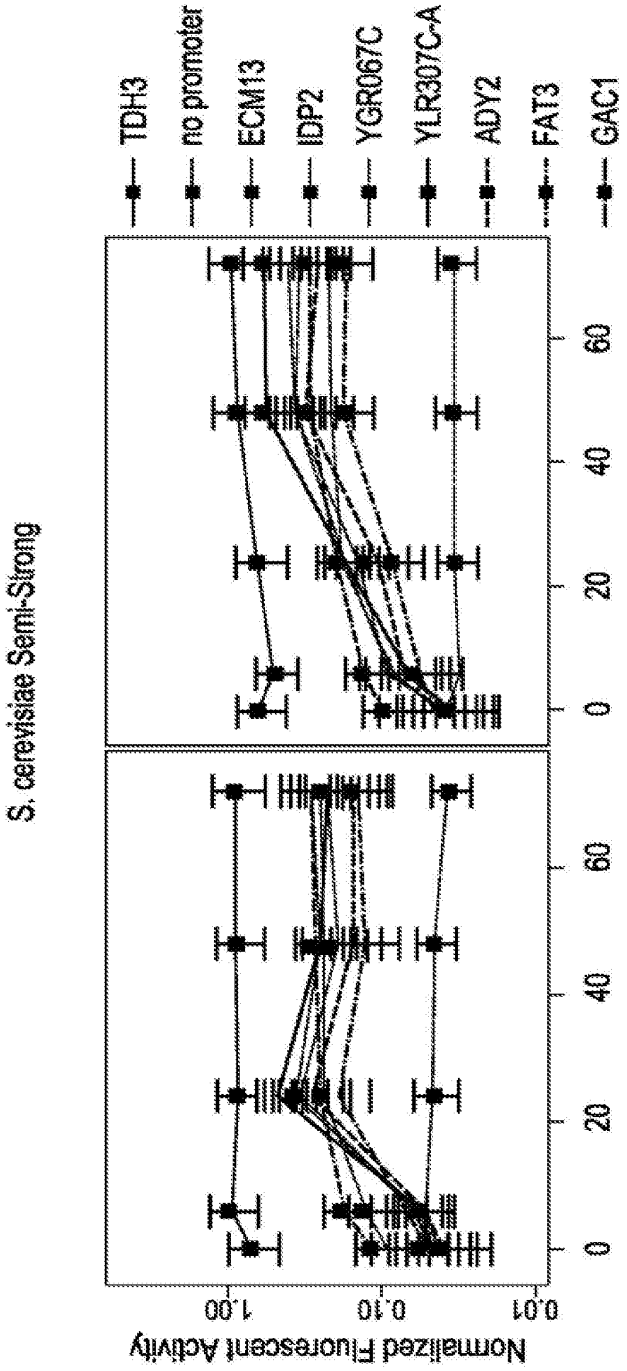


FIG. 17 (part 2/3)

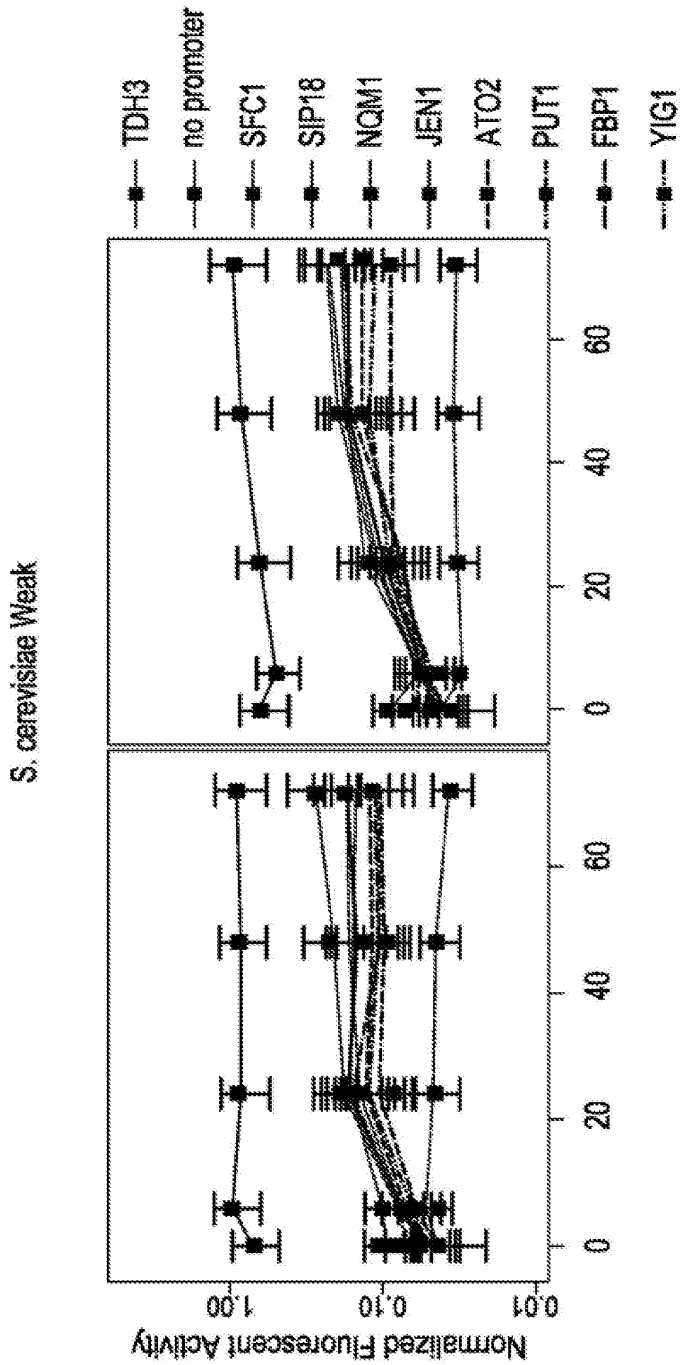


FIG. 17 (part 3/3)

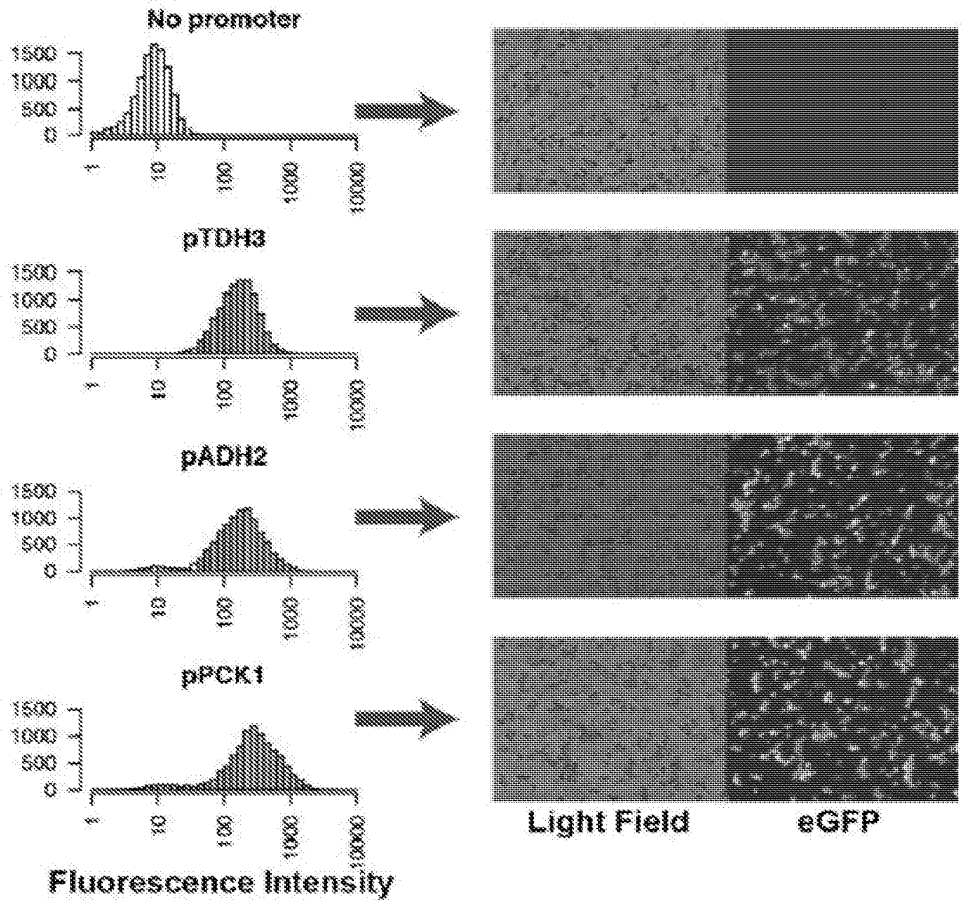

No promoter
pTDH3
pADH2
pPCK1
1500
1000
500
0
1
10
100
1000
10000
Fluorescence Intensity
Light Field
eGFP

FIG. 18

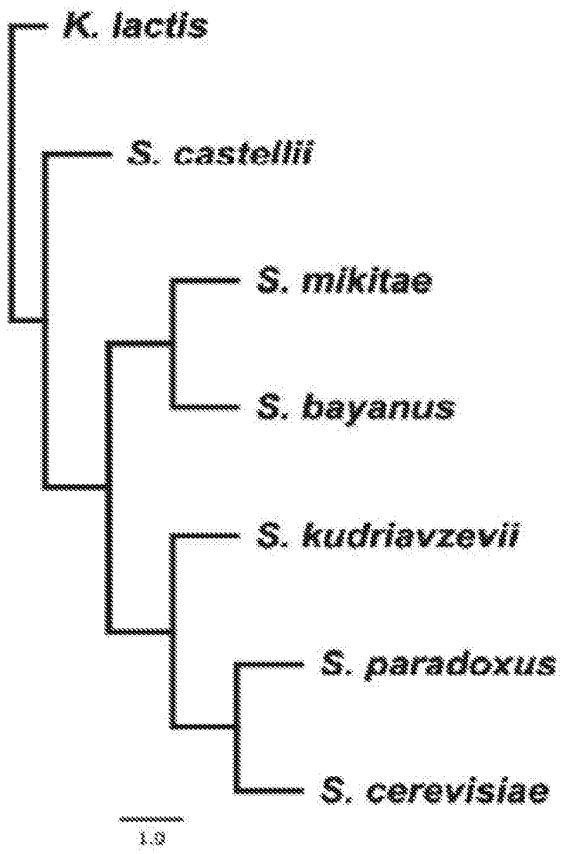

K. lactis
S. castellii
S. mikitae
S. bayanus
S. kudriavzevii
S. paradoxus
S. cerevisiae
1.0

FIG. 19

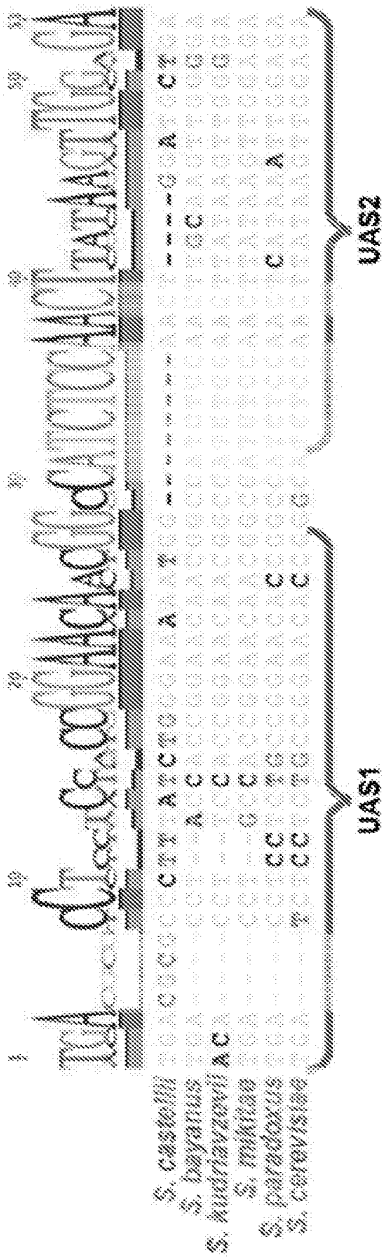


FIG. 20

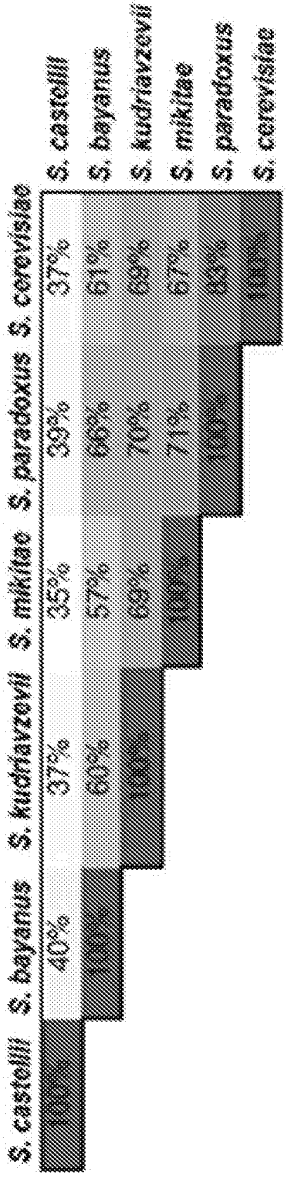

S. castellii
S. bayanus
S. kudriavzevii
S. mikitae
S. paradoxus
S. cerevisiae
100%
40%
37%
35%
39%
37%
100%
60%
57%
66%
61%
100%
63%
68%
70%
69%
100%
68%
71%
67%
100%
85%
100%
85%
S. castellii
S. bayanus
S. kudriavzevii
S. mikitae
S. paradoxus
S. cerevisiae

FIG. 21

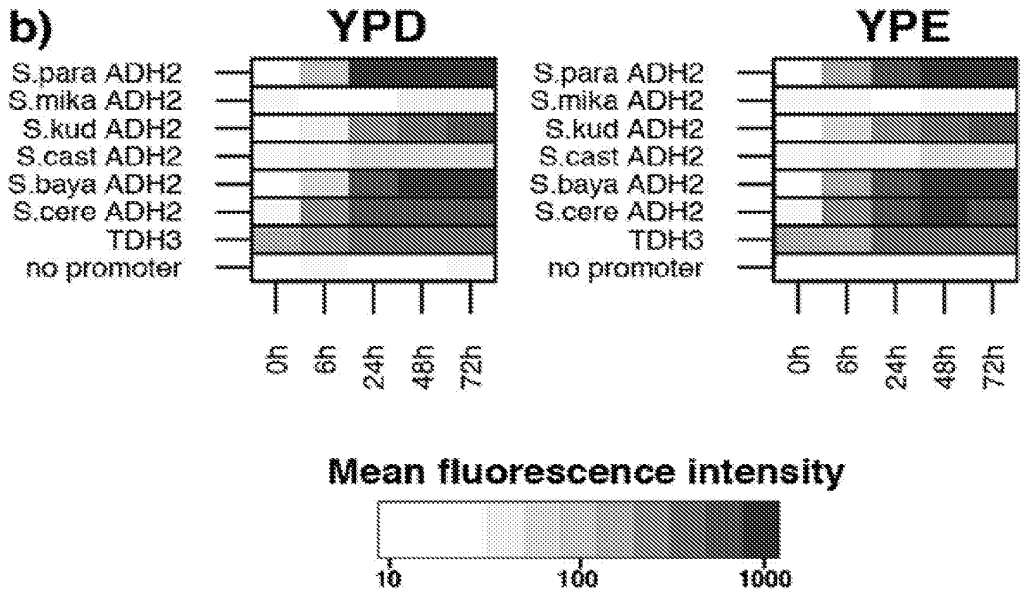

b)
YPD
YPE
S.para ADH2
S.mika ADH2
S.kud ADH2
S.cast ADH2
S.baya ADH2
S.cere ADH2
TDH3
no promoter
0h
6h
24h
48h
72h
Mean fluorescence intensity
10
100
1000

FIG. 22

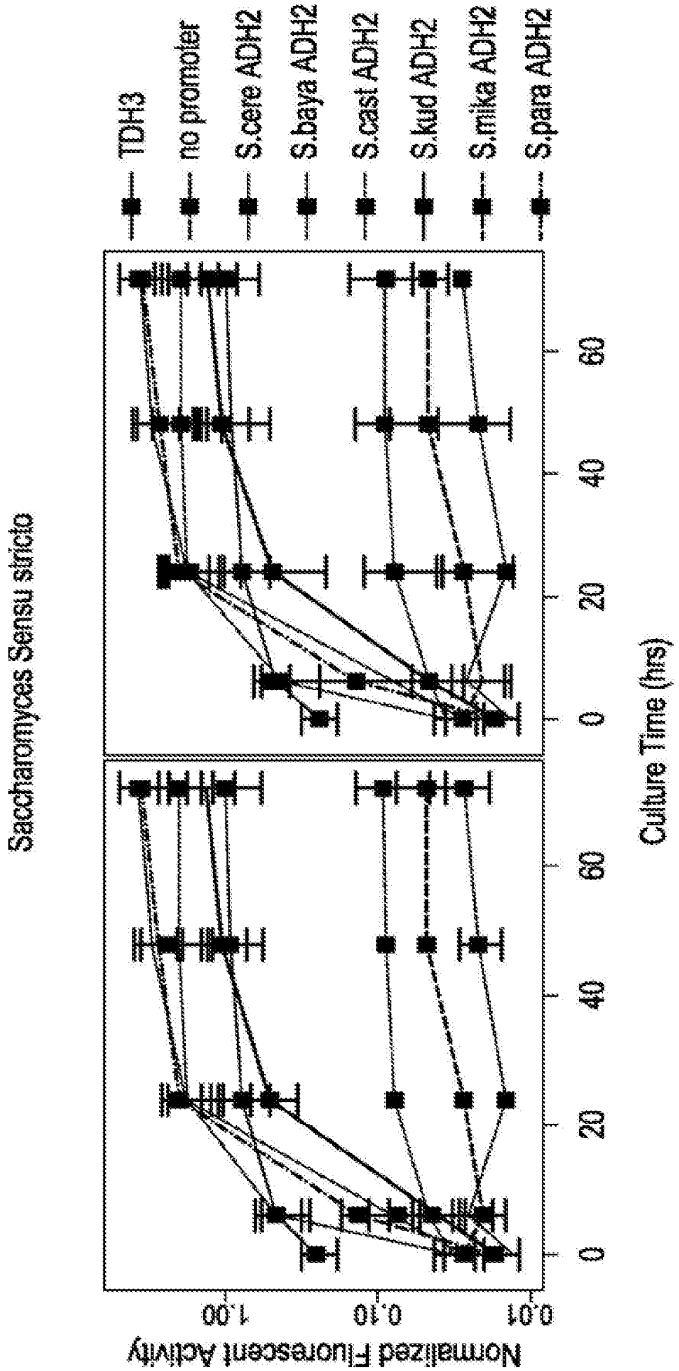

TDH3
no promoter
S.cere ADH2
S.baya ADH2
S.cast ADH2
S.kud ADH2
S.mika ADH2
S.para ADH2
Saccharomyces Sensu stricto
Normalized Fluorescent Activity
0.01
0.10
1.00
0
20
40
60
Culture Time (hrs)

FIG. 23

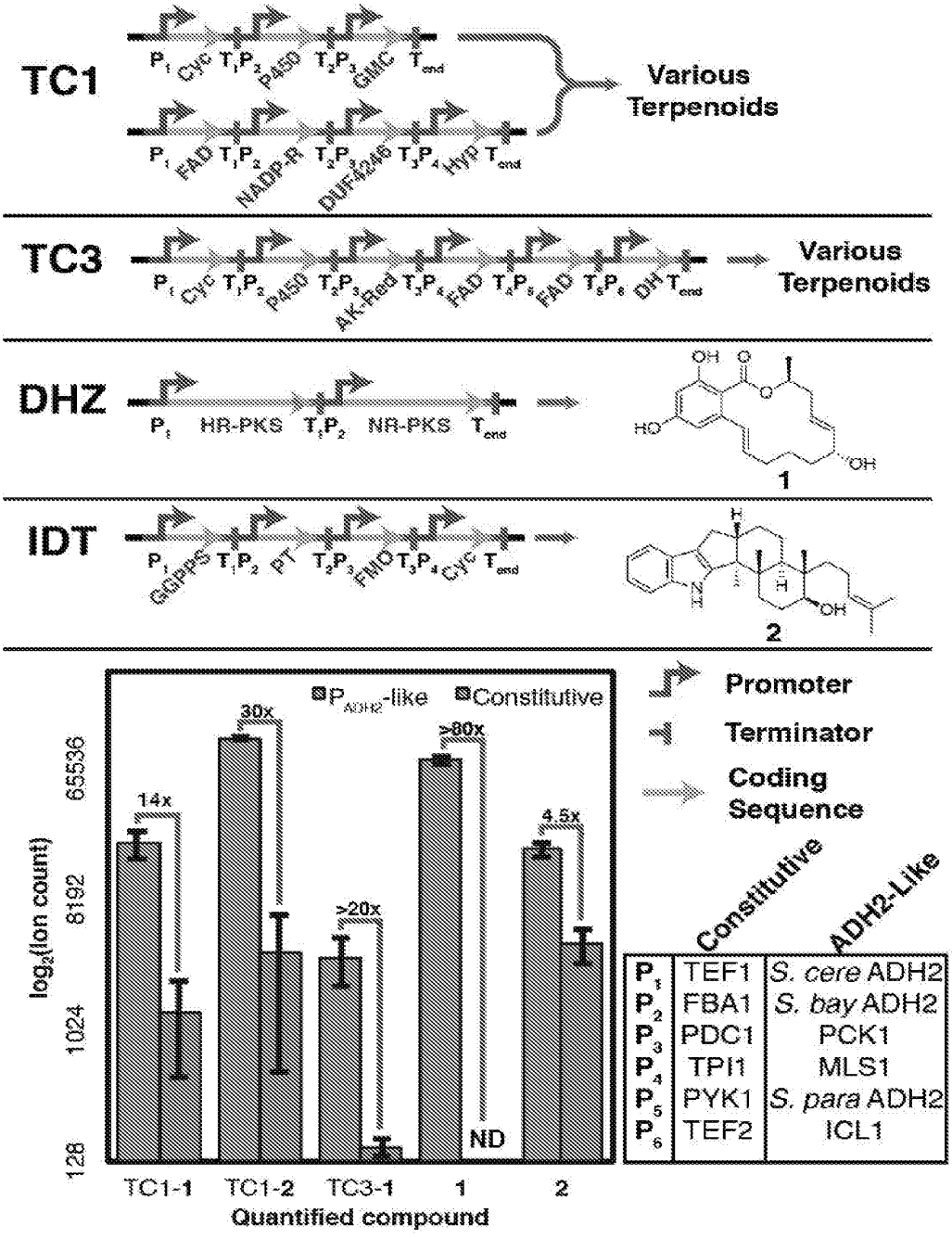

TC1
Various Terpenoids
TC3
Various Terpenoids
DHZ
1
IDT
2
Promoter
Terminator
Coding Sequence
Constitutive
ADH2-Like
P1 TEF1 S. cere ADH2
P2 FBA1 S. bay ADH2
P3 PDC1 PCK1
P4 TPI1 MLS1
P5 PYK1 S. para ADH2
P6 TEF2 ICL1
log2(Ion count)
65536
8192
1024
128
14x
30x
>20x
>80x
4.5x
ND
TC1-1
TC1-2
TC3-1
1
2
Quantified compound

FIG. 24

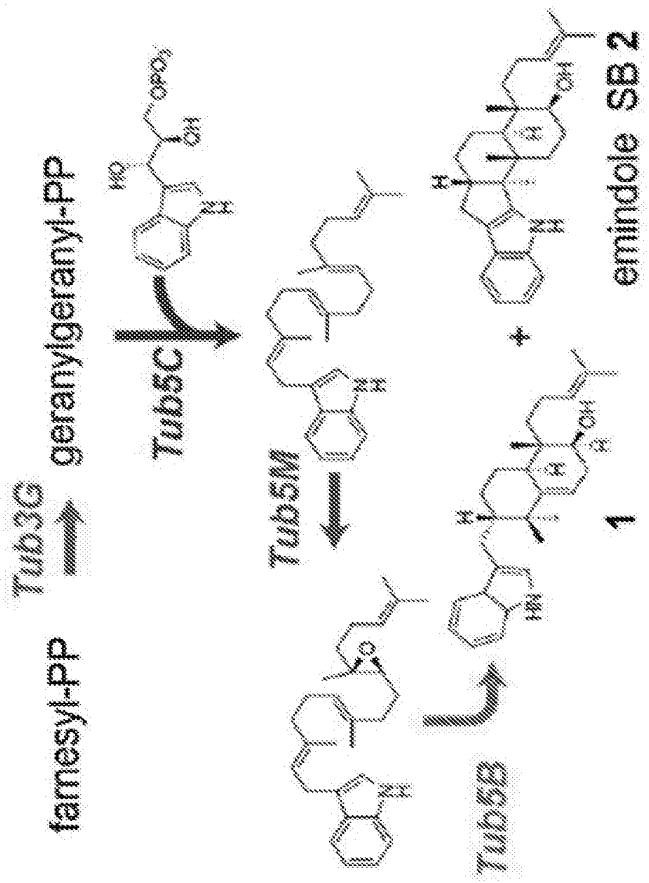

Tub3G
geranylgeranyl-PP
farnesyl-PP
Tub5C
Tub5M
Tub5B
1
+
emindole SB 2

FIG. 25

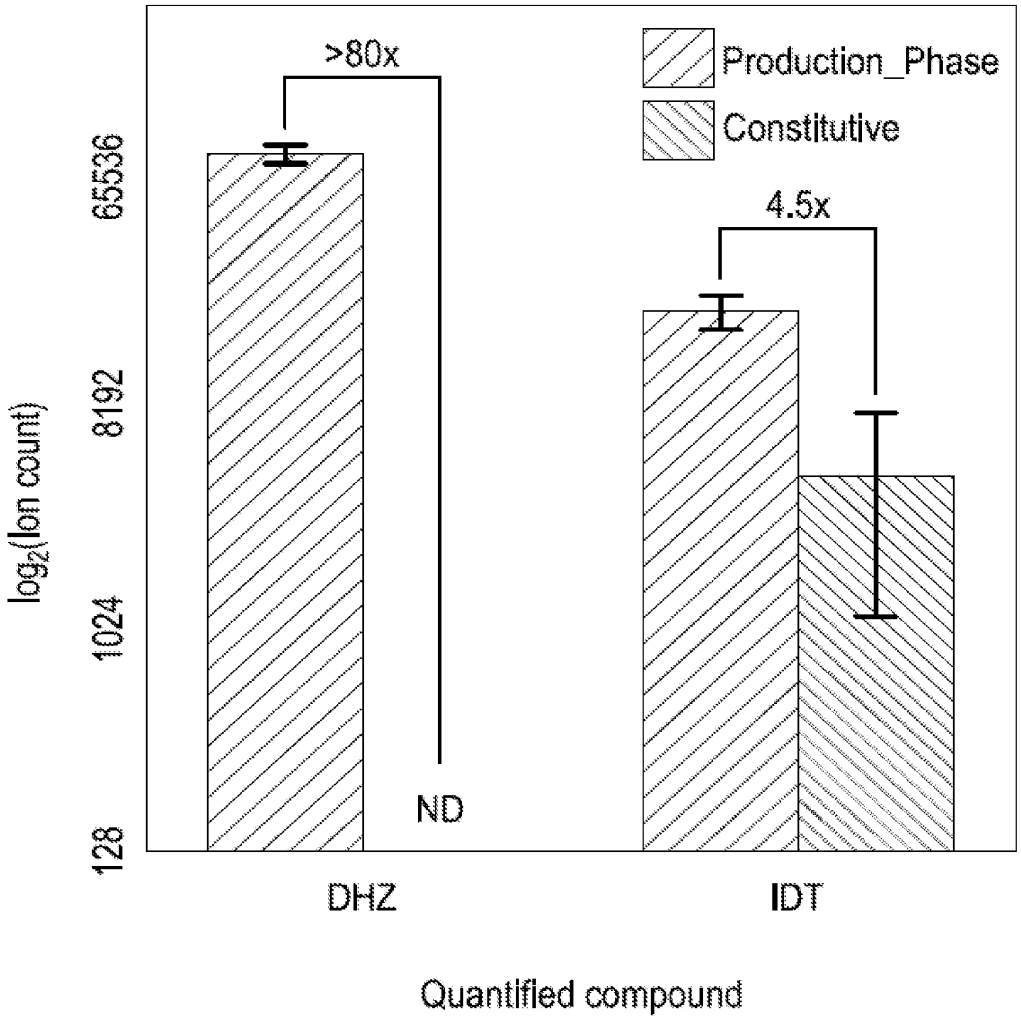

>80x
Production_Phase
Constitutive
4.5x
65536
8192
1024
128
log2(Ion count)
ND
DHZ
IDT
Quantified compound

FIG. 26

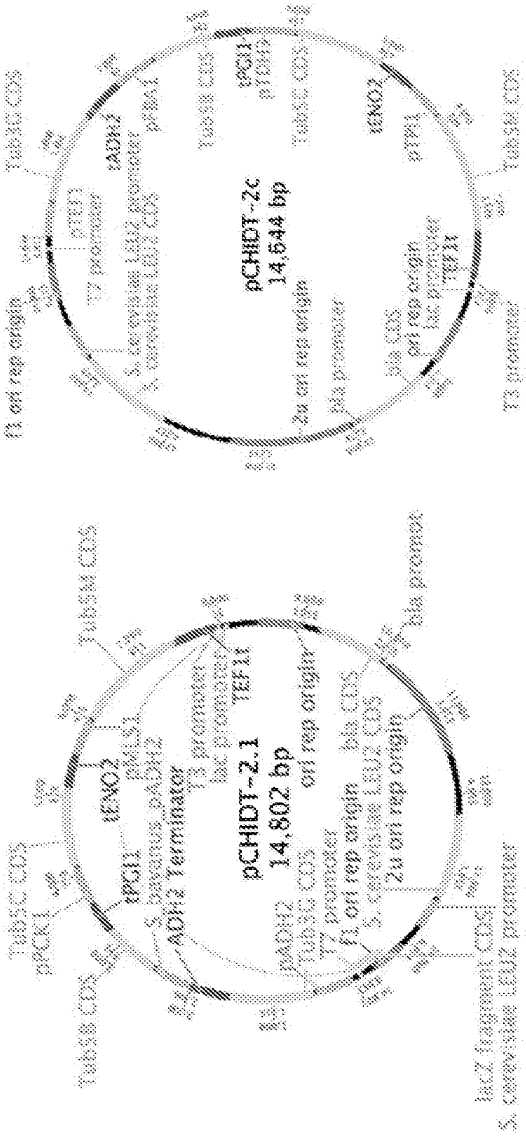


FIG. 27A

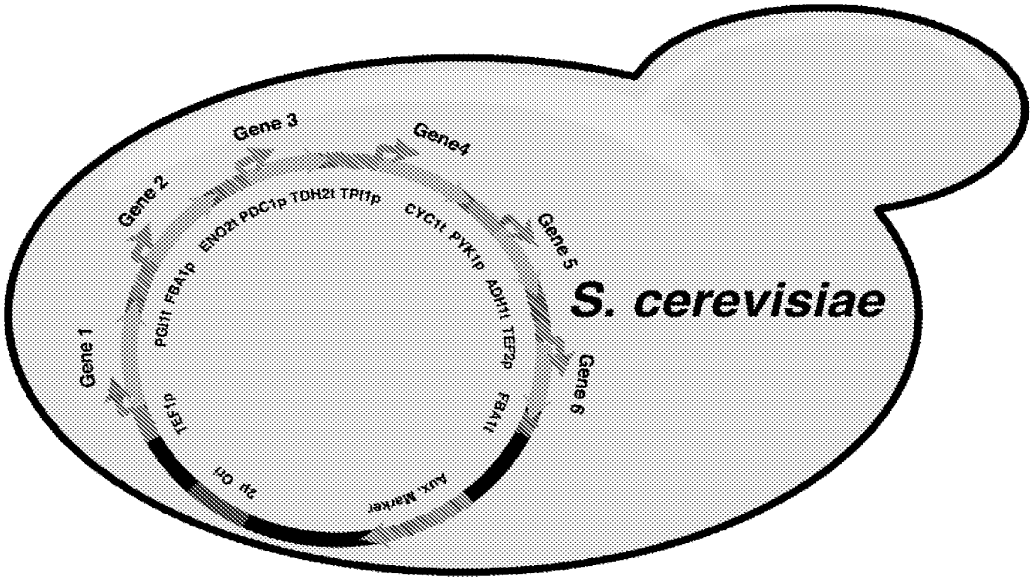

Gene 3
Gene4
Gene 2
Gene 5
S. cerevisiae
Gene 1
Gene 6
ENO2t PDC1p TDH2t TPI1p
CYC1t PYK1p
PGI1t FBA1p
ADH1 TEF2p
FBA1t
TEF1p
Aux. Marker

FIG. 27B

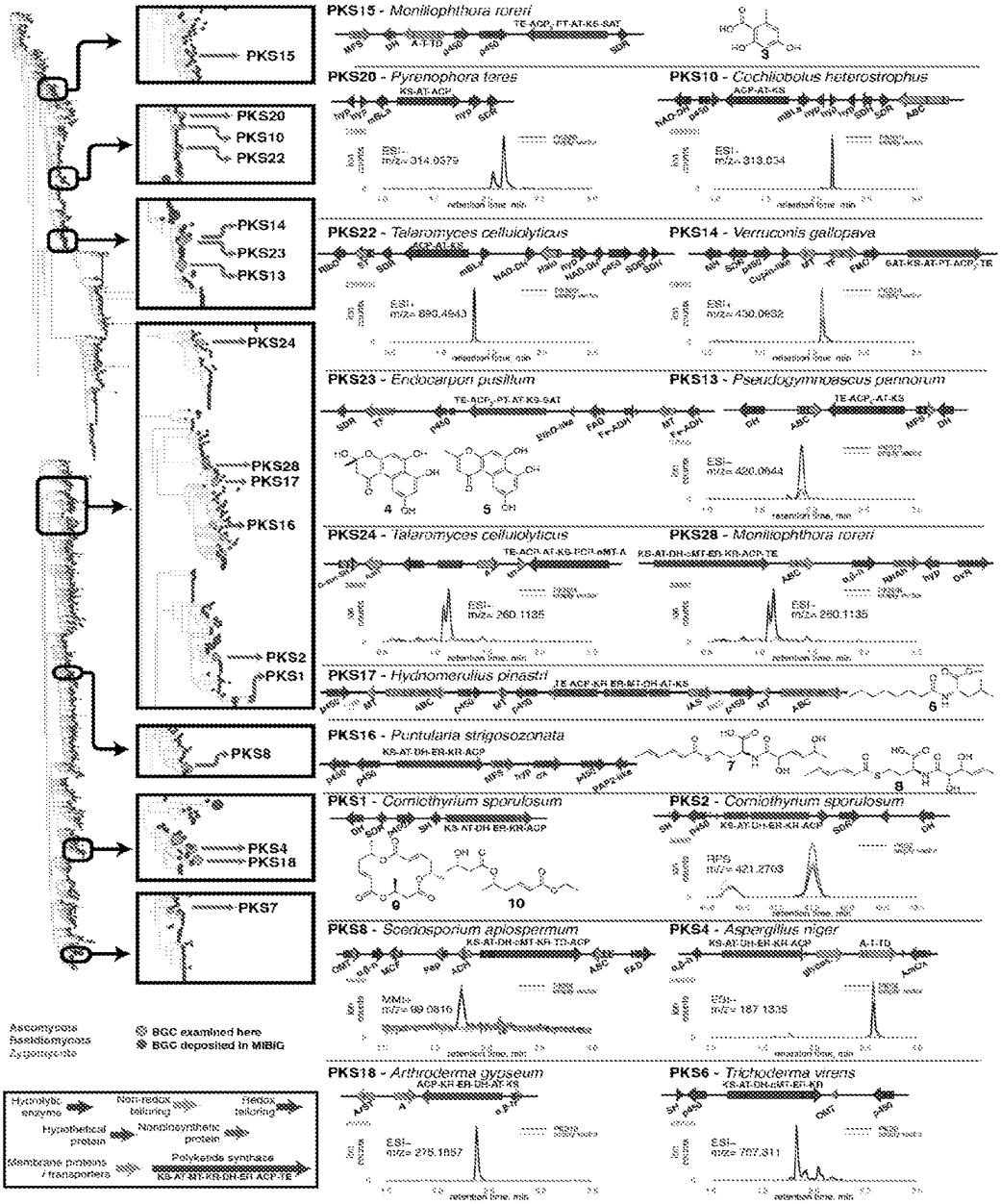

PKS15
PKS20
PKS10
PKS22
PKS14
PKS23
PKS13
PKS24
PKS28
PKS17
PKS16
PKS2
PKS1
PKS8
PKS4
PKS18
PKS7
PKS15 - Moniliophthora roreri
PKS20 - Pyrenophora tares
PKS10 - Cochliobolus heterostrophus
PKS22 - Talaromyces cellulolyticus
PKS14 - Verruconis gallopava
PKS23 - Endocarpon pusillum
PKS13 - Pseudogymnoascus pannorum
PKS24 - Talaromyces cellulolyticus
PKS28 - Moniliophthora roreri
PKS17 - Hydnomerulius pinastri
PKS16 - Puntularia strigosozonata
PKS1 - Corniothyrium sporulosum
PKS2 - Corniothyrium sporulosum
PKS8 - Scedosporium apiospermum
PKS4 - Aspergillus niger
PKS18 - Arthroderma gypseum
PKS6 - Trichoderma virens
Ascomycota
Basidiomycota
Zygomycota
BGC examined here
BGC deposited in MIBiG
Hydrolytic enzyme
Non-redox tailoring
Redox tailoring
Hypothetical protein
Nonbiosynthetic protein
Membrane proteins / transporters
Polyketide synthase
KS-AT-MT-KR-DH-ER-ACP-TE

FIG. 28A

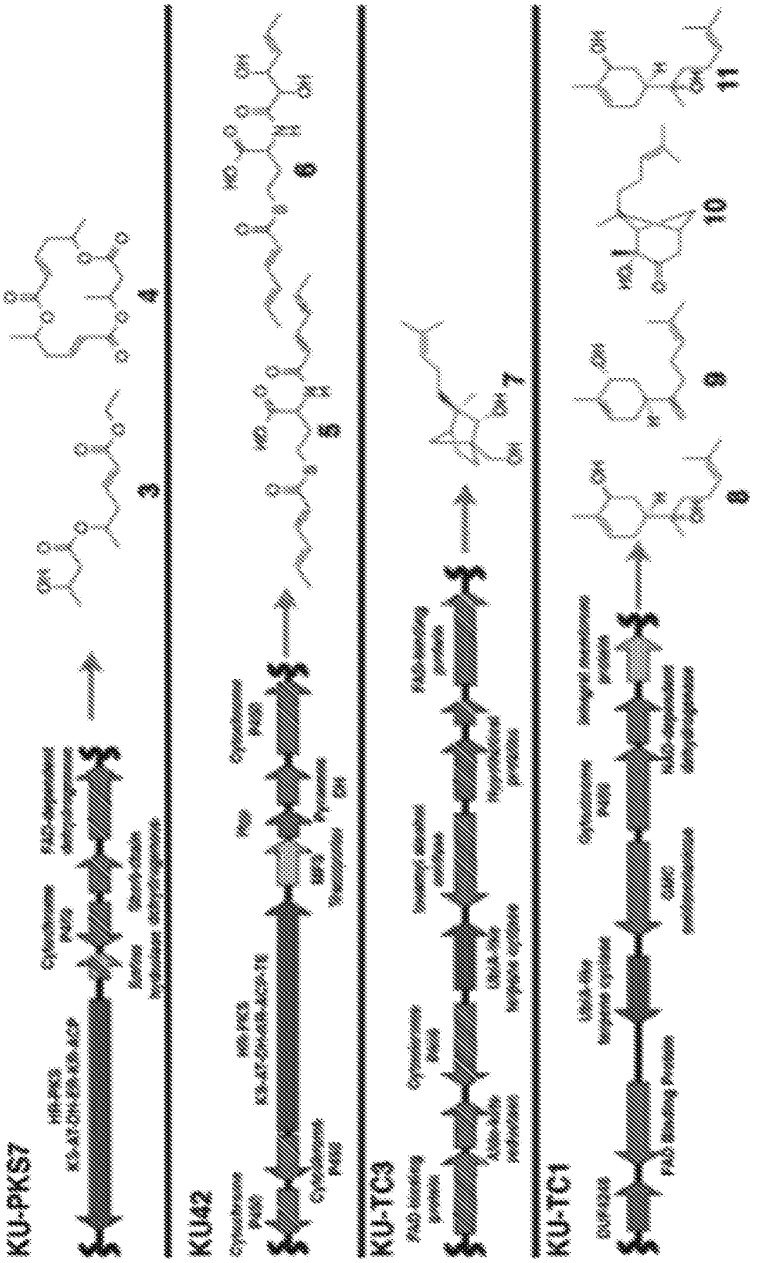


FIG. 28B

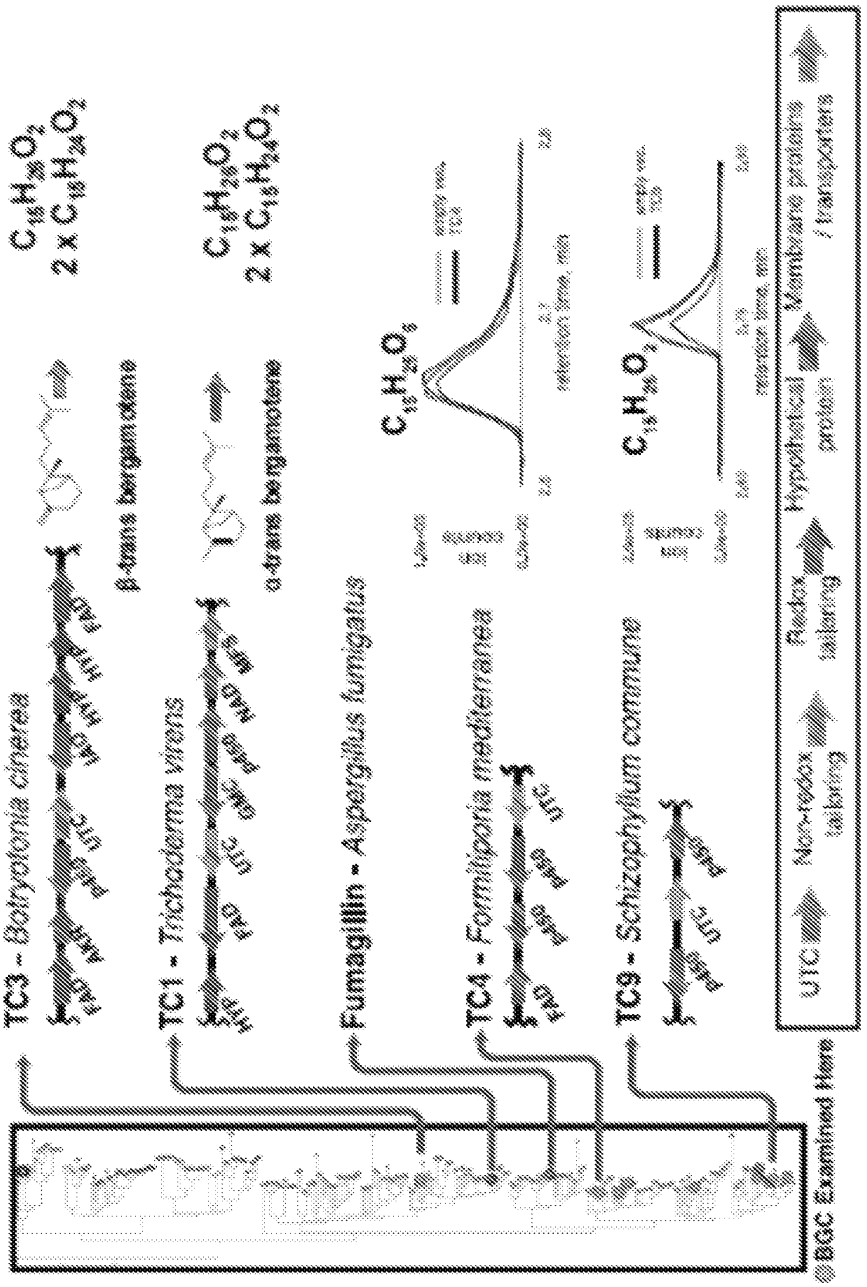

TC3 - Botryotinia cinerea
TC1 - Trichoderma virens
Fumagillin - Aspergillus fumigatus
TC4 - Formitiporia mediterranea
TC9 - Schizophyllum commune
BGC Examined Here
β-trans bergamotene
α-trans bergamotene
UTC
Non-redox tailoring
Redox tailoring
Hypothetical protein
Membrane protein / transporters

FIG. 28C

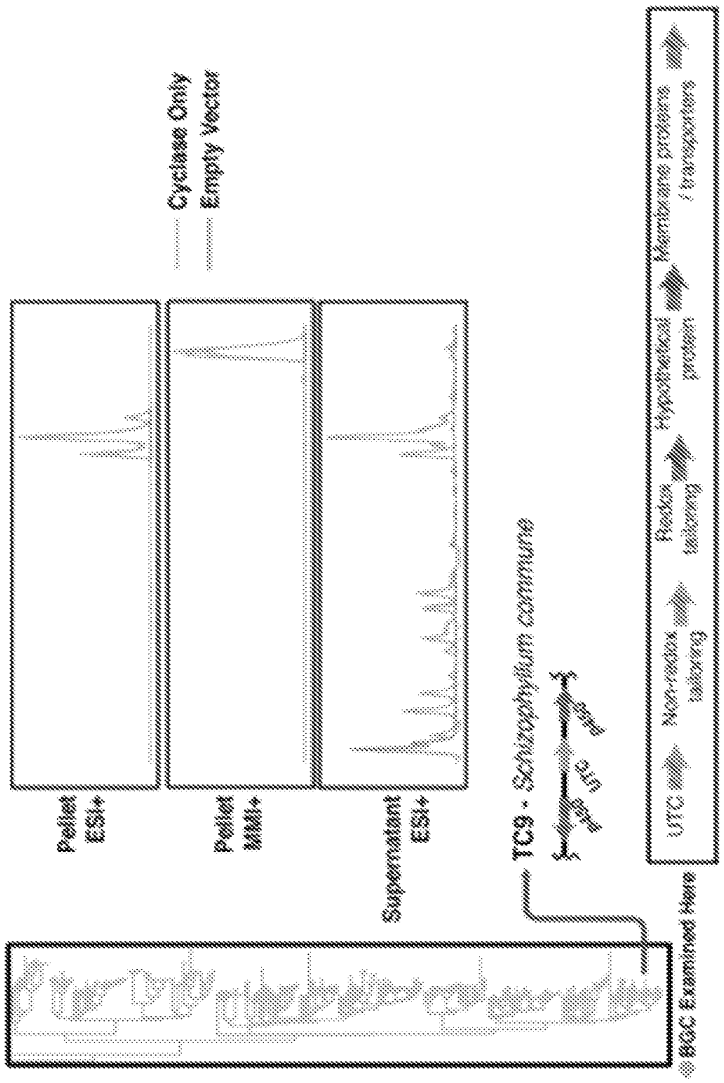

Cyclase Only
Empty Vector
Pellet ESI+
Pellet MMI+
Supernatant ESI+
TC9 - Schizophyllum commune
UTC
Non-redox tailoring
Redox tailoring
Hypothetical protein
Membrane protein
Substrate / Antibiotic
© BGC Examined Here

FIG. 28D

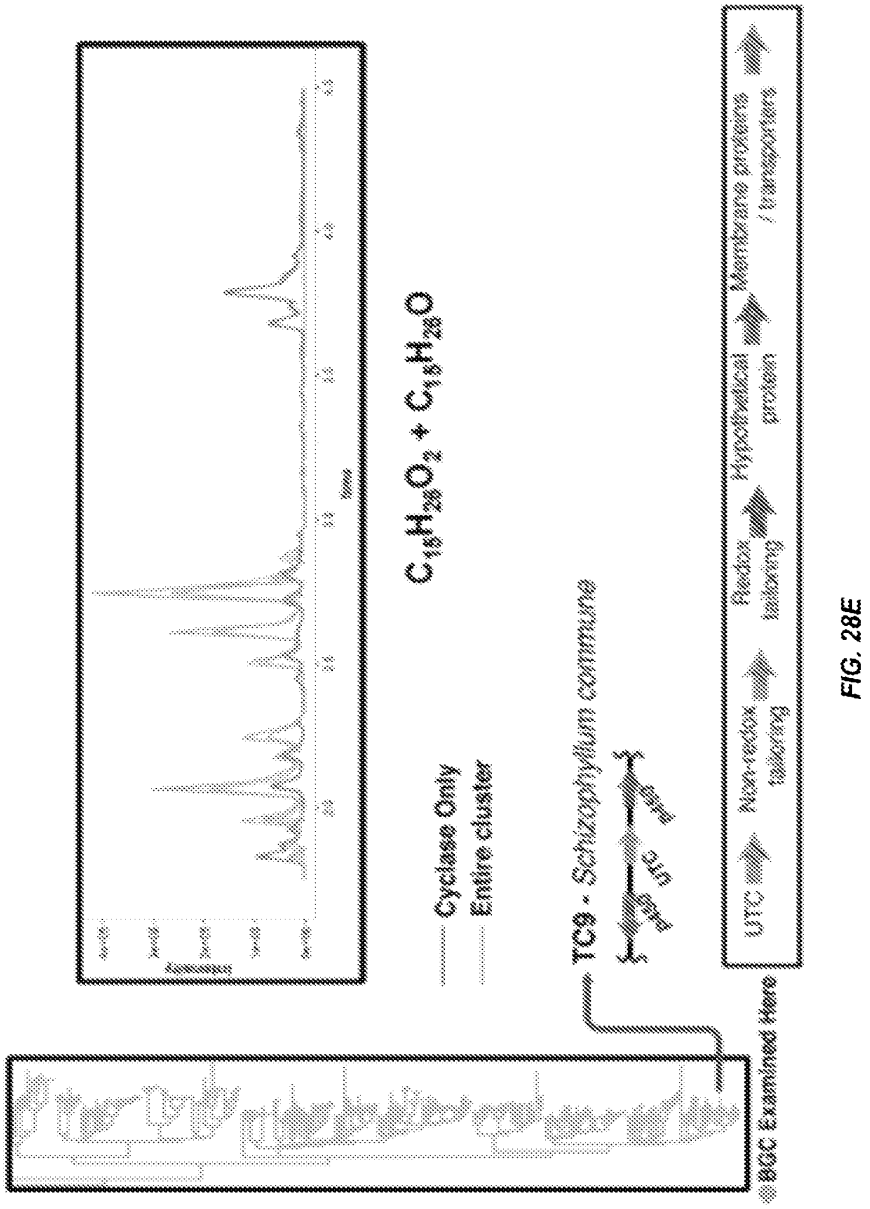

BGC Examined Here
TC9 - Schizophyllum commune
Cyclase Only
Entire cluster
$C_{16}H_{26}O_2 + C_{18}H_{28}O$
UTC
Non-redox tailoring
Redox tailoring
Hypothetical protein
Membrane proteins / transporters

FIG. 28E

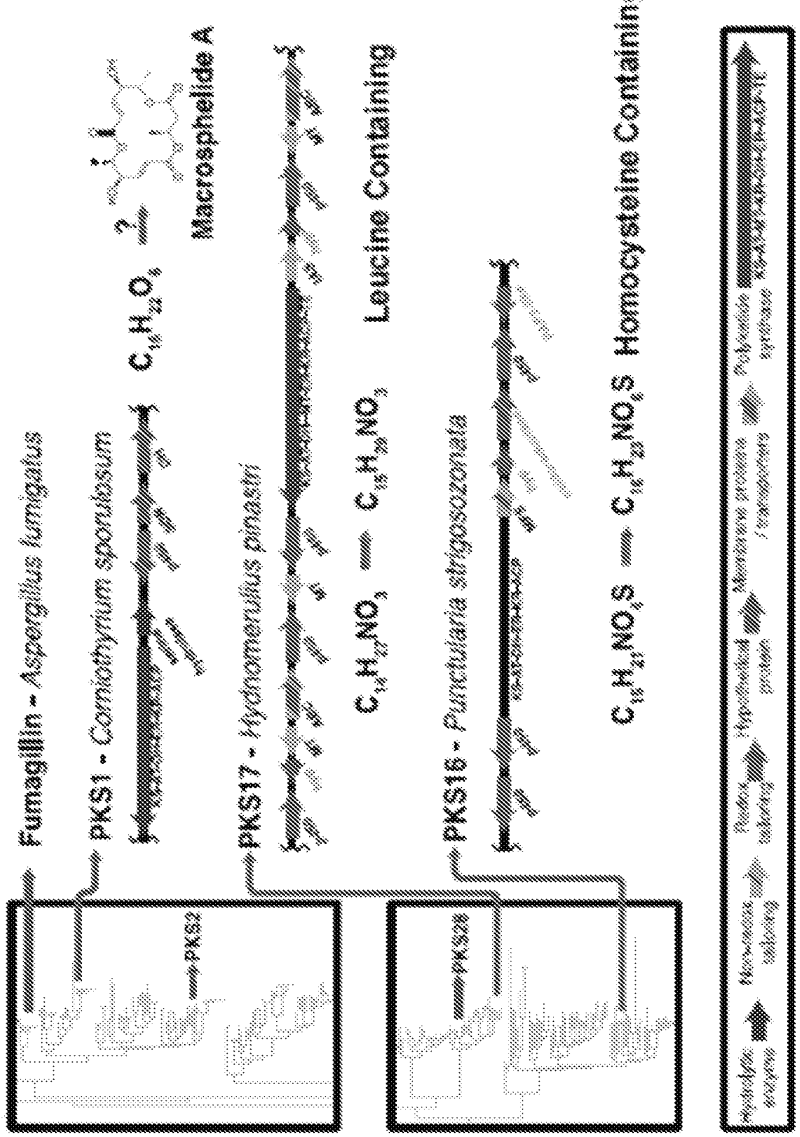

Fumagillin - Aspergillus fumigatus
PKS1 - Coniothyrium sporulosum
PKS2
$C_{15}H_{22}O_5$
Macrosphelide A
PKS17 - Hydnomerulius pinastri
$C_{13}H_{21}NO_3$ — $C_{18}H_{29}NO_3$ Leucine Containing
PKS16 - Punctularia strigosozonata
PKS20
$C_{18}H_{21}NO_4S$ — $C_{18}H_{23}NO_4S$ Homocysteine Containing
Polyketide synthase
Non-reducing polyketide synthase
Redox tailoring
Hypothetical protein
Membrane protein
Transporter
Polyketide synthase
Acyltransferase

FIG. 28F

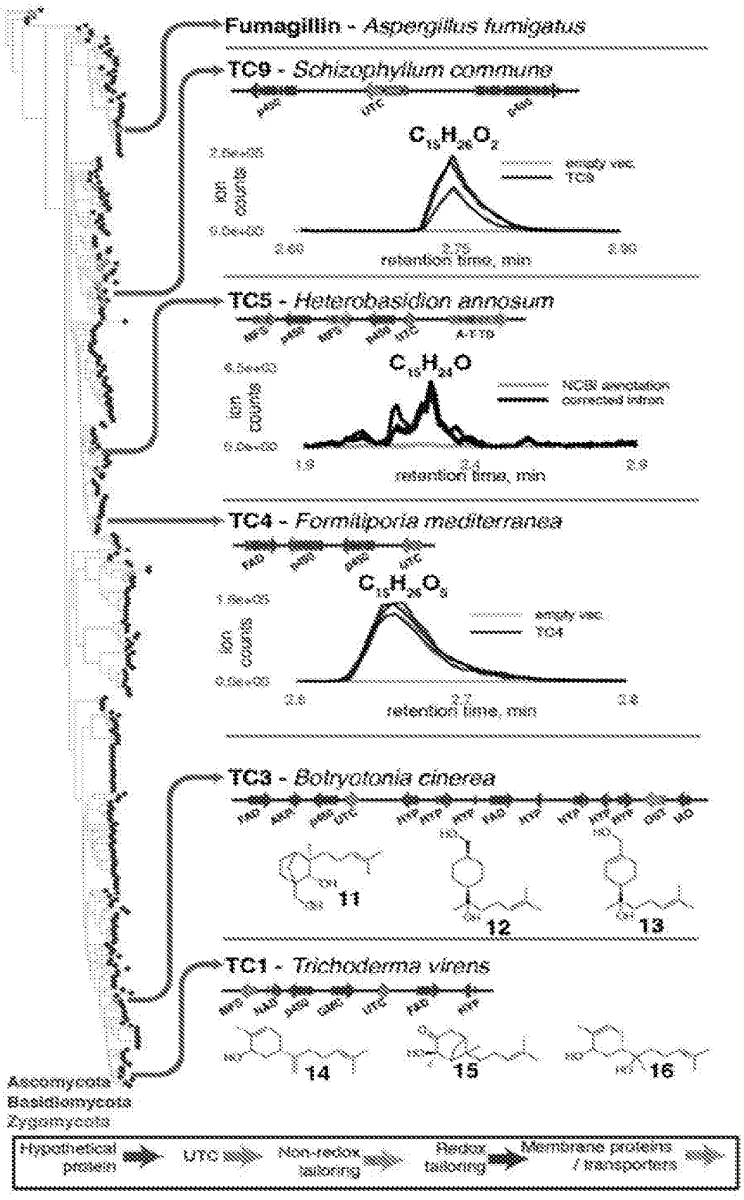

Fumagillin - Aspergillus fumigatus
TC9 - Schizophyllum commune
TC5 - Heterobasidion annosum
TC4 - Formitiporia mediterranea
TC3 - Botryotinia cinerea
TC1 - Trichoderma virens
Ascomycota
Basidiomycota
Zygomycota
Hypothetical protein
UTC
Non-redox tailoring
Redox tailoring
Membrane proteins / transporters

FIG. 29A

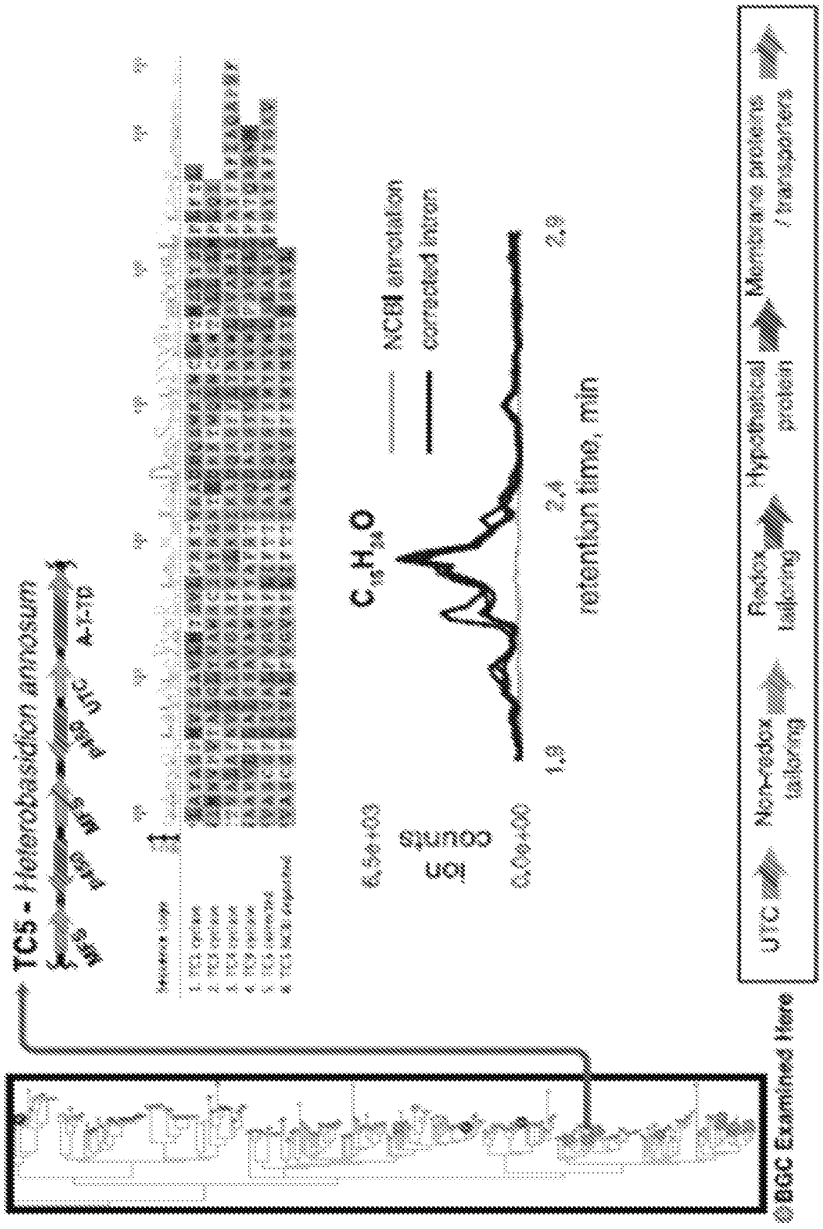

TC5 - Heterobasidion annosum
UTC
A-STD
1
$C_{10}H_{16}O$
NCBI annotation
corrected intron
ion counts
6.5e+03
0.0e+00
1.9
2.4
2.9
retention time, min
UTC
Non-redox tailoring
Redox tailoring
Hypothetical protein
Membrane protein / transporters
BGC Examined Here

FIG. 29B

PKS1-Coniothyrium sporulosum
PKS2-Coniothyrium sporulosum
PKS3-Acremonium Sp. KY4917
PKS4-Aspergillus niger
PKS5-Thielavia terrestris
PKS6-Trichoderma virens
PKS7-Pseudogymnoascus pannorum
PKS8-Scedosporium apiospermum
PKS9-Metarhizium anisopliae
PKS10-Cochliobolus heterostrophus
PKS11-Pseudogymnoascus pannorum
PKS12-Pseudogymnoascus pannorum
PKS13-Pseudogymnoascus pannorum
PKS14-Verruconis gallopava
PKS15-Moniliophthora roreri
PKS16-Punctularia strigosozonata
PKS17-Hydnomerulius pinastri
PKS18-Arthroderma gypseum
PKS19-Setosphaeria turcica
PKS20-Pyrenophora teres
PKS21-Cladophialophora yegresii
PKS22-Talaromyces cellulolyticus
PKS23-Endocarpon pusillum
PKS24-Talaromyces cellulolyticus
PKS25-Moniliophthora roreri
PKS26-Hypholoma sublateritium
PKS27-Ceriporiopsis subvermispora
PKS28-Moniliophthora roreri

FIG. 30A

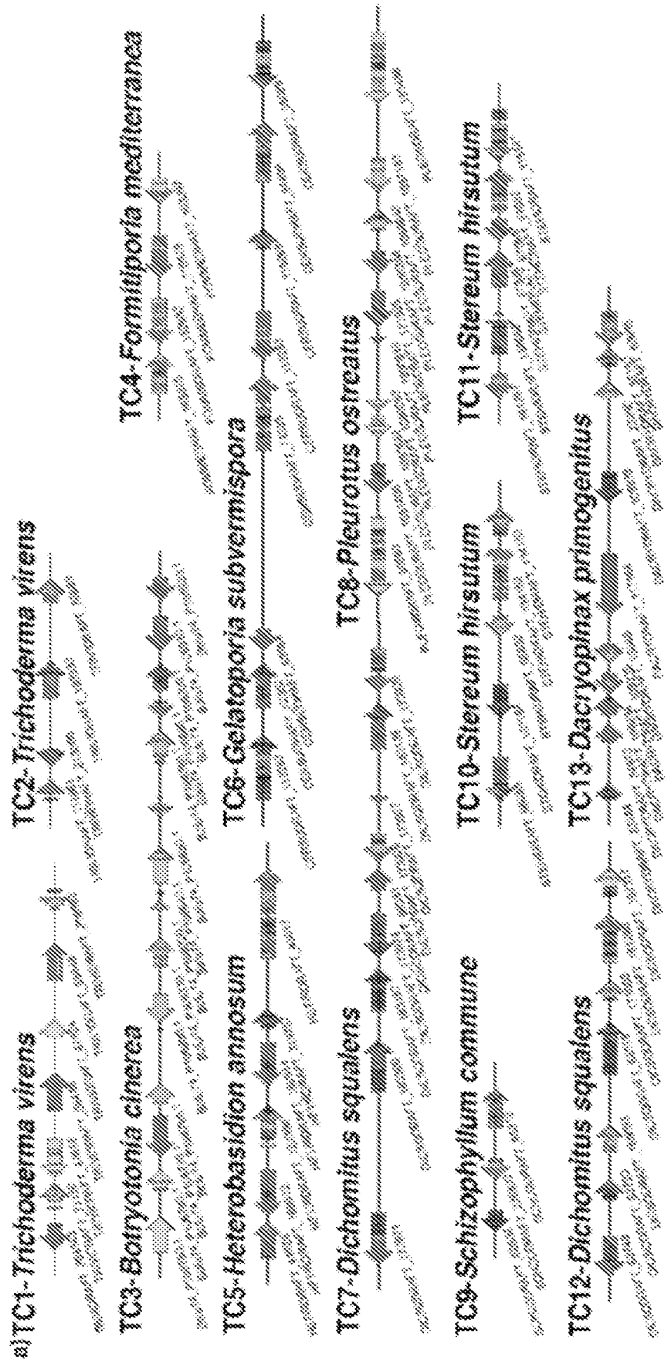

TC1-Trichoderma virens
TC2-Trichoderma virens
TC3-Botryotonia cinerea
TC4-Formitiporia mediterranea
TC5-Heterobasidion annosum
TC6-Gelatoporia subvermispora
TC7-Dichomitus squalens
TC8-Pleurotus ostreatus
TC9-Schizophyllum commune
TC10-Stereum hirsutum
TC11-Stereum hirsutum
TC12-Dichomitus squalens
TC13-Dacryopinax primogenitus

FIG. 30B

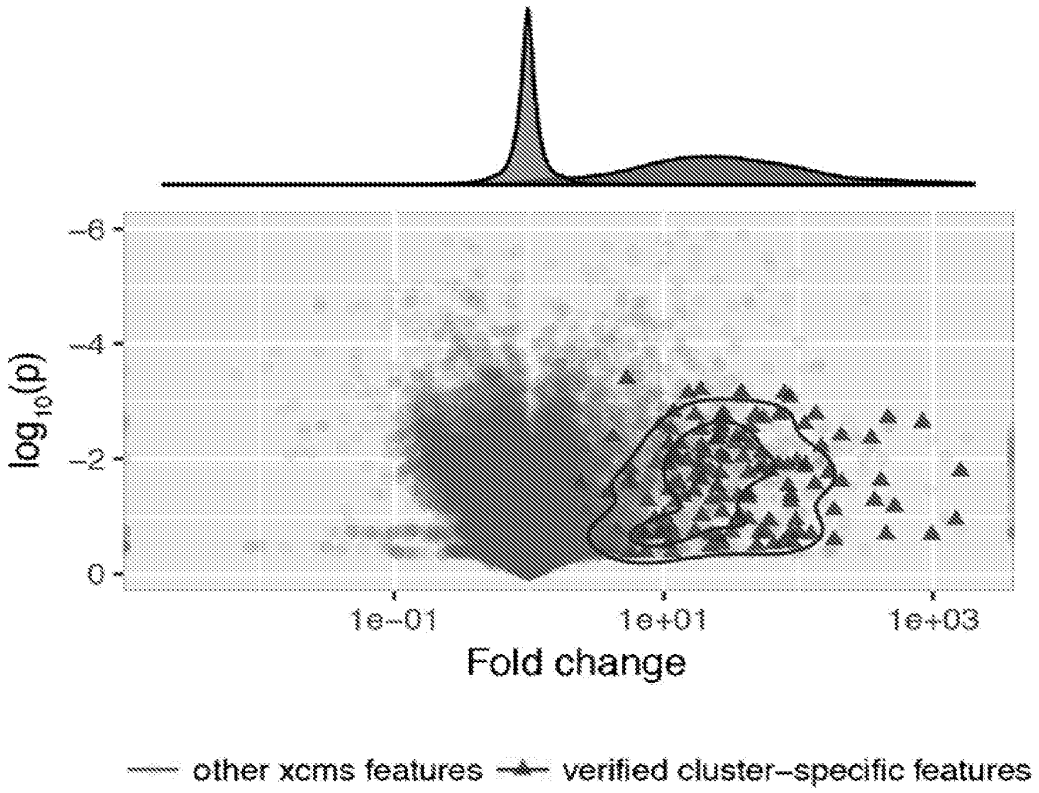

6
4
2
0
log10(p)
1e−01
1e+01
1e+03
Fold change
other xcms features
verified cluster−specific features

FIG. 31

LC-MS Data:.

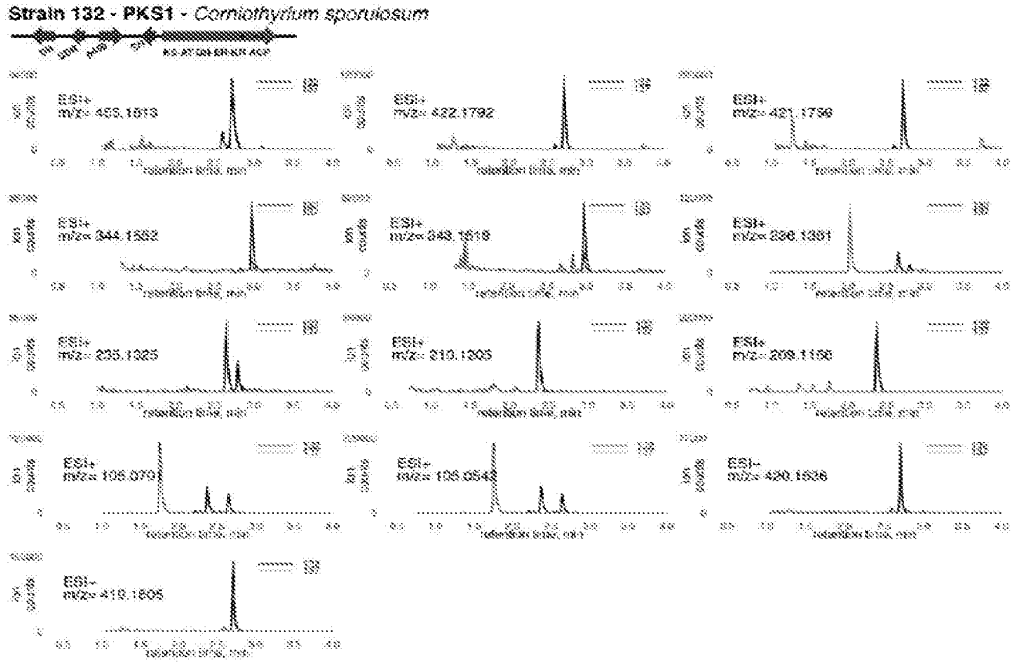


FIG. 32

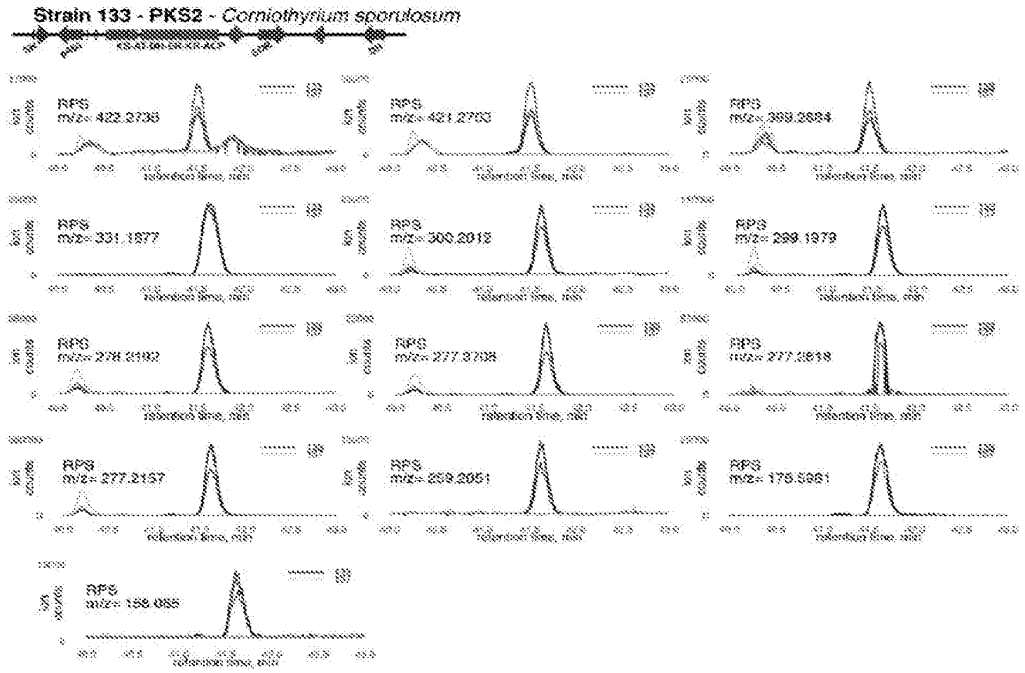

Strain 133 - PKS2 - Corniothyrium sporulosum
RPS
m/z= 423.2736
RPS
m/z= 421.2703
RPS
m/z= 399.2884
RPS
m/z= 331.1877
RPS
m/z= 300.2012
RPS
m/z= 299.1979
RPS
m/z= 278.0192
RPS
m/z= 277.9708
RPS
m/z= 277.2618
RPS
m/z= 277.2157
RPS
m/z= 259.2051
RPS
m/z= 178.5961
RPS
m/z= 158.055

FIG. 33

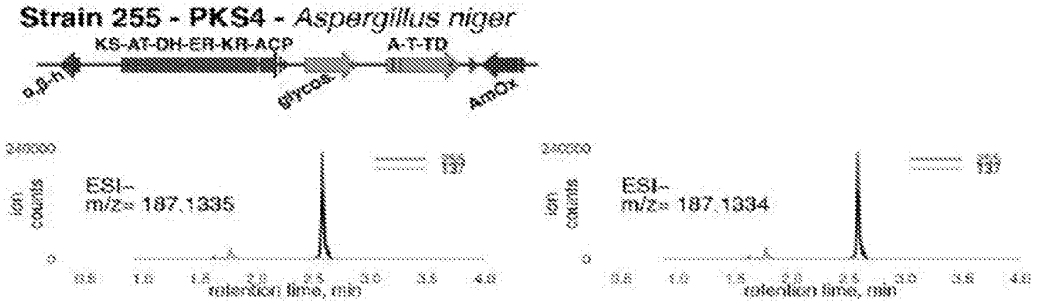

Strain 255 - PKS4 - Aspergillus niger
KS-AT-DH-ER-KR-ACP
A-T-TD
g.p.n
gPrcos
AmOx
ESI-
m/z= 187.1335
ESI-
m/z= 187.1334
retention time, min
retention time, min

FIG. 34

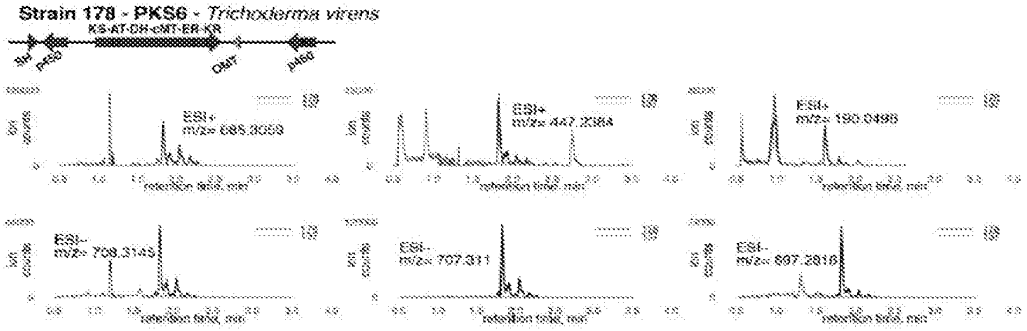

Strain 178 - PKS6 - Trichoderma virens
KS-AT-DH-cMT-ER-KR
SK
PKS6
cMT
PKS6
ESI+
m/z= 685.3059
ESI+
m/z= 447.2384
ESI+
m/z= 190.0498
ESI-
m/z= 709.3145
ESI-
m/z= 707.311
ESI-
m/z= 697.2816

FIG. 35

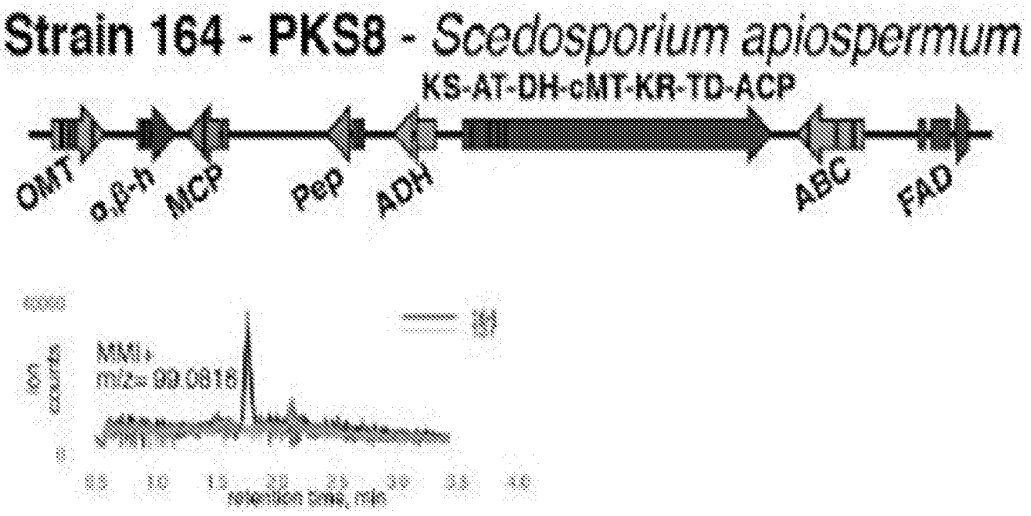

Strain 164 - PKS8 - Scedosporium apiospermum
KS-AT-DH-cMT-KR-TD-ACP
OMT
α,β-h
MCP
PcP
ADH
ABC
FAD
M1H+
m/z= 99.0818

FIG. 36

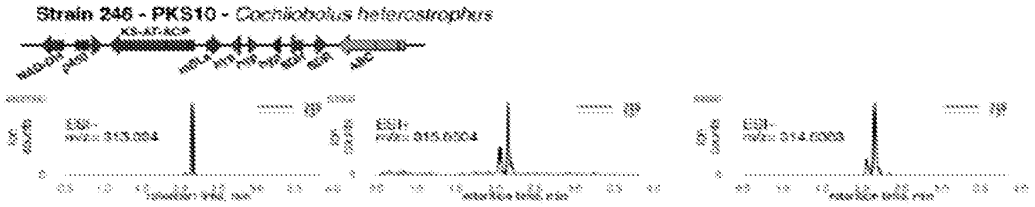


FIG. 37

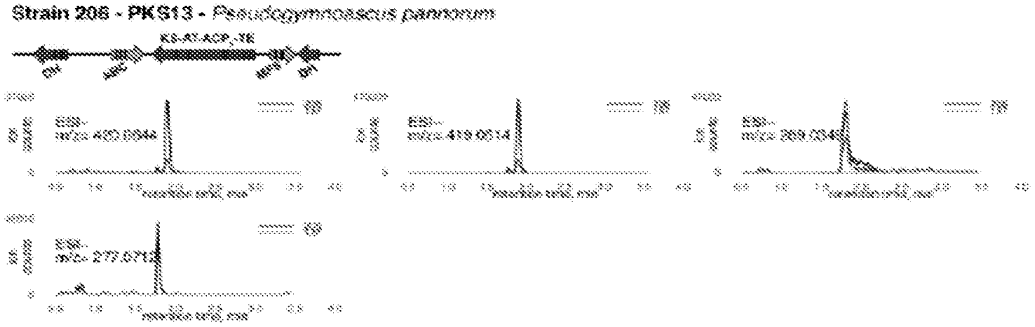


FIG. 38

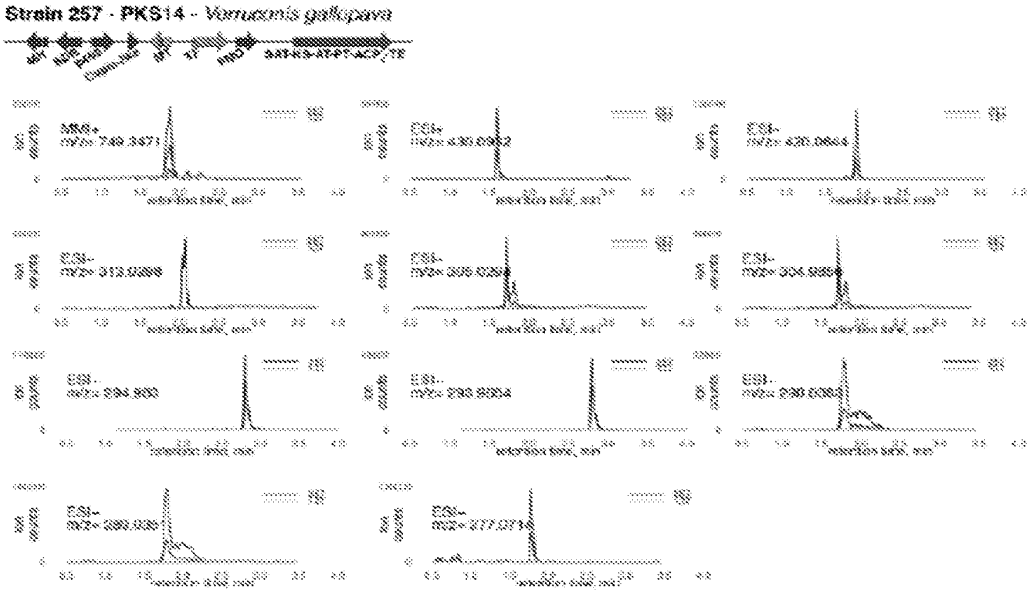


FIG. 39

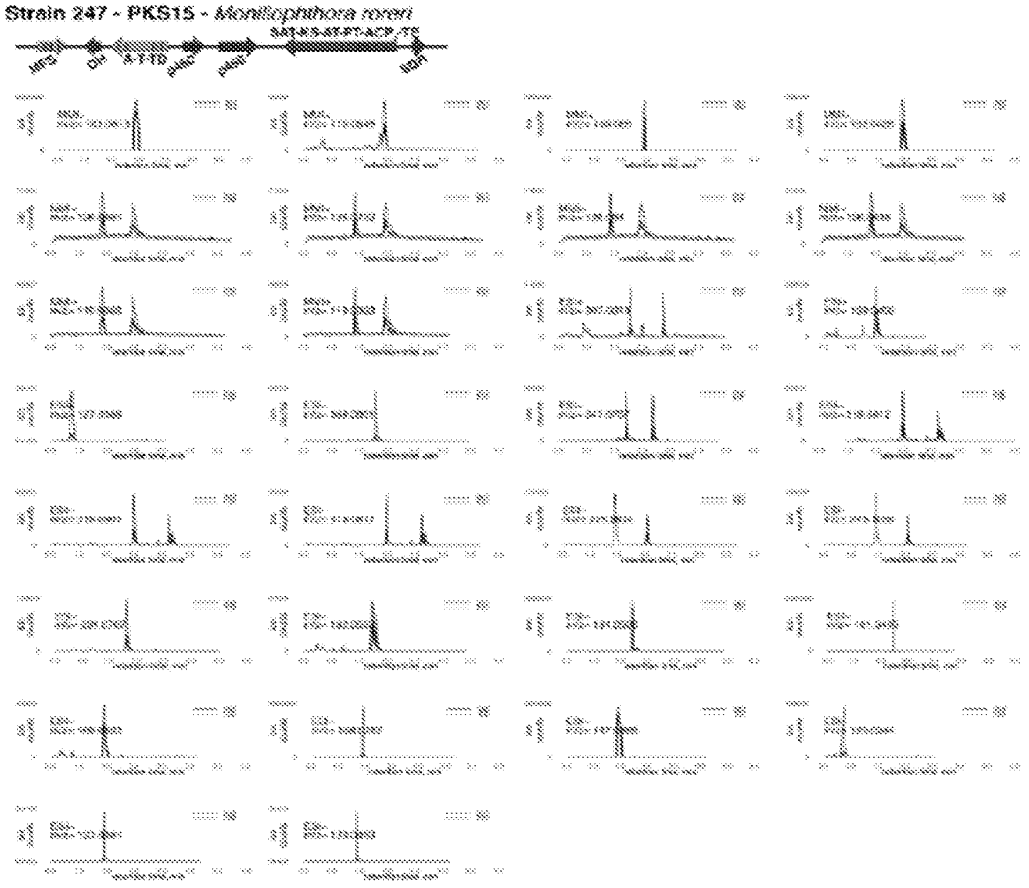

Strain 247 - PKS15 - Moniliophthora roreri
MFS1
DH
KR
PS1
PS2
TE/R

FIG. 40

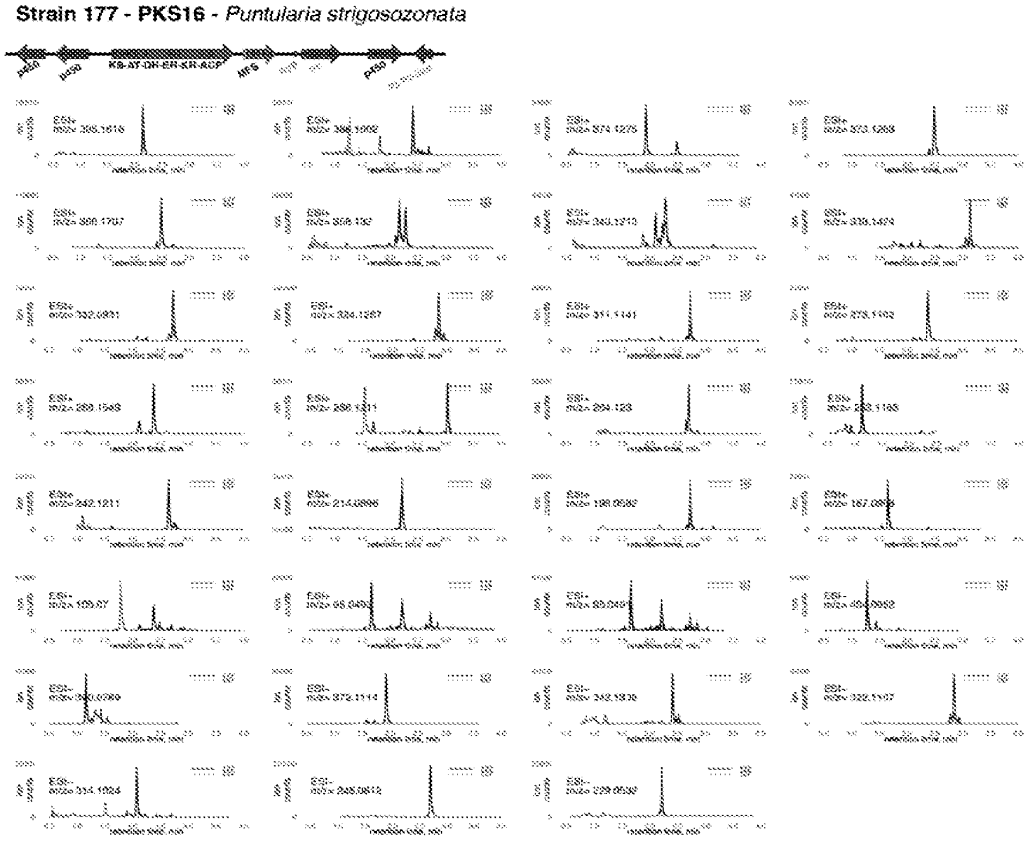


FIG. 41

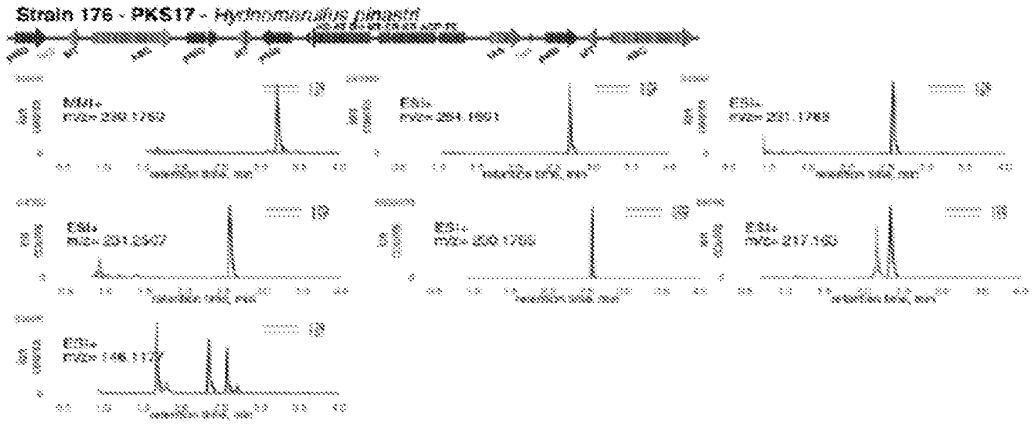


FIG. 42

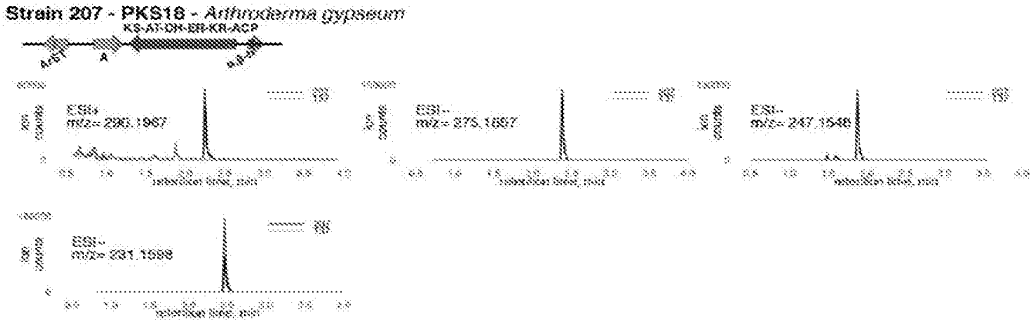

Strain 207 - PKS18 - Arthroderma gypseum
KS-AT-DH-ER-KR-ACP
ESI-
m/z= 231.1698

FIG. 43

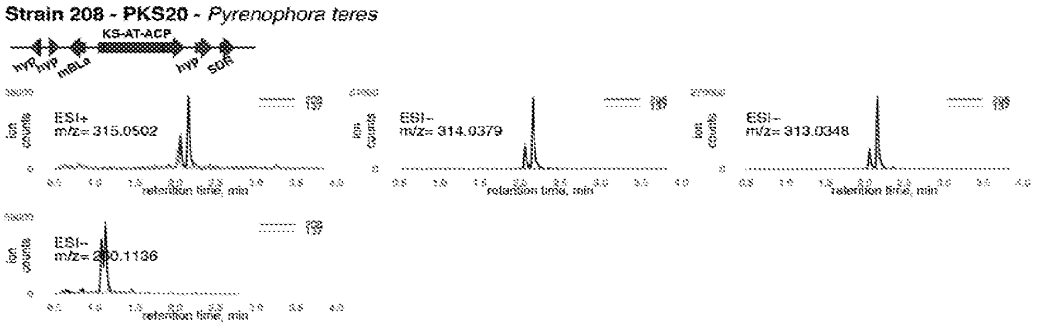

Strain 208 - PKS20 - Pyrenophora teres
KS-AT-ACP
hyp
hyp
mtLS
hyp
SDR
ESI+
m/z= 315.0502
ESI-
m/z= 314.0379
ESI-
m/z= 313.0348
ESI-
retention time, min

FIG. 44

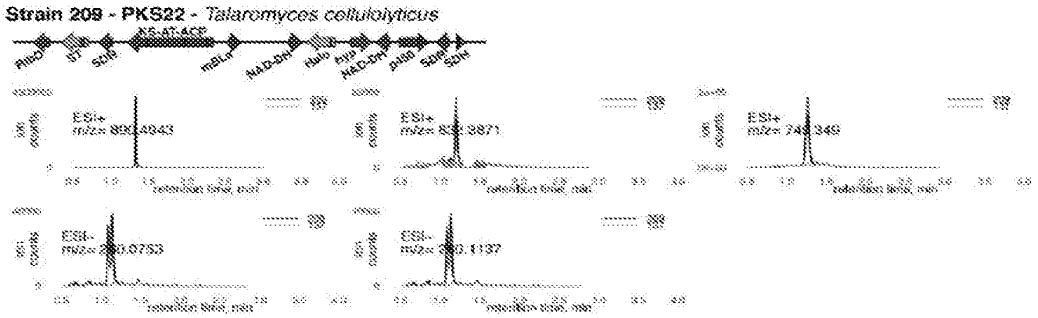

Strain 209 - PKS22 - Talaromyces cellulolyticus
KS-AT-ACP
mdlC
NAD-BH
dabD
FAD-BH
phiB
SDR
SDH
ESI+
ESI+
ESI+
ESI-
ESI-

FIG. 45

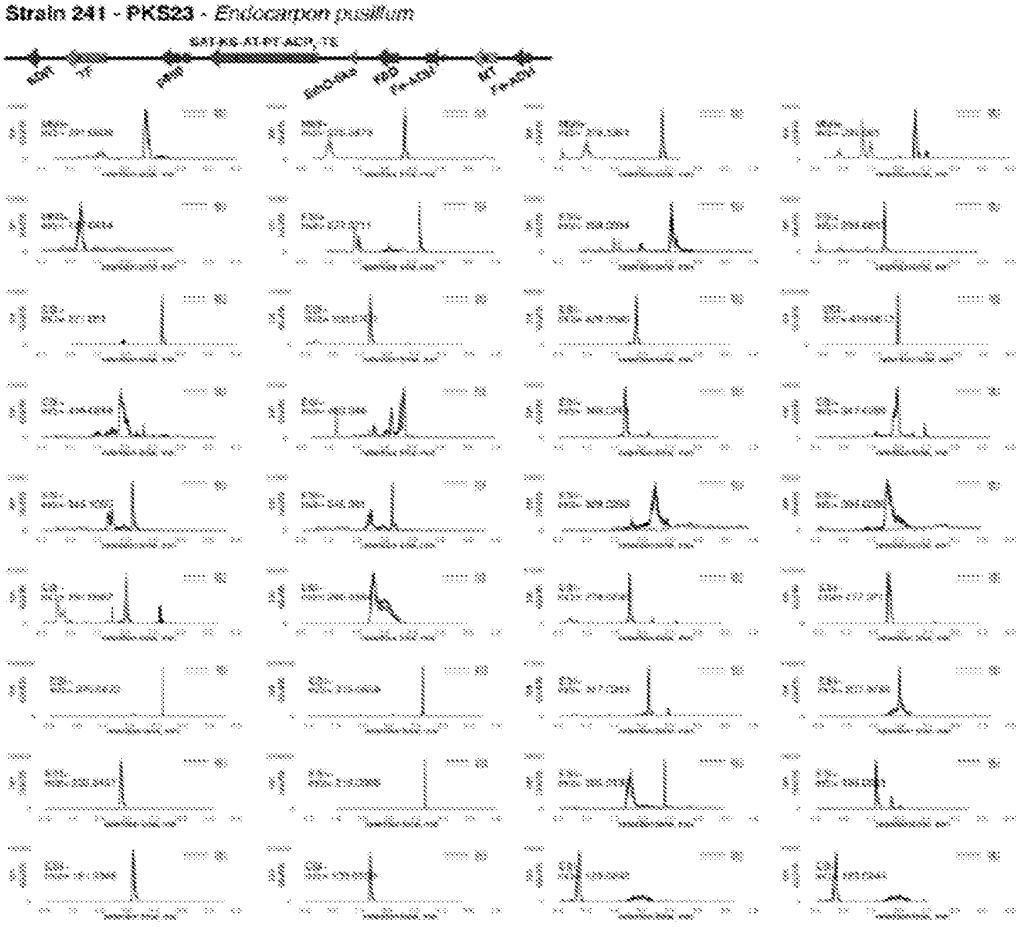


FIG. 46

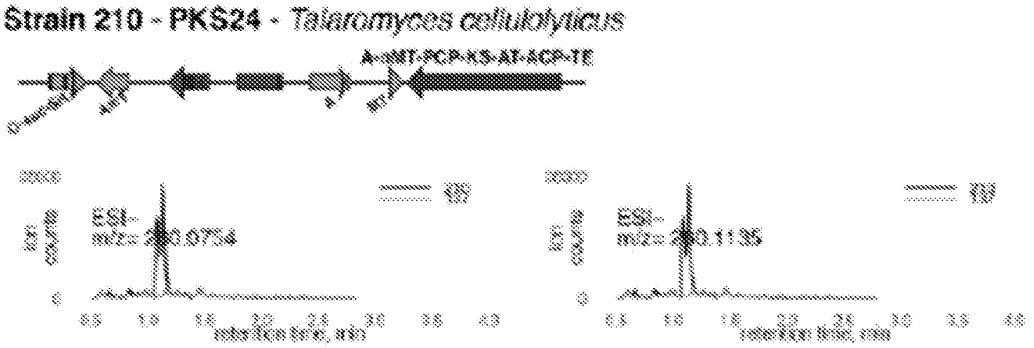

Strain 210 - PKS24 - Talaromyces cellulolyticus
A-cMT-PCP-KS-AT-ACP-TE
ESI-
ESI-

FIG. 47

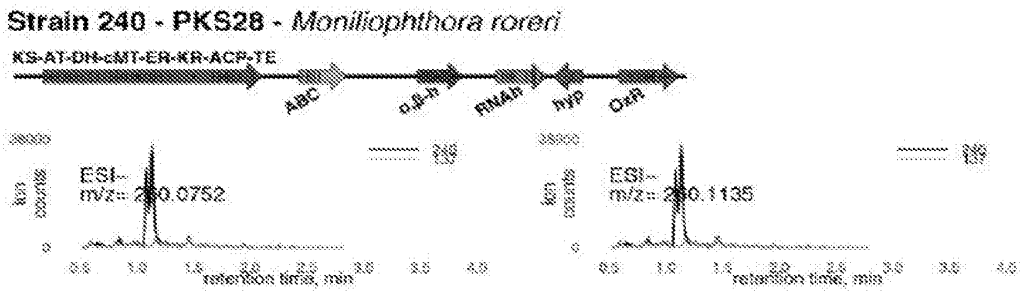
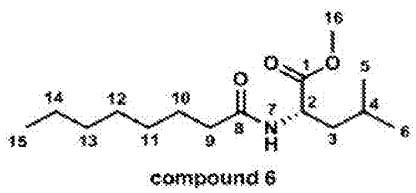
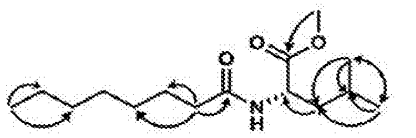

Strain 240 - PKS28 - Moniliophthora roreri
KS-AT-DH-cMT-ER-KR-ACP-TE
ABC
o.g.-n
RNAb
hyp
OxR
ESI-
m/z= 240.0752
ESI-
m/z= 240.1135
retention time, min
retention time, min

FIG. 48

^{1}H NMR spectrum (500 MHz), ^{13}C NMR spectrum (125 MHz), $CDCl_3$



compound 6



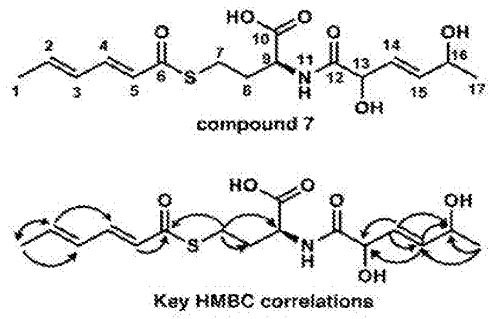
Key HMBC correlations

position	δ_H, mult (J in Hz)	δ_C	COSY
1		173.92	
2	4.66, td (4.9, 8.5)	50.85	H3, H7
3	1.51, m, overlap 1.64, m, overlap	41.96	H2, H4
4	1.64, m, overlap	25.04	
5[a]	0.84, m, overlap	22.96	H4
6[a]	0.84, m, overlap	22.14	H4
7-NH	5.76, d (8.5)		H2
8		173.02	
9	2.21, m	36.77	H8
10	1.64, m, overlap	25.73	H9
11	1.28, m, overlap	29.32	
12	1.28, m, overlap	29.19	
13	1.28, m, overlap	31.83	
14	1.28, m, overlap	22.75	
15	0.88, t (7.2)	14.22	H14
16	3.73, s	52.42	

Note: [a] Assignments denoted by superscript letters are interchangeable.

FIG. 49

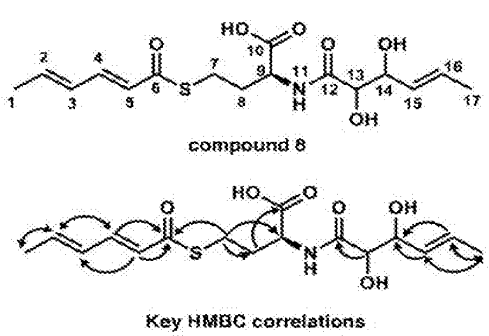
^{1}H NMR spectrum (500 MHz), ^{13}C NMR spectrum (125 MHz), Acetone-d_6



position	δ_H, mult (J in Hz)	δ_C	COSY
1	1.87, d (6.2)	18.85	H2
2	6.26, m	143.36	H1, H3
3	7.35, m	129.90	H2, H4
4	6.45, t (11.2)	142.64	H3, H5
5	5.88, d (11.3)	122.27	H4
6		185.98	
7	2.99, m	25.86	H8
8	2.08, m 2.19, m	33.12	H7, H11
9	4.59, m, overlap	51.89	H11
10		172.86	
11-NH	7.98, d (7.2)		H9
12		173.11	
13	4.59, m, overlap	72.71	H14
14	5.80, m	127.68	H13, H14
15	5.91, m	137.57	H14, H16
16	4.26, m	67.72	H15, H17
17	1.17, d (6.4)	23.95	H16

FIG. 50

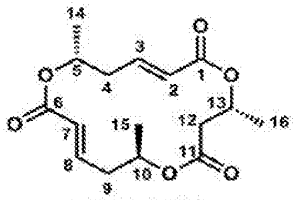
^{1}H NMR spectrum (500 MHz), ^{13}C NMR spectrum (125 MHz), Acetone-d_6



position	δ_H, mult (J in Hz)	δ_C	COSY
1	1.86, dd (7.0, 1.7)	18.84	H2
2	6.26, m	143.35	H1, H3
3	7.36, m	129.90	H2, H4
4	6.45, t (11.2)	142.83	H3, H5
5	5.89, dd (11.2, 0.85)	122.27	H4
6		186.97	
7	3.01, m	25.88	H8
8	2.09, m 2.19, m	33.03	H7, H11
9	4.60, m	52.03	H11
10		172.89	
11-NH	7.64, d (8.1)		H9
12		173.38	
13	4.06, d (3.6)	74.69	H14
14	4.34, m	73.73	H13, H14
15	5.61, m	131.40	H14, H16
16	5.75, m	127.51	H15, H17
17	1.66, dt (6.5, 1.4)	17.95	H16

FIG. 51

^{1}H NMR spectrum (500 MHz), ^{13}C NMR spectrum (125 MHz), $CDCl_3$



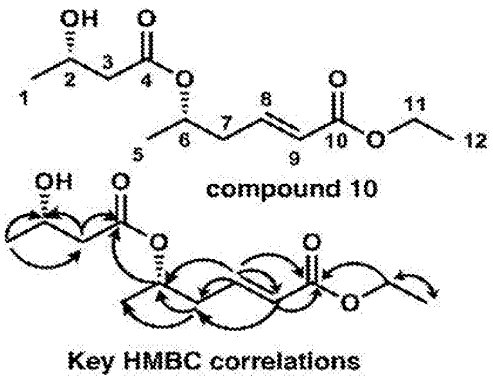
compound 9

Key HMBC correlations

position	δ_H, mult (J in Hz)	δ_C	COSY
1		165.07	
2	5.80, m	125.02	H3
3	6.84, m, overlap	143.61	H2, H4
4	2.31, m 2.49, m	38.99	H3, H5
5	5.05, m	66.72	H4, H14
6		165.46	
7	5.77, m	123.84	H8
8	6.80, m, overlap	144.65	H7, H9
9	2.39, m	39.50	H8, H10
10	5.11, m	70.21	H9, H15
11		170.30	
12	2.47, dd (15.9, 2.8) 2.57, dd (15.3, 10.6)	41.63	H13
13	5.35, m	67.81	H12, H16
14	1.94, d (6.3)	21.16	H5
15	1.27, d (6.3)	20.68	H10
16	1.30, d (6.3)	20.21	H13

FIG. 52

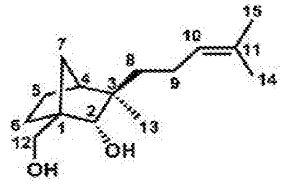
^{1}H NMR spectrum (500 MHz), ^{13}C NMR spectrum (125 MHz), $CDCl_3$



position	δ_H, mult (J in Hz)	δ_C	COSY
1	1.22, d (6.3)	22.56	H2
2	4.19, m, overlap	64.41	H1, H3
3	2.45, m, overlap	43.70	H2
4		172.48	
5	1.28, m, overlap	19.81	H6
6	5.07, m	69.71	H5, H7
7	2.45, m, overlap	38.48	H6, H8
8	6.86, dt (15.5, 7.4)	143.31	H7, H9
9	5.87, dt (15.5, 1.4)	124.51	H8
10		166.27	
11	4.19, m, overlap	60.55	H12
12	1.28, m, overlap	14.38	H11

FIG. 53

^{1}H NMR spectrum (500 MHz), ^{13}C NMR spectrum (125 MHz), $CDCl_3$

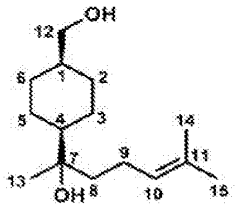


compound 11

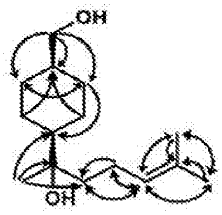
position	δ_H, mult (J in Hz)	δ_C
1		53.82
2	3.63, brs	82.58
3		42.57
4	1.88, m	46.26
5	1.46, m 1.74, m	26.83
6	1.20, m 1.88, m	22.16
7	1.09, d (10.1) 1.45, m	36.59
8	1.31, m	42.61
9	1.96, m	22.95
10	5.09, m	124.95
11		131.47
12	3.75, d (10.5) 3.84, d (10.5)	68.54
13	0.87, s	16.47
14	1.61, s	17.79
15	1.68, s	25.85

FIG. 54

^{1}H NMR spectrum (500 MHz), ^{13}C NMR spectrum (125 MHz), $CDCl_3$



compound 12



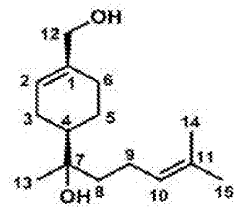
Key HMBC correlations

position	δ_H, mult (J in Hz)	δ_C	COSY
1	1.43, m	40.83	H2, H6, H12
2[a]	0.95, m, overlap 1.87, m, overlap	29.71	H1, H3
3[b]	1.08, m, overlap 1.81, m	26.95	H2, H4
4	1.32, m	47.62	H3, H5
5[b]	1.12, m, overlap 1.89, m, overlap	26.22	H4, H6
6[a]	0.95, m, overlap 1.87, m, overlap	29.76	H1, H5
7		74.51	
8	1.49, dd (9.9, 6.8)	39.90	H9
9	2.05, m	22.27	H8, H10
10	5.13, t (7.3)	124.70	H9
11		131.92	
12	3.46, d (6.3)	68.78	
13	1.11, s	24.03	
14	1.62, s	17.83	
15	1.69, s	25.88	

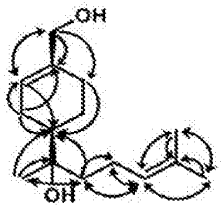
Note: [a,b] Assignments denoted by superscript letters are interchangeable.

FIG. 55

^{1}H NMR spectrum (500 MHz), ^{13}C NMR spectrum (125 MHz), $CDCl_3$



compound 13



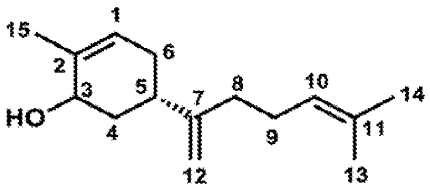
Key HMBC correlations

position	δ_H, mult (J in Hz)	δ_C	COSY
1		137.79	
2	5.68, m	122.56	H3
3	2.19, m, overlap 2.15, m	26.75	H2, H4
4	1.80, m, overlap	43.28	H3, H5
5	1.29, m 1.89, m, overlap	23.41	H4, H6
6	1.86, m 2.05, m, overlap	36.75	H5
7		74.44	
7-OH[a]	1.25, brs		
8	1.51, m	40.21	H9
9	2.06, m, overlap	22.23	H8, H10
10	5.13, t (7.3)	124.59	H9
11		132.03	
12	4.02, m	67.31	
12-OH[a]	1.15, brs		
13	1.12, s	23.41	
14	1.62, s	17.65	
15	1.69, s	25.88	

Note: [a] Assignments denoted by superscript letters are interchangeable.

FIG. 56

^{1}H NMR spectrum (500 MHz), ^{13}C NMR spectrum (125 MHz), $CDCl_3$



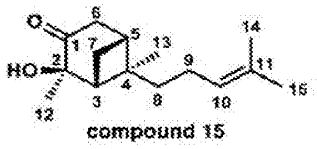
compound 14

Key HMBC correlations

position	δ_H, mult (J in Hz)	δ_C	COSY
1	5.59	125.69	H6
2		134.41	
3	4.02, brs	68.81	H4
4	1.59, m, overlap 1.95, m	37.34	H3, H5
5	2.33, m	34.14	H4, H6
6	1.83, m 2.16, m, overlap	31.74	H1, H5
7		153.42	
8	2.08, m	35.04	H9
9	2.13, m, overlap	26.90	H8, H10
10	5.12	124.30	H9
11		131.79	
12	4.78, s	107.88	
13	1.61, s	17.90	
14	1.68, s	25.85	
15	1.80, s	20.99	

FIG. 57

^{1}H NMR spectrum (500 MHz), ^{13}C NMR spectrum (125 MHz), $CDCl_3$



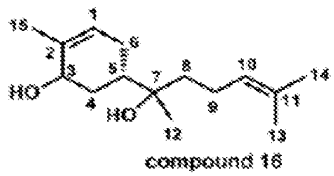
compound 15

Key HMBC correlations

position	δ_H, mult (J in Hz)	δ_C	COSY
1		214.32	
2		77.27	
2-OH	2.14, brs		
3	2.21, m, overlap	48.00	H7
4		42.59	
5	2.18, m, overlap	36.94	H6, H7
6	2.62, m	43.12	H5
7	1.68, m, overlap 2.43, m	28.77	H3, H5
8	1.86, m, overlap 1.80, m	39.12	H9
9	2.00, m	26.78	H8, H10
10	5.16, m	124.51	H9
11		131.82	
12	1.33, s	25.51	
13	0.88, s	19.90	
14	1.63, s	17.77	
15	1.70, s	25.87	

FIG. 58

^{1}H NMR spectrum (500 MHz), ^{13}C NMR spectrum (125 MHz), $CDCl_3$



compound 16

Key HMBC correlations

position	δ_H, mult (J in Hz)	δ_C	COSY
1	3.56, m	125.41	H6
2		134.61	
3	4.04, m	68.76	
4	1.42, td (13.2, 3.5, 1) 2.40, m, overlap	32.18	H5
5	1.76, m, overlap	36.98	H4, H6
6	1.74, m, overlap 2.04, m, overlap	27.32	H1, H5
7		78.86	
8	1.50, m	40.41	H9
9	2.06, m, overlap	22.30	H8, H10
10	5.12, m	124.57	H9
11		131.98	
12	1.11, s	23.39	
13	1.61, s	17.64	
14	1.66, s	25.86	
15	1.78, s	21.00	

FIG. 59

Electronic Patent Application Fee Transmittal

Application Number:	
Filing Date:	
Title of Invention:	BIOLOGICAL METHODS FOR PREPARING TERPENES
First Named Inventor/Applicant Name:	Kimberly Ann Aeling
Filer:	Silvia Salvadori
Attorney Docket Number:	SS118-1026US

Filed as Large Entity

Filing Fees for U.S. National Stage under 35 USC 371

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
NATIONAL STAGE FEE	1631	1	300	300
NATL STAGE SEARCH FEE - U.S. WAS THE ISA	1641	1	140	140
NATIONAL STAGE EXAM - ALL OTHER CASES	1633	1	760	760
Pages:				
NATL STAGE APPL SZ FEE PER 50 PGS >100	1681	11	400	4400
Claims:				
CLAIMS IN EXCESS OF 20	1615	1	100	100
INDEPENDENT CLAIMS IN EXCESS OF 3	1614	6	460	2760

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Miscellaneous-Filing:				
OATH/DECL > 30 MOS FROM 371 COMMENCEMENT	1617	1	140	140
Petition:				
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				
Extension-of-Time:				
Miscellaneous:				
		Total in USD ($)		**8600**

Electronic Acknowledgement Receipt	
EFS ID:	38275060
Application Number:	16630567
International Application Number:	PCT/US2018/041579
Confirmation Number:	5356
Title of Invention:	BIOLOGICAL METHODS FOR PREPARING TERPENES
First Named Inventor/Applicant Name:	Kimberly Ann Aeling
Customer Number:	122515
Filer:	Silvia Salvadori
Filer Authorized By:	
Attorney Docket Number:	SS118-1026US
Receipt Date:	13-JAN-2020
Filing Date:	
Time Stamp:	12:23:47
Application Type:	U.S. National Stage under 35 USC 371

Payment information:

Submitted with Payment	yes
Payment Type	CARD
Payment was successfully received in RAM	$8600
RAM confirmation Number	E20201CC25448054
Deposit Account	506423
Authorized User	Silvia Salvadori
The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows: 37 CFR 1.21 (Miscellaneous fees and charges)	

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Application Data Sheet	ADS.pdf	1874545 09502f5b5e000d0c226d7200b29a6be1a6e05002	no	10
Warnings:					
Information:					
2	Transmittal of New Application	Transmittal.pdf	255161 2414401735151 8fd80822b8aea591222e5d7490e	no	4
Warnings:					
Information:					
3		specification.pdf	14223610 e6ceffbe39bebce8888c9060c6128075fe5f18ed	yes	634

Multipart Description/PDF files in .zip description

Document Description	Start	End
Other Reference-Patent/App/Search documents	632	634
Drawings-other than black and white line drawings	573	631
Claims	552	572
Specification	2	551
Abstract	1	1

Warnings:					
Information:					
4	Documents submitted with 371 Applications	add_inventors.pdf	256548 9cfda26e0b692d4803e6539f90c70235c292b241	no	7
Warnings:					
Information:					

5	Documents submitted with 371 Applications	Change_Applicant.pdf	41899 d3249bac2fc0523321f585d6aaa9de666c49a8f4	no	1
Warnings:					
Information:					
6		Preliminary_Amendment.pdf	768990 675b6b0bc93c29f1d1acb69b5754e480c796df68	yes	29

Multipart Description/PDF files in .zip description

Document Description	Start	End
Drawings-only black and white line drawings	11	29
Applicant Arguments/Remarks Made in an Amendment	10	10
Claims	4	9
Specification	2	3
Preliminary Amendment	1	1

Warnings:					
Information:					
7	Sequence Listing	Sequence_Listing.pdf	788754 65177604f197d03a4bf6833a6a45e3d1d8a0f238	no	398
Warnings:					
Information:					
8	CRF Statement Paper and CRF are the same	Declaration_Sequence.pdf	23056 c4b4cc252a0a19ddda3bee8418e0563f0c5cc579	no	1
Warnings:					
Information:					
9	Information Disclosure Statement (IDS) Form (SB08)	IDS.pdf	1053709 4355fa6847ab7483cdc10300fae01c6a75371791	no	5
Warnings:					
Information:					

10	Foreign Reference	JP2012254044.pdf	3400336 824f3ba122e4e0f97bc1cac433286d2ad3787e56	no	46
Warnings:					
Information:					
11	Other Reference-Patent/App/Search documents	Report_patentability.pdf	2457509 09d7c0dd778066a62cfa9e447019eda42ab6ae72	no	54
Warnings:					
Information:					
12	Other Reference-Patent/App/Search documents	search_Written.pdf	4517705 04e48925792dd6f46eabad66cab4c25bb68bbc52	no	65
Warnings:					
Information:					
13	Other Reference-Patent/App/Search documents	second_written.pdf	185817 21132088b147c66c9cf4688ab0554786e463df4b	no	6
Warnings:					
Information:					
14	Foreign Reference	WO2014100504A2.pdf	4584013 fbb624d274a351b9e30a26df4dbe7adc01c43270	no	386
Warnings:					
Information:					
15	Foreign Reference	WO2018094110A2.pdf	5510842 62665d78b0e44a5f381ea809bab56b9a78b7d7ac	no	199
Warnings:					
Information:					
16	Sequence Listing (Text File)	Sequence_listing.txt	779312	no	-
Warnings:					
Information:					

17	Fee Worksheet (SB06)	fee-info.pdf	41668 e73588edc41ff02dc6b83b9dc3f15606d44bfc5d	no	2
Warnings:					
Information:					
			Total Files Size (in bytes):	40763474	

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

PTO/AIA/14 (02-18)
Approved for use through 11/30/2020. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	SS118-1026US
	Application Number	
Title of Invention	BIOLOGICAL METHODS FOR PREPARING TERPENES	

The application data sheet is part of the provisional or nonprovisional application for which it is being submitted. The following form contains the bibliographic data arranged in a format specified by the United States Patent and Trademark Office as outlined in 37 CFR 1.76.
This document may be completed electronically and submitted to the Office in electronic format using the Electronic Filing System (EFS) or the document may be printed and included in a paper filed application.

Secrecy Order 37 CFR 5.2:

☐ Portions or all of the application associated with this Application Data Sheet may fall under a Secrecy Order pursuant to 37 CFR 5.2 (Paper filers only. Applications that fall under Secrecy Order may not be filed electronically.)

Inventor Information:

Inventor 1 Remove

Legal Name

Prefix	Given Name	Middle Name	Family Name	Suffix
	Kimberly	Ann	Aeling	

Residence Information (Select One) ● US Residency Non US Residency Active US Military Service

City	State/Province	Country of Residence
San Clemente	CA	US

Mailing Address of Inventor:

Address 1	324 Plaza Estival,
Address 2	
City	San Clemente
State/Province	CA
Postal Code	92672
Country	US

Inventor 2 Remove

Legal Name

Prefix	Given Name	Middle Name	Family Name	Suffix
	Pierre-Yves		De Wals	

Residence Information (Select One) ◉ US Residency Non US Residency Active US Military Service

City	State/Province	Country of Residence
Encinitas	CA	US

Mailing Address of Inventor:

Address 1	925 North Vulcan Ave. Apt. #205
Address 2	
City	Encinitas
State/Province	CA
Postal Code	92024
Country	US

Inventor 3 Remove

Legal Name

Prefix	Given Name	Middle Name	Family Name	Suffix
	Eric		Scott	

Residence Information (Select One) ◉ US Residency Non US Residency Active US Military Service

PTO/AIA/14 (02-18)
Approved for use through 11/30/2020. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	SS118-1026US
	Application Number	
Title of Invention	BIOLOGICAL METHODS FOR PREPARING TERPENES	

City	Encinitas	State/Province	CA	Country of Residence	US

Mailing Address of Inventor:

Address 1	1579 Caudor Street		
Address 2			
City	Encinitas	State/Province	CA
Postal Code	92024	Country	US

Inventor 4 Remove

Legal Name

Prefix	Given Name	Middle Name	Family Name	Suffix
	Eric		Knight	

Residence Information (Select One) ◉ US Residency ○ Non US Residency ○ Active US Military Service

City	Carlsbad	State/Province	CA	Country of Residence	US

Mailing Address of Inventor:

Address 1	2744 Loker Ave West, unit 100		
Address 2			
City	Carlsbad	State/Province	CA
Postal Code	92010	Country	US

Inventor 5 Remove

Legal Name

Prefix	Given Name	Middle Name	Family Name	Suffix
	Jian		Yi	

Residence Information (Select One) ◉ US Residency ○ Non US Residency ○ Active US Military Service

City	San Diego	State/Province	CA	Country of Residence	US

Mailing Address of Inventor:

Address 1	13313 Thunderhead Street		
Address 2			
City	San Diego	State/Province	CA
Postal Code	92129	Country	US

Inventor 6 Remove

Legal Name

Prefix	Given Name	Middle Name	Family Name	Suffix
	Michael		Waldbridge	

Residence Information (Select One) ◉ US Residency ○ Non US Residency ○ Active US Military Service

City	Saugus	State/Province	MA	Country of Residence	US

PTO/AIA/14 (02-18)
Approved for use through 11/30/2020. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	SS118-1026US
	Application Number	
Title of Invention	BIOLOGICAL METHODS FOR PREPARING TERPENES	

Mailing Address of Inventor:

Address 1	4 Diane Dr		
Address 2			
City	Saugus	**State/Province**	MA
Postal Code	01906	**Country**	US

Inventor 7 Remove

Legal Name

Prefix	Given Name	Middle Name	Family Name	Suffix
	Thomas		Beardslee	

Residence Information (Select One) (•) US Residency ( ) Non US Residency ( ) Active US Military Service

City	Carlsbad	**State/Province**	CA	**Country of Residence**	US

Mailing Address of Inventor:

Address 1	6827 Carnation Drive		
Address 2			
City	Carlsbad	**State/Province**	CA
Postal Code	92011	**Country**	US

All Inventors Must Be Listed - Additional Inventor Information blocks may be generated within this form by selecting the **Add** button. Add

Correspondence Information:

Enter either Customer Number or complete the Correspondence Information section below.
For further information see 37 CFR 1.33(a).

☐ **An Address is being provided for the correspondence Information of this application.**

Customer Number	122515		
Email Address	silvia@salvadorilaw.com	Add Email	Remove Email

Application Information:

Title of the Invention	BIOLOGICAL METHODS FOR PREPARING TERPENES		
Attorney Docket Number	SS118-1026US	**Small Entity Status Claimed**	☐
Application Type	Nonprovisional		
Subject Matter	Utility		
Total Number of Drawing Sheets (if any)	59	**Suggested Figure for Publication (if any)**	

PTO/AIA/14 (02-18)
Approved for use through 11/30/2020. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	SS118-1026US
	Application Number	
Title of Invention	BIOLOGICAL METHODS FOR PREPARING TERPENES	

Filing By Reference:

Only complete this section when filing an application by reference under 35 U.S.C. 111(c) and 37 CFR 1.57(a). Do not complete this section if application papers including a specification and any drawings are being filed. Any domestic benefit or foreign priority information must be provided in the appropriate section(s) below (i.e., "Domestic Benefit/National Stage Information" and "Foreign Priority Information").

For the purposes of a filing date under 37 CFR 1.53(b), the description and any drawings of the present application are replaced by this reference to the previously filed application, subject to conditions and requirements of 37 CFR 1.57(a).

Application number of the previously filed application	Filing date (YYYY-MM-DD)	Intellectual Property Authority or Country

Publication Information:

☐ Request Early Publication (Fee required at time of Request 37 CFR 1.219)

☐ **Request Not to Publish.** I hereby request that the attached application not be published under 35 U.S.C. 122(b) and certify that the invention disclosed in the attached application **has not and will not** be the subject of an application filed in another country, or under a multilateral international agreement, that requires publication at eighteen months after filing.

Representative Information:

Representative information should be provided for all practitioners having a power of attorney in the application. Providing this information in the Application Data Sheet does not constitute a power of attorney in the application (see 37 CFR 1.32).
Either enter Customer Number or complete the Representative Name section below. If both sections are completed the customer Number will be used for the Representative Information during processing.

Please Select One:	● Customer Number	US Patent Practitioner	◯ Limited Recognition (37 CFR 11.9)
Customer Number	122515		

Domestic Benefit/National Stage Information:

This section allows for the applicant to either claim benefit under 35 U.S.C. 119(e), 120, 121, 365(c), or 386(c) or indicate National Stage entry from a PCT application. Providing benefit claim information in the Application Data Sheet constitutes the specific reference required by 35 U.S.C. 119(e) or 120, and 37 CFR 1.78.
When referring to the current application, please leave the "Application Number" field blank.

Prior Application Status	Pending		Remove
Application Number	Continuity Type	Prior Application Number	Filing or 371(c) Date (YYYY-MM-DD)
	a 371 of international	PCTUS201841579	2018-07-11

PTO/AIA/14 (02-18)
Approved for use through 11/30/2020. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	SS118-1026US
	Application Number	
Title of Invention	BIOLOGICAL METHODS FOR PREPARING TERPENES	

Prior Application Status	Expired		Remove
Application Number	Continuity Type	Prior Application Number	Filing or 371(c) Date (YYYY-MM-DD)
PCTUS201841579	Claims benefit of provisional	62532297	2017-07-13
Additional Domestic Benefit/National Stage Data may be generated within this form by selecting the **Add** button.			Add

Foreign Priority Information:

This section allows for the applicant to claim priority to a foreign application. Providing this information in the application data sheet constitutes the claim for priority as required by 35 U.S.C. 119(b) and 37 CFR 1.55. When priority is claimed to a foreign application that is eligible for retrieval under the priority document exchange program (PDX)[i] the information will be used by the Office to automatically attempt retrieval pursuant to 37 CFR 1.55(i)(1) and (2). Under the PDX program, applicant bears the ultimate responsibility for ensuring that a copy of the foreign application is received by the Office from the participating foreign intellectual property office, or a certified copy of the foreign priority application is filed, within the time period specified in 37 CFR 1.55(g)(1).

			Remove
Application Number	Country[i]	Filing Date (YYYY-MM-DD)	Access Code[i] (if applicable)
Additional Foreign Priority Data may be generated within this form by selecting the **Add** button.			Add

Statement under 37 CFR 1.55 or 1.78 for AIA (First Inventor to File) Transition Applications

☐ This application (1) claims priority to or the benefit of an application filed before March 16, 2013 and (2) also contains, or contained at any time, a claim to a claimed invention that has an effective filing date on or after March 16, 2013.
NOTE: By providing this statement under 37 CFR 1.55 or 1.78, this application, with a filing date on or after March 16, 2013, will be examined under the first inventor to file provisions of the AIA.

PTO/AIA/14 (02-18)
Approved for use through 11/30/2020. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	SS118-1026US
	Application Number	
Title of Invention	BIOLOGICAL METHODS FOR PREPARING TERPENES	

Authorization or Opt-Out of Authorization to Permit Access:

When this Application Data Sheet is properly signed and filed with the application, applicant has provided written authority to permit a participating foreign intellectual property (IP) office access to the instant application-as-filed (see paragraph A in subsection 1 below) and the European Patent Office (EPO) access to any search results from the instant application (see paragraph B in subsection 1 below).

Should applicant choose not to provide an authorization identified in subsection 1 below, applicant **must opt-out** of the authorization by checking the corresponding box A or B or both in subsection 2 below.

NOTE: This section of the Application Data Sheet is **ONLY** reviewed and processed with the **INITIAL** filing of an application. After the initial filing of an application, an Application Data Sheet cannot be used to provide or rescind authorization for access by a foreign IP office(s). Instead, Form PTO/SB/39 or PTO/SB/69 must be used as appropriate.

1. Authorization to Permit Access by a Foreign Intellectual Property Office(s)

A. Priority Document Exchange (PDX) - Unless box A in subsection 2 (opt-out of authorization) is checked, the undersigned hereby **grants the USPTO authority** to provide the European Patent Office (EPO), the Japan Patent Office (JPO), the Korean Intellectual Property Office (KIPO), the State Intellectual Property Office of the People's Republic of China (SIPO), the World Intellectual Property Organization (WIPO), and any other foreign intellectual property office participating with the USPTO in a bilateral or multilateral priority document exchange agreement in which a foreign application claiming priority to the instant patent application is filed, access to: (1) the instant patent application-as-filed and its related bibliographic data, (2) any foreign or domestic application to which priority or benefit is claimed by the instant application and its related bibliographic data, and (3) the date of filing of this Authorization. See 37 CFR 1.14(h)(1).

B. Search Results from U.S. Application to EPO - Unless box B in subsection 2 (opt-out of authorization) is checked, the undersigned hereby **grants the USPTO authority** to provide the EPO access to the bibliographic data and search results from the instant patent application when a European patent application claiming priority to the instant patent application is filed. See 37 CFR 1.14(h)(2).

The applicant is reminded that the EPO's Rule 141(1) EPC (European Patent Convention) requires applicants to submit a copy of search results from the instant application without delay in a European patent application that claims priority to the instant application.

2. Opt-Out of Authorizations to Permit Access by a Foreign Intellectual Property Office(s)

☐ A. Applicant **DOES NOT** authorize the USPTO to permit a participating foreign IP office access to the instant application-as-filed. If this box is checked, the USPTO will not be providing a participating foreign IP office with any documents and information identified in subsection 1A above.

☐ B. Applicant **DOES NOT** authorize the USPTO to transmit to the EPO any search results from the instant patent application. If this box is checked, the USPTO will not be providing the EPO with search results from the instant application.

NOTE: Once the application has published or is otherwise publicly available, the USPTO may provide access to the application in accordance with 37 CFR 1.14.

PTO/AIA/14 (02-18)
Approved for use through 11/30/2020. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	SS118-1026US
	Application Number	
Title of Invention	BIOLOGICAL METHODS FOR PREPARING TERPENES	

Applicant Information:

Providing assignment information in this section does not substitute for compliance with any requirement of part 3 of Title 37 of CFR to have an assignment recorded by the Office.

Applicant 1 Remove

If the applicant is the inventor (or the remaining joint inventor or inventors under 37 CFR 1.45), this section should not be completed. The information to be provided in this section is the name and address of the legal representative who is the applicant under 37 CFR 1.43; or the name and address of the assignee, person to whom the inventor is under an obligation to assign the invention, or person who otherwise shows sufficient proprietary interest in the matter who is the applicant under 37 CFR 1.46. If the applicant is an applicant under 37 CFR 1.46 (assignee, person to whom the inventor is obligated to assign, or person who otherwise shows sufficient proprietary interest) together with one or more joint inventors, then the joint inventor or inventors who are also the applicant should be identified in this section.

Clear

● Assignee	Legal Representative under 35 U.S.C. 117	Joint Inventor
Person to whom the inventor is obligated to assign.	Person who shows sufficient proprietary interest	

If applicant is the legal representative, indicate the authority to file the patent application, the inventor is:

Name of the Deceased or Legally Incapacitated Inventor:

If the Applicant is an Organization check here. ☒

Organization Name	Radici Chimica S.p.A.		
Mailing Address Information For Applicant:			
Address 1	Via G. Verdi, 11		
Address 2			
City	Bergamo	**State/Province**	
Country	IT	Postal Code	24121
Phone Number		Fax Number	
Email Address			

Additional Applicant Data may be generated within this form by selecting the Add button. Add

Assignee Information including Non-Applicant Assignee Information:

Providing assignment information in this section does not substitute for compliance with any requirement of part 3 of Title 37 of CFR to have an assignment recorded by the Office.

PTO/AIA/14 (02-18)
Approved for use through 11/30/2020. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	SS118-1026US
	Application Number	
Title of Invention	BIOLOGICAL METHODS FOR PREPARING TERPENES	

Assignee 1

Complete this section if assignee information, including non-applicant assignee information, is desired to be included on the patent application publication. An assignee-applicant identified in the "Applicant Information" section will appear on the patent application publication as an applicant. For an assignee-applicant, complete this section only if identification as an assignee is also desired on the patent application publication.

Remove

If the Assignee or Non-Applicant Assignee is an Organization check here. ☒

Organization Name	Radici Chimica S.p.A.

Mailing Address Information For Assignee including Non-Applicant Assignee:

Address 1	Via G. Verdi, 11		
Address 2			
City	Bergamo	**State/Province**	
Country i	IT	Postal Code	24121
Phone Number		Fax Number	
Email Address			

Additional Assignee or Non-Applicant Assignee Data may be generated within this form by selecting the Add button. Add

Signature:

Remove

NOTE: This Application Data Sheet must be signed in accordance with 37 CFR 1.33(b). **However, if this Application Data Sheet is submitted with the INITIAL filing of the application and either box A or B is not checked in subsection 2 of the "Authorization or Opt-Out of Authorization to Permit Access" section, then this form must also be signed in accordance with 37 CFR 1.14(c).**

This Application Data Sheet **must** be signed by a patent practitioner if one or more of the applicants is a **juristic entity** (e.g., corporation or association). If the applicant is two or more joint inventors, this form must be signed by a patent practitioner, **all** joint inventors who are the applicant, or one or more joint inventor-applicants who have been given power of attorney (e.g., see USPTO Form PTO/AIA/81) on behalf of **all** joint inventor-applicants.

See 37 CFR 1.4(d) for the manner of making signatures and certifications.

Signature	/Silvia Salvadori/			Date (YYYY-MM-DD)	2020-01-13
First Name	Silvia	Last Name	Salvadori	Registration Number	48265

Additional Signature may be generated within this form by selecting the Add button. Add

PTO/AIA/14 (02-18)
Approved for use through 11/30/2020. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	SS118-1026US
	Application Number	
Title of Invention	BIOLOGICAL METHODS FOR PREPARING TERPENES	

This collection of information is required by 37 CFR 1.76. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 23 minutes to complete, including gathering, preparing, and submitting the completed application data sheet form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

Privacy Act Statement

The Privacy Act of 1974 (P.L. 93-579) requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1 The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C. 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether the Freedom of Information Act requires disclosure of these records.

2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.

3 A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.

4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).

5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent CooperationTreaty.

6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).

7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (i.e., GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.

8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspections or an issued patent.

9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1221 of 3077

DocCode – SCORE

SCORE Placeholder Sheet for IFW Content

Application Number: 16630567 Document Date: 01/13/2020

The presence of this form in the IFW record indicates that the following document type was received in electronic format on the date identified above. This content is stored in the SCORE database.

Since this was an electronic submission, there is no physical artifact folder, no artifact folder is recorded in PALM, and no paper documents or physical media exist. The TIFF images in the IFW record were created from the original documents that are stored in SCORE.

- Drawing

At the time of document entry (noted above):

- USPTO employees may access SCORE content via DAV or via the SCORE web page.
- External customers may access SCORE content via PAIR using the Supplemental Content tab.

Form Revision Date: March 1, 2019

DocCode – SEQ.TXT

SCORE Placeholder Sheet for IFW Content

Application Number: 16630567 Document Date: 01/13/2020

The presence of this form in the IFW record indicates that the following document type was received in electronic format on the date identified above. This content is stored in the SCORE database.

Since this was an electronic submission, there is no physical artifact folder, no artifact folder is recorded in PALM, and no paper documents or physical media exist. The TIFF images in the IFW record were created from the original documents that are stored in SCORE.

- Sequence Listing

At the time of document entry (noted above):

- USPTO employees may access SCORE content via DAV or via the SCORE web page.
- External customers may access SCORE content via PAIR using the Supplemental Content tab.

Form Revision Date: March 1, 2019

==========
Sequence Listing was accepted.
If you need help call the Patent Electronic Business Center at (866) 217-9197 (toll free).
Reviewer: Anjum, Durreshwar
Timestamp: [year=2020; month=1; day=13; hr=16; min=15; sec=15; ms=341;]
==========

Validated By CRFValidator v 1.0.5

Application No: 16630567 Version No: 1.0

Input Set:

Output Set:

Started: 2020-01-13 12:26:00.064
Finished: 2020-01-13 12:26:03.934
Elapsed: 0 hr(s) 0 min(s) 3 sec(s) 870 ms
Total Warnings: 486
Total Errors: 0
No. of SeqIDs Defined: 527
Actual SeqID Count: 527

Error code		Error Description
W	213	Artificial or Unknown found in <213> in SEQ ID (3)
W	213	Artificial or Unknown found in <213> in SEQ ID (4)
W	213	Artificial or Unknown found in <213> in SEQ ID (5)
W	213	Artificial or Unknown found in <213> in SEQ ID (10)
W	213	Artificial or Unknown found in <213> in SEQ ID (11)
W	213	Artificial or Unknown found in <213> in SEQ ID (12)
W	213	Artificial or Unknown found in <213> in SEQ ID (13)
W	213	Artificial or Unknown found in <213> in SEQ ID (16)
W	213	Artificial or Unknown found in <213> in SEQ ID (17)
W	213	Artificial or Unknown found in <213> in SEQ ID (18)
W	213	Artificial or Unknown found in <213> in SEQ ID (19)
W	213	Artificial or Unknown found in <213> in SEQ ID (32)
W	213	Artificial or Unknown found in <213> in SEQ ID (37)
W	213	Artificial or Unknown found in <213> in SEQ ID (38)
W	213	Artificial or Unknown found in <213> in SEQ ID (45)
W	213	Artificial or Unknown found in <213> in SEQ ID (57)
W	213	Artificial or Unknown found in <213> in SEQ ID (58)
W	213	Artificial or Unknown found in <213> in SEQ ID (60)
W	213	Artificial or Unknown found in <213> in SEQ ID (61)
W	213	Artificial or Unknown found in <213> in SEQ ID (62)

Input Set:

Output Set:

Started: 2020-01-13 12:26:00.064
Finished: 2020-01-13 12:26:03.934
Elapsed: 0 hr(s) 0 min(s) 3 sec(s) 870 ms
Total Warnings: 486
Total Errors: 0
No. of SeqIDs Defined: 527
Actual SeqID Count: 527

Error code	Error Description
	This error has occured more than 20 times, will not be displayed
W 447	n or Xaa used, for: SEQID(74) on line number 9581
W 447	n or Xaa used, for: SEQID(127) on line number 14008
W 447	n or Xaa used, for: SEQID(127) on line number 14014
W 447	n or Xaa used, for: SEQID(127) on line number 14022
W 402	Undefined organism found in <213> in SEQ ID (326)
W 402	Undefined organism found in <213> in SEQ ID (327)
W 402	Undefined organism found in <213> in SEQ ID (328)
W 402	Undefined organism found in <213> in SEQ ID (329)
W 402	Undefined organism found in <213> in SEQ ID (330)
W 402	Undefined organism found in <213> in SEQ ID (331)
W 402	Undefined organism found in <213> in SEQ ID (332)
W 402	Undefined organism found in <213> in SEQ ID (333)
W 402	Undefined organism found in <213> in SEQ ID (334)
W 402	Undefined organism found in <213> in SEQ ID (335)
W 402	Undefined organism found in <213> in SEQ ID (336)
W 402	Undefined organism found in <213> in SEQ ID (352)
W 402	Undefined organism found in <213> in SEQ ID (357)
W 402	Undefined organism found in <213> in SEQ ID (358)
W 402	Undefined organism found in <213> in SEQ ID (359)
W 402	Undefined organism found in <213> in SEQ ID (360)
W 402	Undefined organism found in <213> in SEQ ID (361)

Input Set:

Output Set:

Started: 2020-01-13 12:26:00.064
Finished: 2020-01-13 12:26:03.934
Elapsed: 0 hr(s) 0 min(s) 3 sec(s) 870 ms
Total Warnings: 486
Total Errors: 0
No. of SeqIDs Defined: 527
Actual SeqID Count: 527

Error code		Error Description
W	402	Undefined organism found in <213> in SEQ ID (362)
W	402	Undefined organism found in <213> in SEQ ID (363)
W	402	Undefined organism found in <213> in SEQ ID (364)
		This error has occured more than 20 times, will not be displayed
W	447	n or Xaa used, for: SEQID(509) on line number 24753
W	447	n or Xaa used, for: SEQID(509) on line number 24753
W	447	n or Xaa used, for: SEQID(509) on line number 24753
W	447	n or Xaa used, for: SEQID(509) on line number 24753
W	447	n or Xaa used, for: SEQID(509) on line number 24753
W	447	n or Xaa used, for: SEQID(509) on line number 24753
W	447	n or Xaa used, for: SEQID(509) on line number 24753
W	447	n or Xaa used, for: SEQID(509) on line number 24753
W	447	n or Xaa used, for: SEQID(509) on line number 24753
W	447	n or Xaa used, for: SEQID(509) on line number 24753
W	447	n or Xaa used, for: SEQID(509) on line number 24753
W	447	n or Xaa used, for: SEQID(509) on line number 24753
W	447	n or Xaa used, for: SEQID(509) on line number 24753
W	447	n or Xaa used, for: SEQID(509) on line number 24753
W	447	n or Xaa used, for: SEQID(509) on line number 24753
		This error has occured more than 20 times, will not be displayed

SEQUENCE LISTING

<110> VERDEZYNE (ABC), LLC

<120> BIOLOGICAL METHODS FOR PREPARING TERPENES

<130> VAL-3002-PC

<140> US 16/630,567
<141> 2020-01-13

<150> PCT/US2018/041579
<151> 2018-07-11

<150> 62/532,297
<151> 2017-07-13

<160> 527

<170> PatentIn version 3.5

<210> 1
<211> 267
<212> PRT
<213> Candida viswanathii

<400> 1

```
Met Val Ser Thr Lys Thr Tyr Thr Glu Arg Ala Ser Ala His Pro Ser
1               5                   10                  15


Lys Val Ala Gln Arg Leu Phe Arg Leu Met Glu Ser Lys Lys Thr Asn
            20                  25                  30


Leu Cys Ala Ser Ile Asp Val Thr Thr Thr Ala Glu Phe Leu Ser Leu
        35                  40                  45


Ile Asp Lys Leu Gly Pro His Ile Cys Leu Val Lys Thr His Ile Asp
    50                  55                  60


Ile Ile Ser Asp Phe Ser Tyr Glu Gly Thr Ile Glu Pro Leu Leu Val
65                  70                  75                  80


Leu Ala Glu Arg His Gly Phe Leu Ile Phe Glu Asp Arg Lys Phe Ala
                85                  90                  95


Asp Ile Gly Asn Thr Val Met Leu Gln Tyr Thr Ser Gly Val Tyr Arg
            100                 105                 110


Ile Ala Ala Trp Ser Asp Ile Thr Asn Ala His Gly Val Thr Gly Lys
        115                 120                 125
```

```
Gly Val Val Glu Gly Leu Lys Arg Gly Ala Glu Gly Val Glu Lys Glu
    130                 135                 140

Arg Gly Val Leu Met Leu Ala Glu Leu Ser Ser Lys Gly Ser Leu Ala
145                 150                 155                 160

His Gly Glu Tyr Thr Arg Glu Thr Ile Glu Ile Ala Lys Ser Asp Arg
                165                 170                 175

Glu Phe Val Ile Gly Phe Ile Ala Gln Arg Asp Met Gly Gly Arg Glu
            180                 185                 190

Glu Gly Phe Asp Trp Ile Ile Met Thr Pro Gly Val Gly Leu Asp Asp
        195                 200                 205

Lys Gly Asp Ala Leu Gly Gln Gln Tyr Arg Thr Val Asp Glu Val Val
    210                 215                 220

Leu Thr Gly Thr Asp Val Ile Ile Val Gly Arg Gly Leu Phe Gly Lys
225                 230                 235                 240

Gly Arg Asp Pro Glu Val Glu Gly Lys Arg Tyr Arg Asp Ala Gly Trp
                245                 250                 255

Lys Ala Tyr Leu Lys Arg Thr Gly Gln Leu Glu
            260                 265


<210> 2
<211> 627
<212> PRT
<213> Candida viswanathii

<400> 2
Met Phe Asn Phe Lys Leu Ser Gln Gln Val Leu Lys Asn Ser Thr Lys
1               5                   10                  15

Ser Ile Met Pro Ile Leu Lys Lys Pro Phe Ser Thr Ser His Ala Lys
            20                  25                  30

Gly Asp Leu Phe Lys Tyr Gln Ser Gln Leu Pro Lys Leu Pro Val Pro
        35                  40                  45

Thr Leu Glu Glu Thr Ala Ser Lys Tyr Leu Lys Thr Val Glu Pro Phe
```

```
    50                  55                  60

Leu Asn Gln Glu Gln Leu Glu Ser Thr Lys Ala Lys Val Ala Glu Phe
65                  70                  75                  80

Val Arg Pro Gly Gly Ala Gly Glu Ala Leu Gln Ala Arg Leu Asn Asn
                85                  90                  95

Phe Ala Ala Asp Lys Asp Asn Trp Leu Ala Glu Phe Trp Asp Asp Tyr
            100                 105                 110

Ala Tyr Met Ser Tyr Arg Asp Pro Val Val Pro Tyr Val Ser Tyr Phe
        115                 120                 125

Phe Ser His Lys Asp Val Lys Asn Ile Ile Gly Gln Asp Gln Leu Leu
    130                 135                 140

Lys Ala Thr Leu Ile Ala Tyr Tyr Thr Ile Glu Phe Gln Glu Lys Val
145                 150                 155                 160

Leu Asp Glu Ser Leu Asp Pro Glu Val Ile Lys Gly Asn Pro Phe Cys
                165                 170                 175

Met Asn Ala Phe Lys Tyr Met Phe Asn Asn Ser Arg Val Pro Ala Glu
            180                 185                 190

Gly Ser Asp Ile Thr Gln His Tyr Asn Gly Glu Glu Asn Gln Phe Phe
        195                 200                 205

Val Val Ile Tyr Lys Asn Asn Phe Tyr Lys Val Pro Thr His Lys Asn
    210                 215                 220

Gly Gln Arg Leu Thr Lys Gly Glu Ile Tyr Ser Tyr Leu Gln Glu Ile
225                 230                 235                 240

Lys Asn Asp Ala Thr Pro Lys Gly Leu Gly Leu Gly Ala Leu Thr Ser
                245                 250                 255

Leu Asn Arg Asp Glu Trp Leu Ser Ala Tyr Asn Asn Leu Leu Lys Ser
            260                 265                 270

Pro Ile Asn Glu Ala Ser Leu Gly Ser Ile Phe Ala Ser Ser Phe Val
        275                 280                 285
```

```
Ile Ala Leu Asp Ser Asn Asn Pro Val Thr Ile Glu Glu Lys Ser Lys
    290                 295                 300

Asn Cys Trp His Gly Asp Gly Gln Asn Arg Phe Phe Asp Lys Pro Leu
305                 310                 315                 320

Glu Phe Phe Val Ser Ala Asn Gly Asn Ser Gly Phe Leu Gly Glu His
                325                 330                 335

Ser Arg Met Asp Ala Thr Pro Thr Val Gln Leu Asn Asn Thr Ile Tyr
            340                 345                 350

Lys Gln Ile Leu Glu Thr Asn Pro Asn Asp Leu Ile Val Glu Ile Gly
        355                 360                 365

Ser Ser Ala Pro Arg Phe Gly Asn Ala Glu Ile Leu Pro Phe Asp Ile
    370                 375                 380

Asn Pro Thr Thr Arg Ala Asn Ile Lys Asp Ala Ile Ala Lys Phe Asp
385                 390                 395                 400

Ala Thr Ile Ala Ala His Asp Glu Glu Ile Phe Gln His Tyr Gly Tyr
                405                 410                 415

Gly Lys Gly Leu Ile Lys Lys Phe Lys Val Ser Pro Asp Ala Tyr Val
            420                 425                 430

Gln Leu Leu Met Gln Leu Ala Tyr Phe Lys Tyr Thr Gly Lys Ile Arg
        435                 440                 445

Pro Thr Tyr Glu Ser Ala Ala Thr Arg Lys Phe Leu Lys Gly Arg Thr
    450                 455                 460

Glu Thr Gly Arg Thr Val Ser Asn Glu Ser Lys Lys Phe Val Glu Thr
465                 470                 475                 480

Trp Ser Asp Pro Asn Ala Ser Ser Ala Asp Lys Val Ala Thr Phe Gln
                485                 490                 495

Ala Ala Ala Lys Gln His Val Ala Tyr Leu Ser Ala Ala Ala Asp Gly
            500                 505                 510
```

```
Lys Gly Val Asp Arg His Leu Phe Gly Leu Lys Gln Met Ile Gln Pro
        515                 520                 525

Gly Glu Pro Ile Pro Glu Ile Phe Thr Asp Pro Ile Phe Ser Tyr Ser
    530                 535                 540

Gln Thr Trp Tyr Ile Ser Ser Ser Gln Val Pro Ser Glu Phe Phe Gln
545                 550                 555                 560

Ser Trp Gly Trp Ser Gln Val Ile Asp Asp Gly Phe Gly Leu Ala Tyr
                565                 570                 575

Leu Ile Asn Asn Asp Trp Ile His Val His Ile Ser Cys Lys Arg Gly
            580                 585                 590

Asn Gly Leu Gln Ser Asp His Leu Lys Trp Tyr Leu Val Asp Ser Ala
        595                 600                 605

Asn Glu Met Lys Asp Val Leu Thr Lys Gly Leu Leu Thr Asp Ala Lys
    610                 615                 620

Pro Lys Leu
625
```

<210> 3
<211> 609
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 3

```
Met Pro Ile Leu Lys Lys Pro Phe Ser Thr Ser His Ala Lys Gly Asp
1               5                   10                  15

Leu Phe Lys Tyr Gln Ser Gln Leu Pro Lys Leu Pro Val Pro Thr Leu
            20                  25                  30

Glu Glu Thr Ala Ser Lys Tyr Leu Lys Thr Val Glu Pro Phe Leu Asn
        35                  40                  45

Gln Glu Gln Leu Glu Ser Thr Lys Ala Lys Val Ala Glu Phe Val Arg
    50                  55                  60
```

```
Pro Gly Gly Ala Gly Glu Ala Leu Gln Ala Arg Leu Asn Asn Phe Ala
65                  70                  75                  80

Ala Asp Lys Asp Asn Trp Leu Ala Glu Phe Trp Asp Asp Tyr Ala Tyr
                85                  90                  95

Met Ser Tyr Arg Asp Pro Val Val Pro Tyr Val Ser Tyr Phe Phe Ser
            100                 105                 110

His Lys Asp Val Lys Asn Ile Ile Gly Gln Asp Gln Leu Leu Lys Ala
        115                 120                 125

Thr Leu Ile Ala Tyr Tyr Thr Ile Glu Phe Gln Glu Lys Val Leu Asp
    130                 135                 140

Glu Ser Leu Asp Pro Glu Val Ile Lys Gly Asn Pro Phe Cys Met Asn
145                 150                 155                 160

Ala Phe Lys Tyr Met Phe Asn Asn Ser Arg Val Pro Ala Glu Gly Ser
                165                 170                 175

Asp Ile Thr Gln His Tyr Asn Gly Glu Glu Asn Gln Phe Phe Val Val
            180                 185                 190

Ile Tyr Lys Asn Asn Phe Tyr Lys Val Pro Thr His Lys Asn Gly Gln
        195                 200                 205

Arg Leu Thr Lys Gly Glu Ile Tyr Ser Tyr Leu Gln Glu Ile Lys Asn
    210                 215                 220

Asp Ala Thr Pro Lys Gly Leu Gly Leu Gly Ala Leu Thr Ser Leu Asn
225                 230                 235                 240

Arg Asp Glu Trp Leu Ser Ala Tyr Asn Asn Leu Leu Lys Ser Pro Ile
                245                 250                 255

Asn Glu Ala Ser Leu Gly Ser Ile Phe Ala Ser Ser Phe Val Ile Ala
            260                 265                 270

Leu Asp Ser Asn Asn Pro Val Thr Ile Glu Glu Lys Ser Lys Asn Cys
        275                 280                 285
```

```
Trp His Gly Asp Gly Gln Asn Arg Phe Phe Asp Lys Pro Leu Glu Phe
    290                 295                 300

Phe Val Ser Ala Asn Gly Asn Ser Gly Phe Leu Gly Glu His Ser Arg
305                 310                 315                 320

Met Asp Ala Thr Pro Thr Val Gln Leu Asn Asn Thr Ile Tyr Lys Gln
                325                 330                 335

Ile Leu Glu Thr Asn Pro Asn Asp Leu Ile Val Glu Ile Gly Ser Ser
            340                 345                 350

Ala Pro Arg Phe Gly Asn Ala Glu Ile Leu Pro Phe Asp Ile Asn Pro
        355                 360                 365

Thr Thr Arg Ala Asn Ile Lys Asp Ala Ile Ala Lys Phe Asp Ala Thr
    370                 375                 380

Ile Ala Ala His Asp Glu Glu Ile Phe Gln His Tyr Gly Tyr Gly Lys
385                 390                 395                 400

Gly Leu Ile Lys Lys Phe Lys Val Ser Pro Asp Ala Tyr Val Gln Leu
                405                 410                 415

Leu Met Gln Leu Ala Tyr Phe Lys Tyr Thr Gly Lys Ile Arg Pro Thr
            420                 425                 430

Tyr Glu Ser Ala Ala Thr Arg Lys Phe Leu Lys Gly Arg Thr Glu Thr
        435                 440                 445

Gly Arg Thr Val Ser Asn Glu Ser Lys Lys Phe Val Glu Thr Trp Ser
    450                 455                 460

Asp Pro Asn Ala Ser Ser Ala Asp Lys Val Ala Thr Phe Gln Ala Ala
465                 470                 475                 480

Ala Lys Gln His Val Ala Tyr Leu Ser Ala Ala Ala Asp Gly Lys Gly
                485                 490                 495

Val Asp Arg His Leu Phe Gly Leu Lys Gln Met Ile Gln Pro Gly Glu
            500                 505                 510
```

```
Pro Ile Pro Glu Ile Phe Thr Asp Pro Ile Phe Ser Tyr Ser Gln Thr
        515                 520                 525

Trp Tyr Ile Ser Ser Ser Gln Val Pro Ser Glu Phe Phe Gln Ser Trp
    530                 535                 540

Gly Trp Ser Gln Val Ile Asp Asp Gly Phe Gly Leu Ala Tyr Leu Ile
545                 550                 555                 560

Asn Asn Asp Trp Ile His Val His Ile Ser Cys Lys Arg Gly Asn Gly
                565                 570                 575

Leu Gln Ser Asp His Leu Lys Trp Tyr Leu Val Asp Ser Ala Asn Glu
            580                 585                 590

Met Lys Asp Val Leu Thr Lys Gly Leu Leu Thr Asp Ala Lys Pro Lys
        595                 600                 605

Leu
```

<210> 4
<211> 606
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
polypeptide

<400> 4

```
Met Pro Ile Leu Lys Lys Pro Phe Ser Thr Ser His Ala Lys Gly Asp
1               5                   10                  15

Leu Phe Lys Tyr Gln Ser Gln Leu Pro Lys Leu Pro Val Pro Thr Leu
            20                  25                  30

Glu Glu Thr Ala Ser Lys Tyr Leu Lys Thr Val Glu Pro Phe Leu Asn
        35                  40                  45

Gln Glu Gln Leu Glu Ser Thr Lys Ala Lys Val Ala Glu Phe Val Arg
    50                  55                  60

Pro Gly Gly Ala Gly Glu Ala Leu Gln Ala Arg Leu Asn Asn Phe Ala
65                  70                  75                  80
```

```
Ala Asp Lys Asp Asn Trp Leu Ala Glu Phe Trp Asp Asp Tyr Ala Tyr
                85                  90                  95

Met Ser Tyr Arg Asp Pro Val Val Pro Tyr Val Ser Tyr Phe Phe Ser
            100                 105                 110

His Lys Asp Val Lys Asn Ile Ile Gly Gln Asp Gln Leu Leu Lys Ala
        115                 120                 125

Thr Leu Ile Ala Tyr Tyr Thr Ile Glu Phe Gln Glu Lys Val Leu Asp
    130                 135                 140

Glu Ser Leu Asp Pro Glu Val Ile Lys Gly Asn Pro Phe Cys Met Asn
145                 150                 155                 160

Ala Phe Lys Tyr Met Phe Asn Asn Ser Arg Val Pro Ala Glu Gly Ser
                165                 170                 175

Asp Ile Thr Gln His Tyr Asn Gly Glu Glu Asn Gln Phe Phe Val Val
            180                 185                 190

Ile Tyr Lys Asn Asn Phe Tyr Lys Val Pro Thr His Lys Asn Gly Gln
        195                 200                 205

Arg Leu Thr Lys Gly Glu Ile Tyr Ser Tyr Leu Gln Glu Ile Lys Asn
    210                 215                 220

Asp Ala Thr Pro Lys Gly Leu Gly Leu Gly Ala Leu Thr Ser Leu Asn
225                 230                 235                 240

Arg Asp Glu Trp Leu Ser Ala Tyr Asn Asn Leu Leu Lys Ser Pro Ile
                245                 250                 255

Asn Glu Ala Ser Leu Gly Ser Ile Phe Ala Ser Ser Phe Val Ile Ala
            260                 265                 270

Leu Asp Ser Asn Asn Pro Val Thr Ile Glu Glu Lys Ser Lys Asn Cys
        275                 280                 285

Trp His Gly Asp Gly Gln Asn Arg Phe Phe Asp Lys Pro Leu Glu Phe
    290                 295                 300
```

```
Phe Val Ser Ala Asn Gly Asn Ser Gly Phe Leu Gly Glu His Ser Arg
305                 310                 315                 320

Met Asp Ala Thr Pro Thr Val Gln Leu Asn Asn Thr Ile Tyr Lys Gln
                325                 330                 335

Ile Leu Glu Thr Asn Pro Asn Asp Leu Ile Val Glu Ile Gly Ser Ser
            340                 345                 350

Ala Pro Arg Phe Gly Asn Ala Glu Ile Leu Pro Phe Asp Ile Asn Pro
        355                 360                 365

Thr Thr Arg Ala Asn Ile Lys Asp Ala Ile Ala Lys Phe Asp Ala Thr
    370                 375                 380

Ile Ala Ala His Asp Glu Glu Ile Phe Gln His Tyr Gly Tyr Gly Lys
385                 390                 395                 400

Gly Leu Ile Lys Lys Phe Lys Val Ser Pro Asp Ala Tyr Val Gln Leu
                405                 410                 415

Leu Met Gln Leu Ala Tyr Phe Lys Tyr Thr Gly Lys Ile Arg Pro Thr
            420                 425                 430

Tyr Glu Ser Ala Ala Thr Arg Lys Phe Leu Lys Gly Arg Thr Glu Thr
        435                 440                 445

Gly Arg Thr Val Ser Asn Glu Ser Lys Lys Phe Val Glu Thr Trp Ser
    450                 455                 460

Asp Pro Asn Ala Ser Ser Ala Asp Lys Val Ala Thr Phe Gln Ala Ala
465                 470                 475                 480

Ala Lys Gln His Val Ala Tyr Leu Ser Ala Ala Ala Asp Gly Lys Gly
                485
```

PATENT COOPERATION TREATY

PCT

INTERNATIONAL PRELIMINARY REPORT ON PATENTABILITY

(Chapter II of the Patent Cooperation Treaty)

(PCT Article 36 and Rule 70)

Applicant's or agent's file reference **SS118-1026**	**FOR FURTHER ACTION**	See Form PCT/IPEA/416
International application No. **PCT/US2018/041579**	International filing date *(day/month/year)* **11 July 2018 (11.07.2018)**	Priority date *(day/month/year)* **13 July 2017 (13.07.2017)**
International Patent Classification (IPC) or both national classification and IPC IPC: C12N 1/16, 5/10, 9/88, 9/96, 15/09, 15/52; C07H 21/04; C07K 14/415 CPC: See Supplemental Sheet		
Applicant **RADICI CHIMICA S.P.A.**		

1. This report is the international preliminary examination report, established by this International Preliminary Examining Authority under Article 35 and transmitted to the applicant according to Article 36.
2. This REPORT consists of a total of **6** sheets, including this cover sheet.
3. This report is also accompanied by ANNEXES, comprising:
   a. ☒ *(sent to the applicant and to the International Bureau)* a total of 48 sheets, as follows:
      ☒ sheets of the description, claims and/or drawings which have been amended and/or sheets containing rectifications authorized by this Authority, unless those sheets were superseded or cancelled, and any accompanying letters (see Rules 46.5, 66.8, 70.16, 91.2, and Section 607 of the Administrative Instructions).
      ☐ sheets containing rectifications, where the decision was made by this Authority not to take them into account because they were not authorized by or notified to this Authority at the time when this Authority began to draw up this report, and any accompanying letters (Rules 66.4*bis*, 70.2(e), 70.16 and 91.2).
      ☐ superseded sheets and any accompanying letters, where this Authority either considers that the superseding sheets contain an amendment that goes beyond the disclosure in the international application as filed, or the superseding sheets were not accompanied by a letter indicating the basis for the amendments in the application as filed, as indicated in item 4 of Box No. I and the Supplemental Box (see Rule 70.16(b)).
   b. ☐ *(sent to the International Bureau only)* a total of (indicate type and number of electronic carrier(s)) _____ containing a sequence listing, in the form of an Annex C/ST.25 text file, as indicated in the Supplemental Box Relating to Sequence Listing (see paragraph 3*ter* of Annex C of the Administrative Instructions).
4. This report contains indications relating to the following items:
   - ☑ Box No. I Basis of the report
   - ☐ Box No. II Priority
   - ☑ Box No. III Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
   - ☐ Box No. IV Lack of unity of invention
   - ☑ Box No. V Reasoned statement under Article 35(2) with regard to novelty, inventive step and industrial applicability; citations and explanations supporting such statement
   - ☐ Box No. VI Certain documents cited
   - ☑ Box No. VII Certain defects in the international application
   - ☐ Box No. VIII Certain observations on the international application

Date of submission of the demand	Date of completion of this report
13 May 2019 (13.05.2019)	**02 December 2019 (02.12.2019)**
Name and mailing address of the IPEA/US **COMMISSIONER FOR PATENTS MAIL STOP PCT, ATTN: IPEA/US P.O. BOX 1450 ALEXANDRIA, VA 22313-1450, UNITED STATES OF AMERICA** Facsimile No. **(571)273-8300**	Authorized officer **MANJUNATH RAO** Telephone No. **571-272-0939**

Form PCT/IPEA/409 (cover sheet) (January 2015)

INTERNATIONAL PRELIMINARY REPORT ON PATENTABILITY

International application No.
PCT/US2018/041579

Box No. I Basis of the report

1. With regard to the **language**, this report is based on:
   - [x] the international application in the language in which it was filed.
   - [ ] a translation of the international application into _____ which is the language of a translation furnished for the purposes of:
     - [ ] international search (Rules 12.3(a) and 23.1(b))
     - [ ] publication of the international application (Rule 12.4(a))
     - [ ] international preliminary examination (Rules 55.2(a) and/or 55.3(a) and (b))
2. With regard to the **elements** of the international application, this report is based on *(replacement sheets which have been furnished to the receiving Office in response to an invitation under Article 14 are referred to in this report as "originally filed" and are not annexed to this report)*:
   - [ ] the international application as originally filed/furnished, or
   - [x] the description:
     pages **1-550** as originally filed/furnished.
     pages* _____ received by this Authority on _____
     pages* _____ received by this Authority on _____
   - [x] the claims:
     Nos. _____ as originally filed/furnished.
     Nos.* _____ as amended (together with any statement) under Article 19
     Nos.* **1-155** received by this Authority on **10 June 2019 (10.06.2019)**
   - [x] the drawings:
     pages **1/59-59/59** as originally filed/furnished.
     pages* _____ received by this Authority on _____
     pages* _____ received by this Authority on _____
   - [ ] a sequence listing - see Supplemental Box Relating to Sequence Listing.
3. [x] The amendments have resulted in the cancellation of:
   - [ ] the description, pages _____
   - [x] the claims, Nos. **5,44,45,84-86,149**
   - [ ] the drawings, sheets/figs _____
   - [ ] the sequence listing *(specify)*: _____
4. [ ] This report has been established as if (some of) the amendments annexed to this report and listed below had not been made, since either they are considered to go beyond the disclosure as filed, or they were not accompanied by a letter indicating the basis for the amendments in the application as filed, as indicated in the Supplemental Box (Rules 70.2(c) and (c-*bis*)):
   - [ ] the description, pages _____
   - [ ] the claims, Nos. _____
   - [ ] the drawings, sheets/figs _____
   - [ ] the sequence listing *(specify)*: _____
5. [ ] This report has been established:
   - [ ] taking into account the **rectification of an obvious mistake** authorized by or notified to this Authority under Rule 91 (Rules 66.1(d-*bis*) and 70.2(e)).
   - [ ] without taking into account the **rectification of an obvious mistake** authorized by or notified to this Authority under Rule 91(Rules 66.4*bis* and 70.2(e)).
6. With regard to top-up searches (Rules 66.1*ter* and 70.2(f)):
   - [x] A top-up search was carried out by this Authority on **03 September 2019 (03.09.2019)**
     - [ ] Additional relevant documents have been discovered during the top-up search.
   - [ ] No top-up search was carried out by this Authority because it would serve no useful purpose.
7. [ ] Supplementary international search report(s) from Authority(ies) _____ has/ have been received and taken into account in establishing this report (Rule 45*bis*.8(b) and (c)).

** If item 4 applies, some or all of those sheets may be marked "superseded."*

Form PCT/IPEA/409 (Box No. I) (January 2015)

INTERNATIONAL PRELIMINARY REPORT ON PATENTABILITY

International application No.
PCT/US2018/041579

Box No. III Non-establishment of opinion with regard to novelty, inventive step and industrial applicability

The questions whether the claimed invention appears to be novel, to involve an inventive step (to be non obvious), or to be industrially applicable have not been examined in respect of:

☐ the entire international application.

☑ claims Nos. **113 and 114**

because:

☐ the said international application, or the said claims Nos. _____ relate to the following subject matter which does not require an international preliminary examination (*specify*):

☑ the description, claims or drawings *(indicate particular elements below)* or said claims Nos. **113 and 114** are so unclear that no meaningful opinion could be formed (*specify*):

Claims 113 and 114 are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

☐ the claims, or said claims Nos. _____ are so inadequately supported by the description that no meaningful opinion could be formed (*specify*):

☐ no international search report has been established for said claims Nos. _____

☐ a meaningful opinion could not be formed without the sequence listing; the applicant did not, within the prescribed time limit:

☐ furnish a sequence listing in the form of an Annex C/ST.25 text file, and such listing was not available to the International Preliminary Examining Authority in the form and manner acceptable to it; or the sequence listing furnished did not comply with the standard provided for in Annex C of the Administrative Instructions.

☐ furnish a sequence listing on paper or in the form of an image file complying with the standard provided for in Annex C of the Administrative Instructions, and such listing was not available to the International Preliminary Examining Authority in the form and manner acceptable to it; or the sequence listing furnished did not comply with the standard provided for in Annex C of the Administrative Instructions.

☐ pay the required late furnishing fee for the furnishing of a sequence listing in response to an invitation under Rules 13*ter*.1(a) or (b) and 13*ter*.2.

☐ See Supplemental Box for further details.

INTERNATIONAL PRELIMINARY REPORT ON PATENTABILITY

International application No.
PCT/US2018/041579

Box No. V **Reasoned statement under Article 35(2) with regard to novelty, inventive step and industrial applicability; citations and explanations supporting such statement**

1. Statement

Novelty (N)	Claims	1-112 and 115-155	YES
	Claims		NO
Inventive step (IS)	Claims	1-112 and 115-155	YES
	Claims		NO
Industrial applicability (IA)	Claims	1-112 and 115-155	YES
	Claims		NO

2. Citations and explanations (Rule 70.7)

---------- **EXPLANATION** ----------

The claims possess novelty and an inventive step under PCT Article 33(2) & 33(3) as the prior art does not teach nor suggest a genetically modified microorganism, such as a yeast/*Candida* having one or more heterologous terpene biosynthesis enzymes selected from those in claim 1, wherein the expression of said enzymes is provided by a nucleic acid that is responsive to fatty acid or alkane induction (the former for cytochrome P450).

The claims possess industial applicability under PCT Article 33(4) as production of terpenes are important in many industrial fields such as foods, cosmetics, pharmaceutical, etc.

INTERNATIONAL PRELIMINARY REPORT ON PATENTABILITY

International application No.
PCT/US2018/041579

Supplemental Box

In case **the space in any of the preceding boxes is not sufficient.**

Continuation of CPC:
C12N 1/16, 5/10, 9/88, 9/96, 15/09, 15/52, 15/63, 15/70, 15/74, 15/82, 15/8249; C07H 21/04; C07K 14/415; C12P 7/02, 7/26

INTERNATIONAL PRELIMINARY REPORT ON PATENTABILITY

International application No.
PCT/US2018/041579

Box No. VII Certain defects in the international application

The following defects in the form or contents of the international application have been noted:

In claims 140-142, beginning of second lines, a space should exist between "encoding" and the following word.

What is claimed is:

1. A genetically modified microorganism, comprising:

one or more heterologous nucleic acids encoding one or more terpene biosynthesis polypeptides, wherein expression of at least one of the heterologous nucleic acids is regulated by a nucleic acid that provides for fatty acid or alkane induction of expression of the terpene biosynthesis polypeptide:

- wherein the one or more heterologous nucleic acids that encode one or more terpene biosynthesis polypeptides is chosen from terpene synthase, phytoene synthase, geranylgeranyl diphosphate synthase, phytoene desaturase, lycopene cyclase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase, zeaxanthin glucosyltransferase and valencene synthase, or
- wherein the one or more heterologous nucleic acids is a cytochrome P450 reductase and is regulated by a nucleic acid that provides for alkane induction of expression of the cytochrome P450 reductase.

2. The microorganism of claim 1, wherein the microorganism is a fungus.

3. The microorganism of claim 2, wherein the fungus is a yeast.

4. The microorganism of claim 3, wherein the yeast is chosen from *Candida* spp, *Yarrowia* spp, *Rhodotorula* spp, *Rhodosporidium* spp, *Cryptococcus* spp, *Trichosporon* spp, *Lipomyces* spp, and *Blastobotrys* spp.

5. The microorganism of claim 4, wherein the fatty acid is a saturated fatty acid or an unsaturated fatty acid.

6. The microorganism of claim 5, wherein the fatty acid is chosen from one or more of oleic acid, palmitoleic acid, erucic acid, linoleic acid, palmitic acid, caproic acid, enanthic acid, caprylic acid pelargonic acid, capric acid, undecylic acid, lauric acid, myristic acid, pentadecanoic acid, margaric acid, stearic acid arachidic acid, behenic acid, tridecylic acid, and linolenic acid.

7. The microorganism of claim 4, wherein the alkane is chosen from one or more of hexane, heptane, nonane, decane, undecane, dodecane, tridecane, tetradecane, pentadecane, hexadecane, heptadecane, and octadecane

8. The microorganism of claim 7, wherein the nucleic acid that provides for fatty acid or alkane induction of expression of a terpene biosynthesis polypeptide comprises a fatty acid response element or an alkane response element.

9. The microorganism of claim 8, wherein the fatty acid response element comprises an oleic acid response element.

10. The microorganism of claim 8, wherein the alkane response element comprises an alkane response element 1 (ARE1) sequence.

11. The microorganism of claim 10, wherein the nucleic acid that provides for fatty acid induction of expression of a terpene biosynthesis polypeptide comprises a promoter region chosen from promoter regions of genes encoding hydratase-dehydrogenase-epimerase (HDE), acyl co-A oxidase (POX), acyl co-A thiolase (POT), peroxin (PEX) and peroxisomal adenine nucleotide transporter protein (ANT1).

12. The microorganism of claim 11, wherein the microorganism is *Candida viswanathii*.

13. The microorganism of claim 11, wherein the microorganism is *Blastobotrys adeninivorans*.

14. The microorganism of claim 13, wherein the one or more heterologous nucleic acids encode phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase.

15. The microorganism of claim 14, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a promoter region of the gene encoding a hydratase-dehydrogenase-epimerase (HDE).

16. The microorganism of claim 14, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a promoter region of the gene encoding a *Candida* hydratase-dehydrogenase-epimerase (HDE).

17. The microorganism of claim 14, 15 or 16, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a terminator region of the gene encodingan acyl co-A oxidase 4 (POX4).

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1245 of 3077

18. The microorganism of claim 14, 15 or 16, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a terminator region of the gene encodinga *Candida*acyl co-A oxidase 4 (POX4).

19. The microorganism of claim 13, wherein the one or more heterologous nucleic acids encode valencene synthase.

20. The microorganism of claim 19, wherein expression of the heterologous nucleic acids encoding valencene synthase is regulated by a promoter region of the gene encoding a hydratase-dehydrogenase-epimerase (HDE).

21. The microorganism of claim 19, wherein expression of the heterologous nucleic acids encoding valencene synthase is regulated by a promoter region of the gene encoding a *Candida* hydratase-dehydrogenase-epimerase (HDE).

22. The microorganism of claim 19, 20 or 21, wherein expression of the heterologous nucleic acids encoding valencene synthase is regulated by a terminator region of the gene encodingan acyl co-A oxidase 4 (POX4).

23. The microorganism of claim 19, 20 or 21, wherein expression of the heterologous nucleic acids encoding valencene synthase is regulated by a terminator region of the gene encodinga *Candida*acyl co-A oxidase 4 (POX4).

24. The microorganism of claim 13, wherein the one or more heterologous nucleic acids encode geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase.

25. The microorganism of claim 24, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase is regulated by a promoter region of the gene encoding a hydratase-dehydrogenase-epimerase (HDE).

26. The microorganism of claim 24, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1246 of 3077

cyclase/phytoene synthase is regulated by a promoter region of the gene encoding a *Candida* hydratase-dehydrogenase-epimerase (HDE).

27. The microorganism of claim 24, 25 or 26, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase is regulated by a terminator region of the gene encodingan acyl co-A oxidase 4 (POX4).

28. The microorganism of claim 24, 25 or 26, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase is regulated by a terminator region of the gene encodinga *Candida*acyl co-A oxidase 4 (POX4).

29. The microorganism of claim 28, wherein the one or more heterologous nucleic acids encoding the one or more terpene biosynthesis polypeptides are endogenously expressed in a microorganism chosen from *Cronobacter* spp, *Callitropsis* spp, *Xanthophyllomyces* spp, *Agrobacterium* spp, and *Pantoea* spp.

30. The microorganism of claim 29, wherein the amount and/or activity of aRas2 protein has been decreased.

31. The microorganism of claim 30, wherein the microorganism has been genetically modified to reduce or eliminate expression of an endogenous *RAS2* gene.

32. The microorganism of claim 31, wherein the amount and/or activity of an Faa1 protein has been decreased.

33. The microorganism of claim 32, wherein the microorganism has been genetically modified to reduce or eliminate expression of an endogenous *FAA1* gene.

34. The microorganism of claim 33, wherein the amounts and/or activities of one or more proteins chosen from acetyl-CoA C-acetyltransferase, HMG-CoA synthase, HMG-CoA reductase, mevalonate kinase, phosphomevalonate kinase, diphosphomevalonate decarboxylase, and isopentyl diphosphate delta isomerase,have been increased.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1247 of 3077

35. The microorganism of claim 34,wherein the microorganism has been genetically modified to increase expression of one or more endogenous genes chosen from *ERG10*, *ERG13*, *HMG1*, *ERG12*, *ERG8*, *MVD1*, and *IDI1*.

36. The microorganism of claim 35, wherein the amount and/or activity of dimethylallyltranstransferase or farnesyl diphosphate synthetase has been increased.

37. The microorganism of claim 36, wherein the microorganism has been genetically modified to increase expression of an endogenous *ERG20* gene.

38. The microorganism of claim 35, wherein expression of *ERG10*, *ERG13*, *HMG1*, *ERG12*, *ERG8*, *MVD1*, and/or *IDI1*is regulated by a promoter region of the gene encoding oleate-induced peroxisomal protein (POX18).

39. The microorganism of claim 35 or 38, wherein expression of *ERG10*, *ERG13*, *HMG1*, *ERG12*, *ERG8*, *MVD1*, and/or *IDI1*is regulated by a terminator region of the gene encoding oleate-induced peroxisomal protein (POX18).

40. The microorganism of claim 37, wherein expression of*ERG20*is regulated by a promoter region of the gene encoding oleate-induced peroxisomal protein (POX18).

41. The microorganism of claim 37 or 40, wherein expression of*ERG20*is regulated by a terminator region of the gene encoding oleate-induced peroxisomal protein (POX18).

42. A genetically modified microorganism, comprising:

one or more heterologous nucleic acids encoding one or more terpene biosynthesis polypeptides, and

a genetic modification that alters the expression of a polypeptide providing for transport of acetyl-carnitine in the microorganism,

wherein the expression of at least one of the heterologous nucleic acids is regulated by a nucleic acid that provides for glucose, fatty acid or alkane induction of expression of the terpene biosynthesis polypeptide.

43. The microorganism of claim 42, wherein the microorganism is a fungus.

44. The microorganism of claim 43, wherein the fungus is a yeast.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1248 of 3077

45. The microorganism of claim 44, wherein the yeast is chosen from *Candida* spp, *Yarrowia* spp, *Rhodotorula* spp, *Rhodosporidium* spp, *Cryptococcus* spp, *Trichosporon* spp, *Lipomyces* spp, and *Blastobotrys* spp.

46. The microorganism of claim 42, wherein the one or more heterologous nucleic acids encode one or more terpene biosynthesis polypeptides chosen from terpene synthase, phytoene synthase, geranylgeranyl diphosphate synthase, phytoene desaturase, lycopene cyclase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase, zeaxanthin glucosyltransferase, valencene synthase, and cytochrome p450 reductase.

47. The microorganism of claim 46, wherein the genetic modification reduces expression of the polypeptide providing for transport of acetyl-carnitine in the microorganism relative to a microorganism that does not have the genetic modification.

48. The microorganism of claim 47, wherein the genetic modification is a disruption, deletion or knockout of (i) a polynucleotide that encodes a polypeptide providing for transport of acetyl-carnitine, or (ii) a promoter operably linked to a polynucleotide that encodes a polypeptide providing for transport of acetyl-carnitine, whereby endogenous activity of a polypeptide providing for transport of acetyl-carnitine is reduced or abolished.

49. The microorganism of claim 47, further comprising a nucleic acid encoding a polypeptide providing for transport of acetyl-carnitine wherein expression of the nucleic acid encoding a polypeptide providing for transport of acetyl-carnitine is regulated by a promoter that provides for reduced expression relative to endogenous expression.

50. The microorganism of claim 49, wherein the genetic modification comprises replacing the promoter of an endogenous gene encoding the polypeptide providing for transport of acetyl-carnitine in the microorganism with a promoter that provides for reduced expression of the polypeptide in the microorganism relative to a microorganism that does not have the genetic modification.

51. The microorganism of claim 49 or 50, wherein the promoter that provides for reduced expression relative to endogenous expression is a promoter for a glucose-6-phosphate isomerase (G6PI) gene.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1249 of 3077

52. The microorganism of claim 49 or 50, wherein the promoter that provides for reduced expression relative to endogenous expression is a promoter for a *Candida* glucose-6-phosphate isomerase (G6PI) gene.

53. The microorganism of claim 42, wherein the polypeptide providing for transport of acetyl-carnitine in the microorganism is an acetyl-carnitine translocase (CRC).

54. The microorganism of claim 53, wherein the acetyl-carnitine translocase (CRC) is acetyl-carnitine translocase 1 (CRC1).

55. The microorganism of claim 42, wherein the microorganism is *Candida viswanathii.*

56. The microorganism of claim 42, wherein the microorganism is *Blastobotrys adeninivorans.*

57. The microorganism of claim 42, wherein the one or more heterologous nucleic acids encode phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase.

58. The microorganism of claim 57, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a promoter region of the gene encoding a hydratase-dehydrogenase-epimerase (HDE).

59. The microorganism of claim 57, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a promoter region of the gene encoding a *Candida* hydratase-dehydrogenase-epimerase (HDE).

60. The microorganism of claim 57, 58 or 59, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a terminator region of the gene encodingan acyl co-A oxidase 4 (POX4).

61. The microorganism of claim 57, 58 or 59, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a terminator region of the gene encodinga *Candida*acyl co-A oxidase 4 (POX4).

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1250 of 3077

62. The microorganism of claim 42, wherein the one or more heterologous nucleic acids encode geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase.

63. The microorganism of claim 62, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase is regulated by a promoter region of the gene encoding a hydratase-dehydrogenase-epimerase (HDE).

64. The microorganism of claim 62, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase is regulated by a promoter region of the gene encoding a *Candida* hydratase-dehydrogenase-epimerase (HDE).

65. The microorganism of claim 62, 63 or 64, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase is regulated by a terminator region of the gene encodingan acyl co-A oxidase 4 (POX4).

66. The microorganism of claim 62, 63 or 64, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase is regulated by a terminator region of the gene encodinga *Candida*acyl co-A oxidase 4 (POX4).

67. The microorganism of claim 42, wherein the one or more heterologous nucleic acids encoding the one or more terpene biosynthesis polypeptides are endogenously expressed in a microorganism chosen from *Cronobacter* spp, *Callitropsis* spp, *Xanthophyllomyces* spp, *Agrobacterium* spp, and *Pantoea* spp.

68. The microorganism of claim 42, wherein the amount and/or activity of aRas2 protein has been decreased.

69. The microorganism of claim 42, wherein the microorganism has been genetically modified to reduce or eliminate expression of an endogenous *RAS2* gene.

70. The microorganism of claim 42, wherein the amount and/or activity of an Faa1 protein has been decreased.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1251 of 3077

71. The microorganism of claim 42, wherein the microorganism has been genetically modified to reduce or eliminate expression of an endogenous *FAA1* gene.

72. The microorganism of claim 42, wherein the amounts and/or activities of one or more proteins chosen from acetyl-CoA C-acetyltransferase, HMG-CoA synthase, HMG-CoA reductase, mevalonate kinase, phosphomevalonate kinase, diphosphomevalonate decarboxylase, and isopentyl diphosphate delta isomerase, have been increased.

73. The microorganism of claim 42, wherein the microorganism has been genetically modified to increase expression of one or more endogenous genes chosen from *ERG10*, *ERG13*, *HMG1*, *ERG12*, *ERG8*, *MVD1*, and *IDI1*.

74. The microorganism of claim 42, wherein the amount and/or activity of dimethylallyltranstransferase or farnesyl diphosphate synthetase has been increased.

75. The microorganism of claim 42, wherein the microorganism has been genetically modified to increase expression of an endogenous *ERG20* gene.

76. The microorganism of claim 73, wherein expression of *ERG10*, *ERG13*, *HMG1*, *ERG12*, *ERG8*, *MVD1*, and/or *IDI1*is regulated by a promoter region of the gene encoding oleate-induced peroxisomal protein (POX18).

77. The microorganism of claim 73 or 76, wherein expression of *ERG10*, *ERG13*, *HMG1*, *ERG12*, *ERG8*, *MVD1*, and/or *IDI1*is regulated by a terminator region of the gene encoding oleate-induced peroxisomal protein (POX18).

78. The microorganism of claim 75, wherein expression of*ERG20*is regulated by a promoter region of the gene encoding oleate-induced peroxisomal protein (POX18).

79. The microorganism of claim 75 or 78, wherein expression of*ERG20*is regulated by a terminator region of the gene encoding oleate-induced peroxisomal protein (POX18).

80. A genetically modified *Candida viswanathii* yeast, comprising one or more heterologous nucleic acids encoding one or more terpene biosynthesis polypeptides, wherein expression of at least one of the heterologous nucleic acids is regulated by a nucleic acid that provides for glucose, fatty acid or alkane induction of expression of the terpene biosynthesis polypeptide and

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1252 of 3077

- wherein the one or more heterologous nucleic acids encoding one or more terpene biosynthesis polypeptides is chosen from terpene synthase, phytoene synthase, geranylgeranyl diphosphate synthase, phytoene desaturase, lycopene cyclase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase, zeaxanthin glucosyltransferase and valencene synthase, or
- wherein the one or more heterologous nucleic acids is a cytochrome P450 reductase and is regulated by a nucleic acid that provides for glucose or alkane induction of expression of the cytochrome P450 reductase.

81. The genetically modified *Candida viswanathii* yeast of claim 80, wherein the one or more heterologous nucleic acids encode phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase

82. The genetically modified *Candida viswanathii* yeast of claim 81, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a promoter region of the gene encoding a *Candida* hydratase-dehydrogenase-epimerase (HDE).

83. The genetically modified *Candida viswanathii* yeast of claim 82, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a terminator region of the gene encoding a *Candida*acyl co-A oxidase 4 (POX4).

84. The genetically modified *Candida viswanathii* yeast of claim 81, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a promoter region of the gene encoding a *Candida* glyceraldehyde-3-phosphate dehydrogenase (GPD).

85. The genetically modified *Candida viswanathii* yeast of claim 80, wherein the one or more heterologous nucleic acids encode valencene synthase.

86. The genetically modified *Candida viswanathii* yeast of claim 85, wherein expression of the heterologous nucleic acids encoding valencene synthase is regulated by a promoter region of the gene encoding a *Candida* hydratase-dehydrogenase-epimerase (HDE).

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1253 of 3077

87. The genetically modified *Candida viswanathii* yeast of claim 86, wherein expression of the heterologous nucleic acids valencene synthase is regulated by a terminator region of the gene encoding a *Candida*acyl co-A oxidase 4 (POX4).

88. The genetically modified *Candida viswanathii* yeast of claim 85, wherein expression of the heterologous nucleic acids encoding valencene synthase is regulated by a promoter region of the gene encoding a *Candida* glyceraldehyde-3-phosphate dehydrogenase (GPD).

89. The genetically modified *Candida viswanathii* yeast of claim 80, wherein the one or more heterologous nucleic acids encode geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase.

90. The genetically modified *Candida viswanathii* yeast of claim 89, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase is regulated by a promoter region of the gene encoding a *Candida* hydratase-dehydrogenase-epimerase (HDE).

91. The genetically modified *Candida viswanathii* yeast of claim 90, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase is regulated by a terminator region of the gene encoding a *Candida*acyl co-A oxidase 4 (POX4).

92. The genetically modified *Candida viswanathii* yeast of claim 89, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase is regulated by a promoter region of the gene encoding a *Candida* glyceraldehyde-3-phosphate dehydrogenase (GPD).

93. The genetically modified *Candida viswanathii* yeast of claim 80, wherein the one or more heterologous nucleic acids encode geranylgeranyl diphosphate synthase, phytoene desaturase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase,and cytochrome p450 reductase.

94. The genetically modified *Candida viswanathii* yeast of claim 93, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase,and cytochrome p450 reductase is regulated by a promoter region of the gene encoding a *Candida* hydratase-dehydrogenase-epimerase (HDE).

95. The genetically modified *Candida viswanathii* yeast of claim 94, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase,and cytochrome p450 reductase is regulated by a terminator region of the gene encoding a *Candida*acyl co-A oxidase 4 (POX4).

96. The genetically modified *Candida viswanathii* yeast of claim 93, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase,and cytochrome p450 reductase is regulated by a promoter region of the gene encoding a *Candida* glyceraldehyde-3-phosphate dehydrogenase (GPD).

97. The genetically modified *Candida viswanathii* yeast of claim 80, wherein the one or more heterologous nucleic acids encoding the one or more terpene biosynthesis polypeptides are endogenously expressed in a microorganism chosen from *Cronobacter* spp, *Callitropsis* spp, *Xanthophyllomyces* spp, *Agrobacterium* spp, and *Pantoea* spp.

98. The genetically modified *Candida viswanathii* yeast of claim 80, further comprising a genetic modification that alters the expression of one or more nucleic acids encoding one or more endogenous polypeptides.

99. The genetically modified *Candida viswanathii* yeast of claim 98, wherein the one or more endogenous polypeptides are chosen from polypeptides having one or more of the following activities: acyl-CoA synthetase activity, acyl-CoA oxidase activity, ATP-binding cassette transporter activity, carnitine acetyltransferase activity, transport of acetyl-carnitine, acyl-CoA thioesterase activity, acyl-CoA hydrolase activity, aldehyde dehydrogenase activity, monooxygenase activity, or monooxgenase reductase activity.

100. The genetically modified *Candida viswanathii* yeast of claim 80, wherein the amount and/or activity of aRas2 protein has been decreased.

101. The genetically modified *Candida viswanathii*yeast of claim 80, wherein the yeast has been genetically modified to reduce or eliminate expression of an endogenous *RAS2* gene.

102. The genetically modified *Candida viswanathii* yeast of claim 80, wherein the amount and/or activity of an Faa1 protein has been decreased.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1255 of 3077

103. The genetically modified *Candida viswanathii* yeast of claim 80, wherein the yeast has been genetically modified to reduce or eliminate expression of an endogenous *FAA1* gene.

104. The genetically modified *Candida viswanathii* yeast of claim 80, wherein the amounts and/or activities of one or more proteins chosen from acetyl-CoA C-acetyltransferase, HMG-CoA synthase, HMG-CoA reductase, mevalonate kinase, phosphomevalonate kinase, diphosphomevalonate decarboxylase, and isopentyl diphosphate delta isomerase, have been increased.

105. The genetically modified *Candida viswanathii* yeast of claim 80, wherein the yeast has been genetically modified to increase expression of one or more endogenous genes chosen from *ERG10*, *ERG13*, *HMG1*, *ERG12*, *ERG8*, *MVD1*, and *IDI1*.

106. The genetically modified *Candida viswanathii* yeast of claim 80, wherein the amount and/or activity of dimethylallyltranstransferase or farnesyl diphosphate synthetase has been increased.

107. The genetically modified *Candida viswanathii* yeast of claim 80, wherein the yeast has been genetically modified to increase expression of an endogenous *ERG20* gene.

108. The genetically modified *Candida viswanathii* yeast of claim 105, wherein expression of *ERG10*, *ERG13*, *HMG1*, *ERG12*, *ERG8*, *MVD1*, and/or *IDI1* is regulated by a promoter region of the gene encoding oleate-induced peroxisomal protein (POX18).

109. The genetically modified *Candida viswanathii* yeast of claim 105 or 108, wherein expression of *ERG10*, *ERG13*, *HMG1*, *ERG12*, *ERG8*, *MVD1*, and/or *IDI1*is regulated by a terminator region of the gene encoding oleate-induced peroxisomal protein (POX18).

110. The genetically modified *Candida viswanathii*yeast of claim 107, wherein expression of*ERG20* is regulated by a promoter region of the gene encoding oleate-induced peroxisomal protein (POX18).

111. The genetically modified *Candida viswanathii* yeast of claim 107 or 110, wherein expression of*ERG20* is regulated by a terminator region of the gene encoding oleate-induced peroxisomal protein (POX18).

112. A method for producing a terpene comprising:

contacting the genetically modified microorganism of any one of claims 1 to 41 with a feedstock comprising a carbon source, and

culturing the microorganism under conditions in which the terpenes are produced from the feedstock.

113. A method for producing a terpene comprising:

contacting the genetically modified microorganism of any one of claims 42 to 79 with a feedstock comprising a carbon source, and

culturing the microorganism under conditions in which the terpenes are produced from the feedstock.

114. A method for producing a terpene comprising:

contacting the genetically modified *Candida viswanathii*yeast of any one of claims 80 to 111 with a feedstock comprising a carbon source, and

culturing the microorganism under conditions in which the terpenes are produced from the feedstock.

115. The method of claim 112, wherein the feedstock comprises one or more fatty acids.

116. The method of claim 115, wherein the feedstock comprises one or more saturated fatty acids or one or more unsaturated fatty acids; or one or more saturated fatty acids and one or more unsaturated fatty acids.

117. The method of claim 116, wherein the feedstock comprises one or more fatty acids chosen from oleic acid, palmitoleic acid, erucic acid, linoleic acid, palmitic acid, caproic acid, enanthic acid, caprylic acid pelargonic acid, capric acid, undecylic acid, lauric acid, myristic acid, pentadecanoic acid, margaric acid, stearic acid arachidic acid, behenic acid, tridecylic acid, and linolenic acid.

118. The method of claim 117 wherein the feedstock comprises oleic acid.

119. The method of claim 112, wherein the feedstock comprises one or more sugars.

120. The method of 119, wherein the feedstock comprises glucose.

121. The method of claim 112, wherein the feedstock comprises one or more alkane hydrocarbons.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1257 of 3077

122. The method of claim 121, wherein the feedstock comprises one or more alkane hydrocarbons chosen from one or more ofhexane, heptane, nonane, decane, undecane, dodecane, tridecane, tetradecane, pentadecane, hexadecane, heptadecane, and octadecane.

123. The method of claim 121, wherein the feedstock comprises one or more alkane hydrocarbons chosen from C10 to C14 alkane hydrocarbons, nonane and octadecane.

124. The method of claim 112, wherein the feedstock comprises one or more vegetable oils, derivatives thereof, or byproducts thereof.

125. The method of claim 124, wherein the feedstock comprises crude palm oil (CPO).

126. The method of claim 124, wherein the feedstock comprises palm oil fatty acid distillate (PFAD).

127. The method of claim 112, wherein the terpene is chosen from lycopene, beta carotene and astaxanthin.

128. The method of claim 127, wherein the yield or titer of the terpene is between about 0.001 g/L to about 2.0 g/L.

129. The method of claim 127, wherein the yield or titer of the terpene is between about 0.05 g/L to about 0.5 g/L.

130. The method of claim 127, wherein the yield or titer of the terpene is between about 0.001 g/L to about 0.2 g/L.

131. The method of claim 128, wherein the yield or titer of lycopene is between about 0.03 g/L to about 2.0 g/L.

132. The method of claim 129, wherein the yield or titer of lycopene is between about 0.05 g/L to about 0.5 g/L.

133. The method of claim 130, wherein the yield or titer of lycopene is between about 0.001 g/L to about 0.2 g/L.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1258 of 3077

134. The method of claim 128, wherein the yield or titer of beta carotene is between about 0.1 g/L to about 0.2 g/L.

135. The method of claim 130, wherein the yield or titer of beta carotene is between about 0.01 g/L to about 0.1 g/L.

136. The method of claim 130, wherein the yield or titer of beta carotene is between about 0.001 g/L to about 0.01 g/L.

137. The method of claim 128, wherein the yield or titer of astaxanthin is between about 0.005 g/L to about 0.01 g/L.

138. The method of claim 130, wherein the yield or titer of astaxanthin is between about 0.001 g/L to about 0.005 g/L.

139. The method of claim 112, further comprising isolating the terpene.

140. A genetically modified yeast comprising one or more heterologous nucleic acids encodingphytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase, wherein the amount and/or activity of a Ras2 protein has been decreased.

141. A genetically modified yeast comprising one or more heterologous nucleic acids encodinggeranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase, wherein the amount and/or activity of a Ras2 protein has been decreased.

142. A genetically modified yeast comprising one or more heterologous nucleic acids encodinggeranylgeranyl diphosphate synthase, phytoene desaturase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase,and cytochrome p450 reductase, wherein the amount and/or activity of a Ras2 protein has been decreased.

143. The genetically modified yeast of claim 142, wherein the yeast has been genetically modified to reduce or eliminate expression of an endogenous *RAS2* gene.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1259 of 3077

144. The genetically modified yeast of claim 142, wherein the amount and/or activity of an Faa1 protein has been decreased.

145. The genetically modified yeast of claim 142, wherein the yeast has been genetically modified to reduce or eliminate expression of an endogenous *FAA1* gene.

146. The genetically modified yeast of claim 142, wherein the amounts and/or activities of one or more proteins chosen from acetyl-CoA C-acetyltransferase, HMG-CoA synthase, HMG-CoA reductase, mevalonate kinase, phosphomevalonate kinase, diphosphomevalonate decarboxylase, and isopentyl diphosphate delta isomerase,have been increased.

147. The genetically modified yeast of claim 142, wherein the yeast has been genetically modified to increase expression of one or more endogenous genes chosen from *ERG10*, *ERG13*, *HMG1*, *ERG12*, *ERG8*, *MVD1*, and *IDI1*.

148. The genetically modified yeast of claim 142, wherein the amount and/or activity of dimethylallyltranstransferase or farnesyl diphosphate synthetase has been increased.

149. The genetically modified yeast of claim 142, wherein the yeast has been genetically modified to increase expression of an endogenous *ERG20* gene.

150. The genetically modified yeast of claim 142, wherein the yeast is *Candida viswanathii*.

151. The genetically modified yeast of claim 142, wherein the yeast is *Blastobotrys adeninivorans*.

152. Use of the genetically modified yeast of any one of claims 140 to 151 for the production of a terpene.

153. Use of the genetically modified yeast of any one of claims 140 and 143-151 for the production of lycopene.

154. Use of the genetically modified yeast of claim 141 and 143-151 for the production of beta carotene.

155. Use of the genetically modified yeast of any one of claims 142-151 for the production of astaxanthin.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1260 of 3077

Atty. Docket No. SS118-1026

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application	:	Radici Chimica S.p.A.
International Application No.	:	PCT/US2018/41579
International Filing Date	:	11 July 2018
For	:	**BIOLOGICAL METHODS FOR PREPARING TERPENES**

ISA/US Mail Stop PCT
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450
Attn: ISA/US Authorized Officer Shane Thomas

RESPONSE TO WRITTEN OPINION

Dear Sir,

Reference is made to the Demand under Article 31 of the PCT for the above-captioned case. The following is in reply to the outstanding Written Opinion of the ISA dated September 13th, 2018.

Arguments are herein provided to overcome the objections raised in the above cited written opinion.

1. Amendments to the claims:

Claims 6-42, 49-82, 86-145 and 150-162 were believed not in accordance with PCT Rule 6.4(a), and were thus not examined.

A new set of claims 1-155 is provided to replace the currently pending set of claims 1-162. Claims 1-155 are respectfully believed to comply with PCT Rule 6.4(a).

Claim 1: has been amended to include the limitations of pending claim 5, and to specify that when the heterologous nucleic acid is cytochrome P450 reductase, it is regulated by a nucleic acid that provides for alkane induction.

Claims 2-4 have not been amended.

Claim 5 has been deleted.

Claims 6-41 have been renumbered and formally amended to correct dependency.

Claim 42 has been amended to include the limitations of claims 44 and 45.

Pending claims 44 and 45 have been deleted.

Claims 43-79 have been renumbered and formally amended to correct dependency.

Claim 80 has been amended to include the limitations of claims 84-86 and to specify that when the heterologous nucleic acid is cytochrome P450 reductase, it is regulated by a nucleic acid that provides for glucose or alkane induction.

Pending claims 84-86 have been deleted.

Claims 81-155 have been renumbered and formally amended to correct dependency.

Claim 149 has been deleted.

New claims 1-155 are respectfully believed not contain any additional subject matter with respect to the application as filed.

2. NOVELTY (Re Box No. V, 2)

Claims 1-4, 5/1-4 and 83-84 are believed to lack novelty under PCT Art 33(2) as being anticipated by WO 2014/100504 (hereinafter Verdezyne).

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1262 of 3077

Claim 1 has been amended by incorporating the limitations of claim 5, and is now directed to

"A genetically modified microorganism, comprising:

one or more heterologous nucleic acids encoding one or more terpene biosynthesis polypeptides, wherein expression of at least one of the heterologous nucleic acids is regulated by a nucleic acid that provides for fatty acid or alkane induction of expression of the terpene biosynthesis polypeptide:

- wherein the one or more heterologous nucleic acids that encode one or more terpene biosynthesis polypeptides is chosen from terpene synthase, phytoene synthase, geranylgeranyl diphosphate synthase, phytoene desaturase, lycopene cyclase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase, zeaxanthin glucosyltransferase and valencene synthase, or

- wherein the one or more heterologous nucleic acids is a cytochrome P450 reductase and is regulated by a nucleic acid that provides for alkane induction of expression of the cytochrome P450 reductase."

2.1 Novelty of claim 1 in view of Verdezyne

Verdezyne describes "biological methods for producing a fatty dicarboxylic acid and engineered microorganisms capable of such production" (abstract). Verdezyne describes a cytochrome P450 reductse polypeptide (Page 27, third paragraph) and POX4 promoters (page 41, 1st paragraph.).

Verdezyne does not describe microorganisms encoding a nucleic acid encoding one or more terpene biosynthesis polypeptides chosen from the group consisting of terpene synthase, phytoene synthase, geranylgeranyl diphosphate synthase, phytoene

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1263 of 3077

desaturase, lycopene cyclase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase, zeaxanthin glucosyltransferase and valencene synthase, nor does Verdezyne describe microorganisms encoding a cytochrome P450 reductase which is regulated by a nucleic acid that provides for alkane induction of expression.

Claim 1 is thus believed to be novel over Verdezyne.

Conclusions

The novelty objection raised in the outstanding Written Opinion is believed to be overcome in so far as the prior art document Verdezyne does not disclose the specific microorganisms described in amended claim 1.

Since pending claims 2-42 (new claims 2-41) depend on claim 1 and include all its features, new claims 2-41 are also believed to be novel over the prior art document Verdezyne.

2.2 Novelty of claim 42 in view of Verdezyne

Pending claims 44 and 45 are recognized as being novel (Re Box No.V) over the cited prior art.

New claim 42, which previously not examined, incorporates all the limitations of pending claims 44 and 45, and is thus also believed to be novel.

Conclusions

New claim 42 is respectfully believed to be novel over the cited prior art documents.

Since new claims 43-79 depend on claim 42 and include all its features, new claims 43-79 are also believed to be novel over the prior art documents.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1264 of 3077

2.3 Novelty of claim 83 in view of Verdezyne

The Examiner recognizes the novelty of pending claim 85 (Re Box No.V) over the cited prior art.

New claim 80 (pending claim 83) relates to a genetically modified *Candida viswanathii* yeast.

New claim 80 has been amended by incorporating all the limitations of pending claims 84-86.

New claim 80 is now believed to be novel over Verdezyne.

Conclusions

New claim 80 is respectfully believed to be novel over the cited prior art documents.

Since new claims 81-111 depend on claim 80 and include all its features, new claims 81-111 are also believed to be novel over the prior art documents.

2.4 Novelty of new claims 112 - 139

Claims 112 – 139 are method claims which all relate to a method for producing a terpene.

Independent claim 112 relates to a method for producing a terpene by contacting the genetically modified microorganism of claims 1-41 with a feedstock comprising a carbon source.

Since, as indicated in item 2.1, the microorganism of claims 1-41 is novel over the cited prior art, it is respectfully believed that the method for producing a terpene with the microorganism of claims 1-41 is also novel.

Independent claim 113 relates to a method for producing a terpene by contacting the genetically modified microorganism of claims 42-79 with a feedstock comprising a carbon source.

Since, as indicated in item 2.2, the microorganism of claims 42-79 is novel over the cited prior art, it is respectfully believed that the method for producing a terpene with the microorganism of claims 42-79 is also novel over the same prior art.

Independent claim 114 relates to a method for producing a terpene by contacting the genetically modified *Candida viswanathii* of claims 80-111 with a feedstock comprising a carbon source.

Since, as indicated in item 2.3, the microorganism of claims 80-111 is novel over the cited prior art, it is respectfully believed that the method for producing a terpene with the genetically modified *Candida viswanathii* of claims 80-111 is also novel over the same prior art.

Since claims 116-139 are all dependent on claim 112 and have all the limitations thereof, dependent claims 116-139 are also believed to be novel over the cited prior art.

2.5 Novelty of pending claims 146, 147 and 148 in view of Kirin

Pending claims 146-148 (new claims (140, 141 and 142) are believed to lack novelty in view of JP 2012-254044 (hereinafter Kirin).

The Examiner recognizes the novelty of pending claim 149 (Re Box No.V) over the cited prior art.

New claims 140-142 have been amended to include the limitations of pending claim 149 and are thus now believed to be novel over Kirin.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1266 of 3077

Conclusions

New claims 140-142 are respectfully believed to be novel over the cited prior art document.

Since new claims 143-151 depend on claim 142 and include all its features, new claims 143-151 are also believed to be novel over the prior art documents.

2.6 Novelty of new claims 152 - 155

Claims 152 – 155 are use claims which all relate to the use of a genetically modified yeast.

New independent claim 152 relates to the use of the genetically modified yeast of claims 140-151 for producing a terpene.

Since the genetically modified yeast of claims 140-151 is respectfully believed to be novel, as indicated in item 2.5 above, new claim 152 is also believed to be novel.

New independent claim 153 relates to the use of the genetically modified yeast of claims 140, 143-151 for producing a lycopene.

Since the genetically modified yeast of claims 140, 143-151 is respectfully believed to be novel, as indicated in item 2.5 above, new claim 153 is also believed to be novel.

New independent claim 154 relates to the use of the genetically modified yeast of claims 141, 143-151 for producing beta carotene.

Since the genetically modified yeast of claims 141, 143-151 is respectfully believed to be novel, as indicated in item 2.5 above, new claim 154 is also believed to be novel.

New independent claim 155 relates to the use of the genetically modified yeast of claims 142-151 for producing astaxanthin.

Since the genetically modified yeast of claims 142-151 is respectfully believed to be novel, as indicated in item 2.5 above, new claim 155 is also believed to be novel.

3. INVENTIVE STEP (Re Box No. V, 2)

According to the Examiner, the subject matter of claims 44, 46/44, 47/46/44, and 48/47/46/44 lack an inventive step as being obvious over Kirin in view of Verdezyne.

The Applicant respectfully disagrees and will discuss the reasoning in support of inventiveness of claims 1-155 below.

A. Independent claim 1

Kirin was taken to be the closest prior art, for the subject matter of claims 44, 46/44, 47/46/44, and 48/47/46/44.

Kirin describes the provision of a yeast strain capable of producing a metabolite through acetyl-CoA (abstract) and a process for the production of substances utilizing metabolic pathways via acetyl CoA in yeast (par. [0001]).

The subject matter of claim 1 differs from what is known through D1 by the distinguishing features of relating to a genetically modified microorganism, comprising a nucleic acid encoding a terpene biosynthesis polypeptide which is regulated by a nucleic acid that provides for fatty acid or alkane induction of expression.

The objective technical problem addressed by the present invention in the light of the closest prior art (Kirin) is that of obtaining an improved genetically modified microorganism which allows for the production of terpenes. This problem is solved by

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1268 of 3077

the present invention with results which are surprisingly very much higher and ameliorated than the expectations.

The invention of claim 1 solves this problem by providing the microorganism which is characterized by a specific combination of terpene biosynthesis polypeptides and nucleic acids which provide for the induction of expression of the terpene biosynthesis polypeptides.

This specific combination allows for the production of lycopene, B-carotene and astaxanthin, when the microorganisms are grown in the presence of specific substrates.

Evidence that the problem is solved can be found for example in Table 22 (page 467), which shows the results of the analysis of carotenoid production carried out with some of the genetically modified microorganisms.

Examples 67, 68 and 69 (pages 484-486) describe the construction of some of the most performing strains which produce lycopene from an oleic acid substrate. The performance of these strains is shown in Table 22, last 6 rows from the bottom.

There is no suggestion in Kirin that a yeast strain capable of producing a metabolite through acetyl-CoA could produce lycopene from oleic acid.

In fact, Kirin describes yeast strains for the production of isopropanol. As can be seen in Figures 1-7, the results of the analysis carried out in Kirin is that of obtaining a yeast strain which is capable of producing isopropanol when different genes are expressed. In particular, Figure 8 shows a graphical representation of the combination of genes which provides for the highest isopropanol production.

The Examiner indicates that Kirin discloses the microorganism of claim 43 but does not disclose "wherein expression of at least one of the heterologous nucleic acids is

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1269 of 3077

regulated by a nucleic acid that provides for fatty acid or alkane induction of expression of the terpene biosynthesis polypeptide".

The Examiner fails to indicate where Kirin discloses the microorganism of claim 43.

Paragraph [0046] of Kirin states that "it is possible to construct yeast strains that produce certain carotenoids, such as lycopene, specific sterols produced from squalene". In the same paragraph Kirin further states that "Candida utilis expressing a phytoene synthase gene, a phytoene desaturase gene and overexpressing an endogenous HMG-CoA reductase gene (HMG1) has been reported to highly accumulate lycopene (Shimada H, Kondo K, Fraser PD, Miura Y, Saito T, Misawa N. Increased carotenoid production by the food yeast Candida utilis through metabolic engineering of the isoprenoid pathway. Appl Environ Microbiol. 1998 Jul;64(7):2676 - 80)".

It is not clear where the microorganism of claim 43 can be found in the description of Kirin.

The skilled person, faced with the objective technical problem of the present invention, i.e. obtaining a genetically modified microorganism which allows for an improved terpene biosynthesis, and having in mind the disclosure of the Kirin document, would not arrive at the solution of claim 1.

In view of the above, the subject matter of new independent claim 1 is inventive over Kirin.

Combination of Kirin and Verdezyne

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1270 of 3077

Even if the skilled person, starting from Kirin and having in mind the teachings therein comprised, and faced with the objective technical problem described above, were to try to combine Kirin and Verdezyne, he would not arrive at the claimed solution.

Verdezyne describes biological methods and engineered microorganisms for producing a fatty dicarboxylic acid.

Verdezyne does not describe genetically modified microorganisms for the production of terpenes.

The invention of claim 1 is hence not obvious over Kirin or Verdezyne, either taken alone or in combination.

B. Independent claims 42, 80 and 140-142

Like claim 1, also independent claims 42, 80 and 140-142 describe genetically modified microorganisms, have the same distinguishing features of claim 1 and differ from Kirin and from Verdezyne for the same reasoning indicated for claim 1.

Independent claims 42, 80 and 140-142 are thus also believed to be inventive over the cited prior art documents, either taken alone or in combination.

C. Independent claims 112-114, and 152-155

Independent claims 112-114 and 152-155 relate to methods for producing and uses of the genetically modified microorganisms described in claims 1-41.

Independent claims 112-114 and 152-155 are thus also believed to be inventive over Kirin and Verdezyne, either taken alone or in combination.

D. Dependent claims 2-41, 43-79, 81-111, 115-139, 143-151

Since the dependent claims include all the features of the independent claims they depend on, they are respectfully believed to be inventive over the cited documents.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1271 of 3077

Atty. Docket No. SS118-1026

*****

It is respectfully believed that the present response provides basis for the issuance of a positive International Preliminary Report on Patentability of the instant application.

Should any additional items remain not fully attended, it is respectfully requested either the issuance of a second Written Opinion according to R.66.4(b) PCT explaining the Examiner's position, and a further opportunity for amendments and remarks to be filed.

Respectfully submitted

SILVIA SALVADORI, P.C.

June 10, 2019

By: /Silvia Salvadori/

Silvia Salvadori, Reg. No. 48,265
270 Madison Avenue, 8th Floor
New York, NY 10016
Phone: (212) 897-1938
Fax: (646) 219-6529

Attachments:

New claims 1-155 clean and marked-up version.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1272 of 3077

P021168WO_20190608_AMD claims marked up

What is claimed is:

1. A genetically modified microorganism, comprising:

one or more heterologous nucleic acids encoding one or more terpene biosynthesis polypeptides, wherein expression of at least one of the heterologous nucleic acids is regulated by a nucleic acid that provides for fatty acid or alkane induction of expression of the terpene biosynthesis polypeptide;

: wherein the one or more heterologous nucleic acids that encode one or more terpene biosynthesis polypeptides is chosen from terpene synthase, phytoene synthase, geranylgeranyl diphosphate synthase, phytoene desaturase, lycopene cyclase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase, zeaxanthin glucosyltransferase and valencene synthase, or

- wherein the one or more heterologous nucleic acids is a cytochrome P450 reductase and is regulated by a nucleic acid that provides for alkane induction of expression of the cytochrome P450 reductase.

2. The microorganism of claim 1, wherein the microorganism is a fungus.

3. The microorganism of claim 2, wherein the fungus is a yeast.

4. The microorganism of claim 3, wherein the yeast is chosen from *Candida* spp, *Yarrowia* spp, *Rhodotorula* spp, *Rhodosporidium* spp, *Cryptococcus* spp, *Trichosporon* spp, *Lipomyces* spp, and *Blastobotrys* spp.

~~5. The microorganism of any one of claims 1 to 4, wherein the one or more heterologous nucleic acids encode one or more terpene biosynthesis polypeptides chosen from terpene synthase, phytoene synthase, geranylgeranyl diphosphate synthase, phytoene desaturase, lycopene cyclase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase, zeaxanthin glucosyltransferase, valencene synthase, and cytochrome p450 reductase.~~

6. The microorganism ~~of any one~~ of claims ~~1 to 5~~, wherein the fatty acid is a saturated fatty acid or an unsaturated fatty acid.

7. The microorganism ~~of any one~~ of claims ~~1 to 6~~, wherein the fatty acid is chosen from one or more of oleic acid, palmitoleic acid, erucic acid, linoleic acid, palmitic acid, caproic acid, enanthic acid, caprylic acid pelargonic acid, capric acid, undecylic acid, lauric acid, myristic acid,

pentadecanoic acid, margaric acid, stearic acid arachidic acid, behenic acid, tridecylic acid, and linolenic acid.

87. The microorganism ~~of any one of~~ claims ~~1 to 54~~, wherein the alkane is chosen from one or more of hexane, heptane, nonane, decane, undecane, dodecane, tridecane, tetradecane, pentadecane, hexadecane, heptadecane, and octadecane

88. The microorganism ~~of any one of~~ claims ~~1 to 87~~, wherein the nucleic acid that provides for fatty acid or alkane induction of expression of a terpene biosynthesis polypeptide comprises a fatty acid response element or an alkane response element.

~~89~~. The microorganism of claim 88, wherein the fatty acid response element comprises an oleic acid response element.

~~1010~~. The microorganism of claim 88, wherein the alkane response element comprises an alkane response element 1 (ARE1) sequence.

~~1011~~. The microorganism ~~of any one of~~ ~~claimsclaim~~ ~~1 to 1010~~, wherein the nucleic acid that provides for fatty acid induction of expression of a terpene biosynthesis polypeptide comprises a promoter region chosen from promoter regions of genes encoding hydratase-dehydrogenase-epimerase (HDE), acyl co-A oxidase (POX), acyl co-A thiolase (POT), peroxin (PEX) and peroxisomal adenine nucleotide transporter protein (ANT1).

~~1012~~. The microorganism ~~of any one of~~ ~~claimsclaim~~ ~~1 to 1011~~, wherein the microorganism is *Candida viswanathii*.

~~1013~~. The microorganism ~~of any one of~~ ~~claimsclaim~~ ~~1 to 1011~~, wherein the microorganism is *Blastobotrys adeninivorans*.

~~1014~~. The microorganism ~~of any one of~~ ~~claimsclaim~~ ~~1 to 1013~~, wherein the one or more heterologous nucleic acids encode phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase.

~~1015~~. The microorganism of claim ~~1014~~, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a promoter region of the gene encoding a hydratase-dehydrogenase-epimerase (HDE).

~~17~~16. The microorganism of claim ~~15~~14, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a promoter region of the gene encoding a *Candida* hydratase-dehydrogenase-epimerase (HDE).

~~18~~17. The microorganism of claim ~~15, 16 or 17~~14, 15 or 16, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a terminator region of the gene encodingan acyl co-A oxidase 4 (POX4).

~~19~~18. The microorganism of claim 14, 15 or ~~16~~15, ~~16 or 17~~, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a terminator region of the gene encodinga *Candida*acyl co-A oxidase 4 (POX4).

~~20~~19. The microorganism ~~of any one of claim~~claim ~~1 to 14~~13, wherein the one or more heterologous nucleic acids encode valencene synthase.

~~21~~20. The microorganism of claim ~~20~~19, wherein expression of the heterologous nucleic acids encoding valencene synthase is regulated by a promoter region of the gene encoding a hydratase-dehydrogenase-epimerase (HDE).

~~22~~21. The microorganism of claim ~~20~~19, wherein expression of the heterologous nucleic acids encoding valencene synthase is regulated by a promoter region of the gene encoding a *Candida* hydratase-dehydrogenase-epimerase (HDE).

~~23~~22. The microorganism of claim 19, 20 or ~~21~~20, ~~21 or 22~~, wherein expression of the heterologous nucleic acids encoding valencene synthase is regulated by a terminator region of the gene encodingan acyl co-A oxidase 4 (POX4).

~~24~~23. The microorganism of claim ~~20, 21 or 22~~19, 20 or 21, wherein expression of the heterologous nucleic acids encoding valencene synthase is regulated by a terminator region of the gene encodinga *Candida*acyl co-A oxidase 4 (POX4).

P021168WO_20190608_AMD claims marked up

2524. The microorganism ~~of any one of claims~~claim ~~1 to 14~~13, wherein the one or more heterologous nucleic acids encode geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase.

2625. The microorganism of claim 2524, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase is regulated by a promoter region of the gene encoding a hydratase-dehydrogenase-epimerase (HDE).

2726. The microorganism of claim 2524, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase is regulated by a promoter region of the gene encoding a *Candida* hydratase-dehydrogenase-epimerase (HDE).

2827. The microorganism of claim ~~25, 26 or 27~~24, 25 or 26, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase is regulated by a terminator region of the gene encodingan acyl co-A oxidase 4 (POX4).

2928. The microorganism of claim 24, 25 or ~~26, 26 or 27~~, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase is regulated by a terminator region of the gene encodinga *Candida*acyl co-A oxidase 4 (POX4).

3029. The microorganism ~~of any one of claims~~claim ~~1 to 29~~28, wherein the one or more heterologous nucleic acids encoding the one or more terpene biosynthesis polypeptides are endogenously expressed in a microorganism chosen from *Cronobacter* spp, *Callitropsis* spp, *Xanthophyllomyces* spp, *Agrobacterium* spp, and *Pantoea* spp.

3130. The microorganism ~~of any one of claims~~claim ~~1 to 30~~29, wherein the amount and/or activity of aRas2 protein has been decreased.

3231. The microorganism ~~of any one of claims~~claim ~~1 to 31~~30, wherein the microorganism has been genetically modified to reduce or eliminate expression of an endogenous *RAS2* gene.

3332. The microorganism ~~of any one of claims~~claim ~~1 to 32~~31, wherein the amount and/or activity of an Faa1 protein has been decreased.

~~3433~~. The microorganism ~~of any one of claims~~claim ~~1 to 3432~~, wherein the microorganism has been genetically modified to reduce or eliminate expression of an endogenous *FAA1* gene.

~~3534~~. The microorganism ~~of any one of claims~~claim ~~1 to 3433~~, wherein the amounts and/or activities of one or more proteins chosen from acetyl-CoA C-acetyltransferase, HMG-CoA synthase, HMG-CoA reductase, mevalonate kinase, phosphomevalonate kinase, diphosphomevalonate decarboxylase, and isopentyl diphosphate delta isomerase,have been increased.

~~3635~~. The microorganism ~~of any one of claims~~claim ~~1 to 3634~~,wherein the microorganism has been genetically modified to increase expression of one or more endogenous genes chosen from *ERG10*, *ERG13*, *HMG1*, *ERG12*, *ERG8*, *MVD1*, and *IDI1*.

~~3736~~. The microorganism ~~of any one of claims~~claim ~~1 to 3635~~, wherein the amount and/or activity of dimethylallyltranstransferase or farnesyl diphosphate synthetase has been increased.

~~3837~~. The microorganism ~~of any one of claims~~claim ~~1 to 3736~~, wherein the microorganism has been genetically modified to increase expression of an endogenous *ERG20* gene.

~~3938~~. The microorganism of claim ~~3635~~, wherein expression of *ERG10*, *ERG13*, *HMG1*, *ERG12*, *ERG8*, *MVD1*, and/or *IDI1*is regulated by a promoter region of the gene encoding oleate-induced peroxisomal protein (POX18).

~~4039~~. The microorganism of claim ~~36 or 39~~35 or 38, wherein expression of *ERG10*, *ERG13*, *HMG1*, *ERG12*, *ERG8*, *MVD1*, and/or *IDI1*is regulated by a terminator region of the gene encoding oleate-induced peroxisomal protein (POX18).

~~4140~~. The microorganism of claim ~~3837~~, wherein expression of*ERG20*is regulated by a promoter region of the gene encoding oleate-induced peroxisomal protein (POX18).

~~4241~~. The microorganism of claim ~~38 or 41~~37 or 40, wherein expression of*ERG20*is regulated by a terminator region of the gene encoding oleate-induced peroxisomal protein (POX18).

~~4342~~. A genetically modified microorganism, comprising:

one or more heterologous nucleic acids encoding one or more terpene biosynthesis polypeptides, and

a genetic modification that alters the expression of a polypeptide providing for transport of acetyl-carnitine in the microorganism.

wherein the expression of at least one of the heterologous nucleic acids is regulated by a nucleic acid that provides for glucose, fatty acid or alkane induction of expression of the terpene biosynthesis polypeptide.

~~44. The microorganism of claim 43, wherein expression of at least one of the heterologous nucleic acids is regulated by a nucleic acid that provides for fatty acid or alkane induction of expression of the terpene biosynthesis polypeptide.~~

~~45. The microorganism of claim 43, wherein expression of at least one of the heterologous nucleic acids is regulated by a nucleic acid that provides for glucose induction of expression of the terpene biosynthesis polypeptide.~~

~~46~~43. The microorganism of claim ~~42~~43, ~~44 or 45~~, wherein the microorganism is a fungus.

~~47~~44. The microorganism of claim ~~43~~46, wherein the fungus is a yeast.

~~48~~45. The microorganism of claim ~~44~~47, wherein the yeast is chosen from *Candida* spp, *Yarrowia* spp, *Rhodotorula* spp, *Rhodosporidium* spp, *Cryptococcus* spp, *Trichosporon* spp, *Lipomyces* spp, and *Blastobotrys* spp.

~~49~~46. The microorganism ~~of any one of claims~~claim ~~42~~43 to ~~48~~, wherein the one or more heterologous nucleic acids encode one or more terpene biosynthesis polypeptides chosen from terpene synthase, phytoene synthase, geranylgeranyl diphosphate synthase, phytoene desaturase, lycopene cyclase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase, zeaxanthin glucosyltransferase, valencene synthase, and cytochrome p450 reductase.

~~50~~47. The microorganism ~~of any one of claims~~claim ~~46~~43 to ~~49~~, wherein the genetic modification reduces expression of the polypeptide providing for transport of acetyl-carnitine in the microorganism relative to a microorganism that does not have the genetic modification.

~~51~~48. The microorganism of claim ~~47~~50, wherein the genetic modification is a disruption, deletion or knockout of (i) a polynucleotide that encodes a polypeptide providing for transport of acetyl-carnitine, or (ii) a promoter operably linked to a polynucleotide that encodes a polypeptide

providing for transport of acetyl-carnitine, whereby endogenous activity of a polypeptide providing for transport of acetyl-carnitine is reduced or abolished.

6248. The microorganism of claim 4760, further comprising a nucleic acid encoding a polypeptide providing for transport of acetyl-carnitine wherein expression of the nucleic acid encoding a polypeptide providing for transport of acetyl-carnitine is regulated by a promoter that provides for reduced expression relative to endogenous expression.

6350. The microorganism of claim 4962, wherein the genetic modification comprises replacing the promoter of an endogenous gene encoding the polypeptide providing for transport of acetyl-carnitine in the microorganism with a promoter that provides for reduced expression of the polypeptide in the microorganism relative to a microorganism that does not have the genetic modification.

6451. The microorganism of claim ~~62 or 6349 or 50~~, wherein the promoter that provides for reduced expression relative to endogenous expression is a promoter for a glucose-6-phosphate isomerase (G6PI) gene.

6652. The microorganism of claim ~~49 or 5063 or 63~~, wherein the promoter that provides for reduced expression relative to endogenous expression is a promoter for a *Candida* glucose-6-phosphate isomerase (G6PI) gene.

6853. The microorganism ~~of any one of~~ claim ~~43 to 6642~~, wherein the polypeptide providing for transport of acetyl-carnitine in the microorganism is an acetyl-carnitine translocase (CRC).

6754. The microorganism of claim 5356, wherein the acetyl-carnitine translocase (CRC) is acetyl-carnitine translocase 1 (CRC1).

6855. The microorganism ~~of any one of~~ claim ~~43 to 6742~~, wherein the microorganism is *Candida viswanathii*.

6856. The microorganism ~~of any one of~~ claim ~~43 to 6742~~, wherein the microorganism is *Blastobotrys adeninivorans*.

6957. The microorganism ~~of any one of~~ claim ~~43 to 6942~~, wherein the one or more heterologous nucleic acids encode phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase.

~~6458~~. The microorganism of claim ~~5760~~, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a promoter region of the gene encoding a hydratase-dehydrogenase-epimerase (HDE).

~~6259~~. The microorganism of claim ~~5760~~, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a promoter region of the gene encoding a *Candida* hydratase-dehydrogenase-epimerase (HDE).

~~6360~~. The microorganism of claim ~~60, 61 or 6257, 58 or 59~~, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a terminator region of the gene encodingan acyl co-A oxidase 4 (POX4).

~~6461~~. The microorganism of claim ~~57, 58 or 5960, 61 or 62~~, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a terminator region of the gene encodinga *Candida*acyl co-A oxidase 4 (POX4).

~~6562~~. The microorganism ~~of any one of claimsclaim 4243 to 59~~, wherein the one or more heterologous nucleic acids encode geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase.

~~6663~~. The microorganism of claim ~~6265~~, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase is regulated by a promoter region of the gene encoding a hydratase-dehydrogenase-epimerase (HDE).

~~6764~~. The microorganism of claim ~~6265~~, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase is regulated by a promoter region of the gene encoding a *Candida* hydratase-dehydrogenase-epimerase (HDE).

~~6865~~. The microorganism of clairn ~~62, 63 or 6465, 66 or 67~~, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase

and bifunctional lycopene cyclase/phytoene synthase is regulated by a terminator region of the gene encodingan acyl co-A oxidase 4 (POX4).

6966. The microorganism of claim 62, 63 or 6465, 66 or 67, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase is regulated by a terminator region of the gene encodinga *Candida*acyl co-A oxidase 4 (POX4).

7067. The microorganism of any one of claimsclaim 4243 to 69, wherein the one or more heterologous nucleic acids encoding the one or more terpene biosynthesis polypeptides are endogenously expressed in a microorganism chosen from *Cronobacter* spp, *Callitropsis* spp, *Xanthophyllomyces* spp, *Agrobacterium* spp, and *Pantoea* spp.

7168. The microorganism of any one of claimsclaim 4243 to 70, wherein the amount and/or activity of aRas2 protein has been decreased.

7269. The microorganism of any one of claimsclaim 4243 to 71, wherein the microorganism has been genetically modified to reduce or eliminate expression of an endogenous *RAS2* gene.

7370. The microorganism of any one of claimsclaim 4243 to 72, wherein the amount and/or activity of an Faa1 protein has been decreased.

7471. The microorganism of any one of claimsclaim 4243 to 73, wherein the microorganism has been genetically modified to reduce or eliminate expression of an endogenous *FAA1* gene.

7572. The microorganism of any one of claimsclaim 4243 to 74, wherein the amounts and/or activities of one or more proteins chosen from acetyl-CoA C-acetyltransferase, HMG-CoA synthase, HMG-CoA reductase, mevalonate kinase, phosphomevalonate kinase, diphosphomevalonate decarboxylase, and isopentyl diphosphate delta isomerase, have been increased.

7673. The microorganism of any one of claimsclaim 4243 to 75, wherein the microorganism has been genetically modified to increase expression of one or more endogenous genes chosen from *ERG10*, *ERG13*, *HMG1*, *ERG12*, *ERG8*, *MVD1*, and *IDI1*.

7774. The microorganism of any one of claimsclaim 4243 to 76, wherein the amount and/or activity of dimethylallyltranstransferase or farnesyl diphosphate synthetase has been increased.

~~75~~. The microorganism ~~of any one of claims~~claim ~~42~~43 ~~to 77~~, wherein the microorganism has been genetically modified to increase expression of an endogenous *ERG20* gene.

~~76~~. The microorganism of claim ~~73~~74, wherein expression of *ERG10*, *ERG13*, *HMG1*, *ERG12*, *ERG8*, *MVD1*, and/or *IDI1*is regulated by a promoter region of the gene encoding oleate-induced peroxisomal protein (POX18).

~~77~~. The microorganism of claim ~~73 or 76~~74 ~~or 76~~, wherein expression of *ERG10*, *ERG13*, *HMG1*, *ERG12*, *ERG8*, *MVD1*, and/or *IDI1*is regulated by a terminator region of the gene encoding oleate-induced peroxisomal protein (POX18).

~~78~~. The microorganism of claim ~~75~~76, wherein expression of*ERG20*is regulated by a promoter region of the gene encoding oleate-induced peroxisomal protein (POX18).

~~79~~. The microorganism of claim ~~75 or 78~~76 ~~or 78~~, wherein expression of*ERG20*is regulated by a terminator region of the gene encoding oleate-induced peroxisomal protein (POX18).

~~80~~. A genetically modified *Candida viswanathii* yeast, comprising one or more heterologous nucleic acids encoding one or more terpene biosynthesis polypeptides, wherein expression of at least one of the heterologous nucleic acids is regulated by a nucleic acid that provides for glucose, fatty acid or alkane induction of expression of the terpene biosynthesis polypeptide and

, wherein the one or more heterologous nucleic acids encoding one or more terpene biosynthesis polypeptides is chosen from terpene synthase, phytoene synthase, geranylgeranyl diphosphate synthase, phytoene desaturase, lycopene cyclase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase, zeaxanthin glucosyltransferase and valencene synthase, or

- wherein the one or more heterologous nucleic acids is a cytochrome P450 reductase and is regulated by a nucleic acid that provides for glucose or alkane induction of expression of the cytochrome P450 reductase.

~~81. The genetically modified *Candida viswanathii* yeast of claim 80, wherein expression of at least one of the heterologous nucleic acids is regulated by a nucleic acid that provides for fatty acid or alkane induction of expression of the terpene biosynthesis polypeptide.~~

P021168WO_20190608_AMD claims marked up

~~85. The genetically modified *Candida viswanathii* yeast of claim 83, wherein expression of at least one of the heterologous nucleic acids is regulated by a nucleic acid that provides for glucose induction of expression of the terpene biosynthesis polypeptide.~~

~~86. The genetically modified *Candida viswanathii* yeast of claim83, 84 or 85, wherein the one or more heterologous nucleic acids encode one or more terpene biosynthesis polypeptides chosen from terpene synthase, phytoene synthase, geranylgeranyl diphosphate synthase, phytoene desaturase, lycopene cyclase,bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase, zeaxanthin glucosyltransferase, valencene synthase, and cytochrome p450 reductase.~~

8781. The genetically modified *Candida viswanathii* yeast ~~of any one of claimsclaim 8083 to86~~, wherein the one or more heterologous nucleic acids encode phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase

8882. The genetically modified *Candida viswanathii* yeast of claim 8187, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a promoter region of the gene encoding a *Candida* hydratase-dehydrogenase-epimerase (HDE).

8983. The genetically modified *Candida viswanathii* yeast of claim 8288, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a terminator region of the gene encoding a *Candida*acyl co-A oxidase 4 (POX4).

9084. The genetically modified *Candida viswanathii* yeast of claim 8187, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a promoter region of the gene encoding a *Candida* glyceraldehyde-3-phosphate dehydrogenase (GPD).

9185. The genetically modified *Candida viswanathii* yeast ~~of any one of claimsclaim 8083 to86~~, wherein the one or more heterologous nucleic acids encode valencene synthase.

9286. The genetically modified *Candida viswanathii* yeast of claim 8591, wherein expression of the heterologous nucleic acids encoding valencene synthase is regulated by a promoter region of the gene encoding a *Candida* hydratase-dehydrogenase-epimerase (HDE).

P021168WO_20190608_AMD claims marked up

8387. The genetically modified *Candida viswanathii* yeast of claim 8682, wherein expression of the heterologous nucleic acids valencene synthase is regulated by a terminator region of the gene encoding a *Candida*acyl co-A oxidase 4 (POX4).

8488. The genetically modified *Candida viswanathii* yeast of claim 8584, wherein expression of the heterologous nucleic acids encoding valencene synthase is regulated by a promoter region of the gene encoding a *Candida* glyceraldehyde-3-phosphate dehydrogenase (GPD).

8589. The genetically modified *Candida viswanathii* yeast ~~of any one~~ of ~~claims~~claim 8083-to86, wherein the one or more heterologous nucleic acids encode geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase.

8690. The genetically modified *Candida viswanathii* yeast of claim 8085, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase is regulated by a promoter region of the gene encoding a *Candida* hydratase-dehydrogenase-epimerase (HDE).

8791. The genetically modified *Candida viswanathii* yeast of claim 8086, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase is regulated by a terminator region of the gene encoding a *Candida*acyl co-A oxidase 4 (POX4).

8892. The genetically modified *Candida viswanathii* yeast of claim 8085, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase is regulated by a promoter region of the gene encoding a *Candida* glyceraldehyde-3-phosphate dehydrogenase (GPD).

8993. The genetically modified *Candida viswanathii* yeast ~~of any one~~ of ~~claims~~claim 8083-to86, wherein the one or more heterologous nucleic acids encode geranylgeranyl diphosphate synthase, phytoene desaturase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase,and cytochrome p450 reductase.

10094. The genetically modified *Candida viswanathii* yeast of claim 9300, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase,and cytochrome p450 reductase is regulated by a promoter region of the gene encoding a *Candida* hydratase-dehydrogenase-epimerase (HDE).

P021168WO_20190608_AMD claims marked up

~~104~~95. The genetically modified *Candida viswanathii* yeast of claim ~~94~~100, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase,and cytochrome p450 reductase is regulated by a terminator region of the gene encoding a *Candida*acyl co-A oxidase 4 (POX4).

~~104~~96. The genetically modified *Candida viswanathii* yeast of claim ~~93~~95, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase,and cytochrome p450 reductase is regulated by a promoter region of the gene encoding a *Candida* glyceraldehyde-3-phosphate dehydrogenase (GPD).

~~103~~97. The genetically modified *Candida viswanathii* yeast ~~of any one of claims~~claim ~~80~~83 to ~~102~~, wherein the one or more heterologous nucleic acids encoding the one or more terpene biosynthesis polypeptides are endogenously expressed in a microorganism chosen from *Cronobacter* spp, *Callitropsis* spp, *Xanthophyllomyces* spp, *Agrobacterium* spp, and *Pantoea* spp.

~~104~~98. The genetically modified *Candida viswanathii* yeast ~~of any one of claims~~claim ~~80~~83 to ~~103~~, further comprising a genetic modification that alters the expression of one or more nucleic acids encoding one or more endogenous polypeptides.

~~106~~99. The genetically modified *Candida viswanathii* yeast of claim ~~98~~104, wherein the one or more endogenous polypeptides are chosen from polypeptides having one or more of the following activities: acyl-CoA synthetase activity, acyl-CoA oxidase activity, ATP-binding cassette transporter activity, carnitine acetyltransferase activity, transport of acetyl-carnitine, acyl-CoA thioesterase activity, acyl-CoA hydrolase activity, aldehyde dehydrogenase activity, monooxygenase activity, or monooxgenase reductase activity.

~~106~~100. The genetically modified *Candida viswanathii* yeast ~~of any one of claims~~claim ~~80~~83 to ~~105~~, wherein the amount and/or activity of aRas2 protein has been decreased.

~~107~~101. The genetically modified *Candida viswanathii*yeast ~~of any one of claims~~claim ~~80~~83 to ~~106~~, wherein the yeast has been genetically modified to reduce or eliminate expression of an endogenous *RAS2* gene.

P021168WO_20190608_AMD claims marked up

~~98~~102. The genetically modified *Candida viswanathii* yeast ~~of any one of claims~~claim ~~80~~83 ~~to 97~~, wherein the amount and/or activity of an Faa1 protein has been decreased.

~~99~~103. The genetically modified *Candida viswanathii* yeast ~~of any one of claims~~claim ~~80~~83 ~~to 98~~, wherein the yeast has been genetically modified to reduce or eliminate expression of an endogenous *FAA1* gene.

~~100~~104. The genetically modified *Candida viswanathii* yeast ~~of any one of claims~~claim ~~80~~83 ~~to 99~~, wherein the amounts and/or activities of one or more proteins chosen from acetyl-CoA C-acetyltransferase, HMG-CoA synthase, HMG-CoA reductase, mevalonate kinase, phosphomevalonate kinase, diphosphomevalonate decarboxylase, and isopentyl diphosphate delta isomerase, have been increased.

~~101~~105. The genetically modified *Candida viswanathii* yeast ~~of any one of claims~~claim ~~80~~83 ~~to 100~~, wherein the yeast has been genetically modified to increase expression of one or more endogenous genes chosen from *ERG10*, *ERG13*, *HMG1*, *ERG12*, *ERG8*, *MVD1*, and *IDI1*.

~~102~~106. The genetically modified *Candida viswanathii* yeast ~~of any one of claims~~claim ~~80~~83 ~~to 101~~, wherein the amount and/or activity of dimethylallyltranstransferase or farnesyl diphosphate synthetase has been increased.

~~103~~107. The genetically modified *Candida viswanathii* yeast ~~of any one of claims~~claim ~~80~~83 ~~to 102~~, wherein the yeast has been genetically modified to increase expression of an endogenous *ERG20* gene.

~~104~~108. The genetically modified *Candida viswanathii* yeast of claim ~~101~~105, wherein expression of *ERG10*, *ERG13*, *HMG1*, *ERG12*, *ERG8*, *MVD1*, and/or *IDI1* is regulated by a promoter region of the gene encoding oleate-induced peroxisomal protein (POX18).

~~105~~109. The genetically modified *Candida viswanathii* yeast of claim 105 or ~~108~~~~101 or 104~~, wherein expression of *ERG10*, *ERG13*, *HMG1*, *ERG12*, *ERG8*, *MVD1*, and/or *IDI1*is regulated by a terminator region of the gene encoding oleate-induced peroxisomal protein (POX18).

~~106~~110. The genetically modified *Candida viswanathii*yeast of claim 107~~103~~, wherein expression of*ERG20* is regulated by a promoter region of the gene encoding oleate-induced peroxisomal protein (POX18).

P021168WO_20190608_AMD claims marked up

~~117~~111. The genetically modified *Candida viswanathii* yeast of claim 107 or 110~~113 or 116~~, wherein expression of *ERG20* is regulated by a terminator region of the gene encoding oleate-induced peroxisomal protein (POX18).

~~118~~112. A method for producing a terpene comprising:

contacting the genetically modified microorganism of any one of claims 1 to ~~42~~42 with a feedstock comprising a carbon source, and

culturing the microorganism under conditions in which the terpenes are produced from the feedstock.

~~118~~113. A method for producing a terpene comprising:

contacting the genetically modified microorganism of any one of claims ~~42~~42 to 79~~43 to 82~~ with a feedstock comprising a carbon source, and

culturing the microorganism under conditions in which the terpenes are produced from the feedstock.

~~120~~114. A method for producing a terpene comprising:

contacting the genetically modified *Candida viswanathii*yeast of any one of claims ~~80~~80 to 111~~83 to 117~~ with a feedstock comprising a carbon source, and

culturing the microorganism under conditions in which the terpenes are produced from the feedstock.

~~124~~115. The method ~~of any one of~~ claimsclaim 112~~118 to 120~~, wherein the feedstock comprises one or more fatty acids.

~~122~~116. The method of claim 115~~121~~, wherein the feedstock comprises one or more saturated fatty acids or one or more unsaturated fatty acids; or one or more saturated fatty acids and one or more unsaturated fatty acids.

~~123~~117. The method of claim 116~~121 or 122~~, wherein the feedstock comprises one or more fatty acids chosen from oleic acid, palmitoleic acid, erucic acid, linoleic acid, palmitic acid, caproic acid, enanthic acid, caprylic acid pelargonic acid, capric acid, undecylic acid, lauric acid, myristic acid, pentadecanoic acid, margaric acid, stearic acid arachidic acid, behenic acid, tridecylic acid, and linolenic acid.

~~124~~118. The method of claim 117~~123~~ wherein the feedstock comprises oleic acid.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1287 of 3077

125119. The method ~~of any one of~~ claimsclaim 112~~118 to 120~~, wherein the feedstock comprises one or more sugars.

126120. The method of 119~~125~~, wherein the feedstock comprises glucose.

127121. The method ~~of any one of~~ claimsclaim 112~~118 to 120~~, wherein the feedstock comprises one or more alkane hydrocarbons.

128122. The method of claim 121~~127~~, wherein the feedstock comprises one or more alkane hydrocarbons chosen from one or more ofhexane, heptane, nonane, decane, undecane, dodecane, tridecane, tetradecane, pentadecane, hexadecane, heptadecane, and octadecane.

129123. The method of claim 121~~127~~, wherein the feedstock comprises one or more alkane hydrocarbons chosen from C10 to C14 alkane hydrocarbons, nonane and octadecane.

130124. The method ~~of any one of~~ claimsclaim 112~~118 to 120~~, wherein the feedstock comprises one or more vegetable oils, derivatives thereof, or byproducts thereof.

131125. The method of claim 124~~130~~, wherein the feedstock comprises crude palm oil (CPO).

132126. The method of claim 124~~130~~, wherein the feedstock comprises palm oil fatty acid distillate (PFAD).

133127. The method ~~of any one of~~ claimsclaim 112~~118 to 132~~, wherein the terpene is chosen from lycopene, beta carotene and astaxanthin.

134128. The method ~~of any one of~~ claimsclaim 127~~118 to 133~~, wherein the yield or titer of the terpene is between about 0.001 g/L to about 2.0 g/L.

135129. The method ~~of any one of~~ claimsclaim 127~~118 to 133~~, wherein the yield or titer of the terpene is between about 0.05 g/L to about 0.5 g/L.

136130. The method ~~of any one of~~ claimsclaim 127~~118 to 133~~, wherein the yield or titer of the terpene is between about 0.001 g/L to about 0.2 g/L.

137131. The method of claim 128~~134~~, wherein the yield or titer of lycopene is between about 0.03 g/L to about 2.0 g/L.

~~130~~132. The method of claim 129~~130~~, wherein the yield or titer of lycopene is between about 0.05 g/L to about 0.5 g/L.

~~131~~133. The method of claim 130~~131~~, wherein the yield or titer of lycopene is between about 0.001 g/L to about 0.2 g/L.

~~132~~134. The method of claim 128~~129~~, wherein the yield or titer of beta carotene is between about 0.1 g/L to about 0.2 g/L.

~~133~~135. The method of claim 130~~131~~, wherein the yield or titer of beta carotene is between about 0.01 g/L to about 0.1 g/L.

~~134~~136. The method of claim 130~~131~~, wherein the yield or titer of beta carotene is between about 0.001 g/L to about 0.01 g/L.

137~~143~~. The method of claim 128~~129~~, wherein the yield or titer of astaxanthin is between about 0.005 g/L to about 0.01 g/L.

~~144~~138. The method of claim ~~136~~130, wherein the yield or titer of astaxanthin is between about 0.001 g/L to about 0.005 g/L.

~~145~~139. The method ~~of any one of~~ claim~~s~~ 112~~113 to 144~~, further comprising isolating the terpene.

~~146~~140. A genetically modified yeast comprising one or more heterologous nucleic acids encodingphytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase, wherein the amount and/or activity of a Ras2 protein has been decreased.

~~147~~141. A genetically modified yeast comprising one or more heterologous nucleic acids encodinggeranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase, wherein the amount and/or activity of a Ras2 protein has been decreased.

~~148~~142. A genetically modified yeast comprising one or more heterologous nucleic acids encodinggeranylgeranyl diphosphate synthase, phytoene desaturase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin

synthase,and cytochrome p450 reductase, wherein the amount and/or activity of a Ras2 protein has been decreased.

~~148. The genetically modified yeast of claim146, 147 or 148, wherein the amount and/or activity of aRas2 protein has been decreased.~~

~~160~~143. The genetically modified yeast ~~of any one of claims~~claim 142~~146 to 149~~, wherein the yeast has been genetically modified to reduce or eliminate expression of an endogenous *RAS2* gene.

~~151~~144. The genetically modified yeast ~~of any one of claims~~claim 142~~146 to 150~~, wherein the amount and/or activity of an Faa1 protein has been decreased.

~~152~~145. The genetically modified yeast ~~of any one of claims~~claim 142~~146 to 151~~, wherein the yeast has been genetically modified to reduce or eliminate expression of an endogenous *FAA1* gene.

~~153~~146. The genetically modified yeast ~~of any one of claims~~claim 142~~146 to 152~~, wherein the amounts and/or activities of one or more proteins chosen from acetyl-CoA C-acetyltransferase, HMG-CoA synthase, HMG-CoA reductase, mevalonate kinase, phosphomevalonate kinase, diphosphomevalonate decarboxylase, and isopentyl diphosphate delta isomerase,have been increased.

~~154~~147. The genetically modified yeast ~~of any one of claims~~claim 142~~146 to 153~~, wherein the yeast has been genetically modified to increase expression of one or more endogenous genes chosen from *ERG10*, *ERG13*, *HMG1*, *ERG12*, *ERG8*, *MVD1*, and *IDI1*.

~~155~~148. The genetically modified yeast ~~of any one of claims~~claim 142~~146 to 154~~, wherein the amount and/or activity of dimethylallyltranstransferase or farnesyl diphosphate synthetase has been increased.

~~156~~149. The genetically modified yeast ~~of any one of claims~~claim 142~~146 to 155~~, wherein the yeast has been genetically modified to increase expression of an endogenous *ERG20* gene.

~~157~~150. The genetically modified yeast ~~of any one of claims~~claim 142~~146 to 156~~, wherein the yeast is *Candida viswanathii*.

P021168WO_20190608_AMD claims marked up

~~158~~151. The genetically modified yeast ~~of any one of~~ claim~~claim~~ ~~142~~140 ~~to 156~~, wherein the yeast is *Blastobotrys adeninivorans*.

~~159~~152. Use of the genetically modified yeast of any one of claims ~~146 to 158~~140 to 151 for the production of a terpene.

~~160~~153. Use of the genetically modified yeast of any one of claims ~~140~~146 and ~~140 to 159~~143-151 for the production of lycopene.

~~161~~154. Use of the genetically modified yeast ~~of any one of~~ claim~~claim~~ ~~141~~147 and ~~140 to 159~~143-151 for the production of beta carotene.

~~162~~155. Use of the genetically modified yeast of any one of claims ~~140 to 159~~142-151 for the production of astaxanthin.

ADVANCE E-MAIL

From the INTERNATIONAL BUREAU

PCT

NOTIFICATION OF THE RECORDING OF A CHANGE

(PCT Rule 92*bis*.1 and Administrative Instructions, Section 422)

To:

SALVADORI, Silvia
Silvia Salvadori, P.C.
270 Madison Avenue, 8th Floor
New York, NY 10016
ÉTATS-UNIS D'AMÉRIQUE

Date of mailing *(day/month/year)*
15 May 2019 (15.05.2019)

Applicant's or agent's file reference
SS118-1026

IMPORTANT NOTIFICATION

International application No.
PCT/US2018/041579

International filing date *(day/month/year)*
11 July 2018 (11.07.2018)

1. The following indications appeared on record concerning:

☐ the applicant ☐ the inventor ☒ the agent ☐ the common representative

Name and Address
DICKINSON, Kari A.
Grant IP, Inc.
C/o CPA Global
900 Second Avenue S., Suite 600
Minneapolis, MN 55402
United States of America

State of Nationality | State of Residence

Telephone No.
858-412-1084

Facsimile No.
858-623-3224

E-mail address
docketing@cpaglobal.com

2. The International Bureau hereby notifies the applicant that the following change has been recorded concerning:

☒ the person ☐ the name ☐ the address ☐ the nationality ☐ the residence

Name and Address
SALVADORI, Silvia
Silvia Salvadori, P.C.
270 Madison Avenue, 8th Floor
New York, NY 10016
United States of America

State of Nationality | State of Residence

Telephone No.
212-897-1938

Facsimile No.
646-219-6529

E-mail address
silvia@salvadorilaw.com
☒ Notifications by e-mail authorized

3. Further observations, if necessary:
The new agent's file reference is: SS118-1026

4. A copy of this notification has been sent to:

☒ the receiving Office
☐ the International Searching Authority
☐ the Authority(ies) specified for supplementary search
☐ the International Preliminary Examining Authority
☒ the designated Offices concerned
☐ the elected Offices concerned
☒ other: DICKINSON, Kari A.

The International Bureau of WIPO
34, chemin des Colombettes
1211 Geneva 20, Switzerland

Authorized officer
Chahine Leila
e-mail pct.team4@wipo.int
Telephone No. +41 22 338 74 04

Form PCT/IB/306 (January 2009) 1/HUS4DSGPN2ETA0

ADVANCE E-MAIL

From the INTERNATIONAL BUREAU

PCT

NOTIFICATION OF THE RECORDING OF A CHANGE

(PCT Rule 92*bis*.1 and Administrative Instructions, Section 422)

To:

SALVADORI, Silvia
Silvia Salvadori, P.C.
270 Madison Avenue, 8th Floor
New York, NY 10016
ÉTATS-UNIS D'AMÉRIQUE

Date of mailing *(day/month/year)*
07 January 2020 (07.01.2020)

Applicant's or agent's file reference
SS118-1026

IMPORTANT NOTIFICATION

International application No.
PCT/US2018/041579

International filing date *(day/month/year)*
11 July 2018 (11.07.2018)

1. The following indications appeared on record concerning:

☐ the applicant ☒ the inventor ☐ the agent ☐ the common representative

Name and Address

State of Nationality | State of Residence

Telephone No.

Facsimile No.

E-mail address

2. The International Bureau hereby notifies the applicant that the following change has been recorded concerning:

☒ the person ☐ the name ☐ the address ☐ the nationality ☐ the residence

Name and Address
SCOTT, Erin
1579 Caudor Street
Encinitas, CA 92024
United States of America

State of Nationality | State of Residence

Telephone No.

Facsimile No.

E-mail address
☐ Notifications by e-mail authorized

3. Further observations, if necessary:
The person identified in Box 2 has been added as inventor only.

4. A copy of this notification has been sent to:

☒ the receiving Office
☐ the International Searching Authority
☐ the Authority(ies) specified for supplementary search
☐ the International Preliminary Examining Authority
☐ the designated Offices concerned
☒ the elected Offices concerned
☐ other:

The International Bureau of WIPO
34, chemin des Colombettes
1211 Geneva 20, Switzerland

Authorized officer
Chahine Leila
e-mail pct.team4@wipo.int
Telephone No. +41 22 338 74 04

Form PCT/IB/306 (January 2009) 1/SZFMR6WT332M50

ADVANCE E-MAIL

From the INTERNATIONAL BUREAU

PCT

NOTIFICATION OF THE RECORDING OF A CHANGE

(PCT Rule 92*bis*.1 and Administrative Instructions, Section 422)

To:

SALVADORI, Silvia
Silvia Salvadori, P.C.
270 Madison Avenue, 8th Floor
New York, NY 10016
ÉTATS-UNIS D'AMÉRIQUE

Date of mailing *(day/month/year)*
07 January 2020 (07.01.2020)

Applicant's or agent's file reference
SS118-1026

IMPORTANT NOTIFICATION

International application No.
PCT/US2018/041579

International filing date *(day/month/year)*
11 July 2018 (11.07.2018)

1. The following indications appeared on record concerning:

☐ the applicant ☒ the inventor ☐ the agent ☐ the common representative

Name and Address

State of Nationality | State of Residence

Telephone No.

Facsimile No.

E-mail address

2. The International Bureau hereby notifies the applicant that the following change has been recorded concerning:

☒ the person ☐ the name ☐ the address ☐ the nationality ☐ the residence

Name and Address
BEARDSLEE, Thomas
6827 Carnation Drive
Carlsbad, CA 92011
United States of America

State of Nationality | State of Residence

Telephone No.

Facsimile No.

E-mail address
☐ Notifications by e-mail authorized

3. Further observations, if necessary:
The person identified in Box 2 has been added as inventor only.

4. A copy of this notification has been sent to:

☒ the receiving Office
☐ the International Searching Authority
☐ the Authority(ies) specified for supplementary search
☐ the International Preliminary Examining Authority
☐ the designated Offices concerned
☒ the elected Offices concerned
☐ other:

The International Bureau of WIPO
34, chemin des Colombettes
1211 Geneva 20, Switzerland

Authorized officer
Chahine Leila
e-mail pct.team4@wipo.int
Telephone No. +41 22 338 74 04

Form PCT/IB/306 (January 2009) 1/NVM9TUPVGXTNW0

ADVANCE E-MAIL

From the INTERNATIONAL BUREAU

PCT

NOTIFICATION OF THE RECORDING OF A CHANGE

(PCT Rule 92*bis*.1 and Administrative Instructions, Section 422)

To:

SALVADORI, Silvia
Silvia Salvadori, P.C.
270 Madison Avenue, 8th Floor
New York, NY 10016
ÉTATS-UNIS D'AMÉRIQUE

Date of mailing *(day/month/year)*
07 January 2020 (07.01.2020)

Applicant's or agent's file reference
SS118-1026

IMPORTANT NOTIFICATION

International application No.
PCT/US2018/041579

International filing date *(day/month/year)*
11 July 2018 (11.07.2018)

1. The following indications appeared on record concerning:

☐ the applicant ☒ the inventor ☐ the agent ☐ the common representative

Name and Address

State of Nationality | State of Residence

Telephone No.

Facsimile No.

E-mail address

2. The International Bureau hereby notifies the applicant that the following change has been recorded concerning:

☒ the person ☐ the name ☐ the address ☐ the nationality ☐ the residence

Name and Address
WALBRIDGE, Michael
4 Diane Dr.
Saugus, MA 01906
United States of America

State of Nationality | State of Residence

Telephone No.

Facsimile No.

E-mail address
☐ Notifications by e-mail authorized

3. Further observations, if necessary:
The person identified in Box 2 has been added as inventor only.

4. A copy of this notification has been sent to:

☒ the receiving Office
☐ the International Searching Authority
☐ the Authority(ies) specified for supplementary search
☐ the International Preliminary Examining Authority
☐ the designated Offices concerned
☒ the elected Offices concerned
☐ other:

The International Bureau of WIPO
34, chemin des Colombettes
1211 Geneva 20, Switzerland

Authorized officer
Chahine Leila
e-mail pct.team4@wipo.int
Telephone No. +41 22 338 74 04

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1295 of 3077

ADVANCE E-MAIL

From the INTERNATIONAL BUREAU

PCT

NOTIFICATION OF THE RECORDING OF A CHANGE

(PCT Rule 92*bis*.1 and Administrative Instructions, Section 422)

To:

SALVADORI, Silvia
Silvia Salvadori, P.C.
270 Madison Avenue, 8th Floor
New York, NY 10016
ÉTATS-UNIS D'AMÉRIQUE

Date of mailing *(day/month/year)*
07 January 2020 (07.01.2020)

Applicant's or agent's file reference
SS118-1026

IMPORTANT NOTIFICATION

International application No.
PCT/US2018/041579

International filing date *(day/month/year)*
11 July 2018 (11.07.2018)

1. The following indications appeared on record concerning:

☐ the applicant ☒ the inventor ☐ the agent ☐ the common representative

Name and Address

State of Nationality | State of Residence

Telephone No.

Facsimile No.

E-mail address

2. The International Bureau hereby notifies the applicant that the following change has been recorded concerning:

☒ the person ☐ the name ☐ the address ☐ the nationality ☐ the residence

Name and Address
YI, Jian
13313 Thunderhead Street
San Diego, CA 92129
United States of America

State of Nationality | State of Residence

Telephone No.

Facsimile No.

E-mail address
☐ Notifications by e-mail authorized

3. Further observations, if necessary:
The person identified in Box 2 has been added as inventor only.

4. A copy of this notification has been sent to:

☒ the receiving Office
☐ the International Searching Authority
☐ the Authority(ies) specified for supplementary search
☐ the International Preliminary Examining Authority
☐ the designated Offices concerned
☒ the elected Offices concerned
☐ other:

The International Bureau of WIPO
34, chemin des Colombettes
1211 Geneva 20, Switzerland

Authorized officer
Chahine Leila
e-mail pct.team4@wipo.int
Telephone No. +41 22 338 74 04

Form PCT/IB/306 (January 2009) 1/RJD34N3SALYAR0

BIOLOGICAL METHODS FOR PREPARING TERPENES

Related Patent Applications

This patent application claims the benefit of U.S. provisional patent application no. 62/532,297, filed on July 13, 2017, entitled BIOLOGICAL METHODS FOR PREPARING TERPENES, naming Kimberly Ann Aeling as inventor, and designated by Attorney Docket No. VRD-3002-PV. This patent application is related to an International patent application, filed simultaneously herewith, entitled BIOLOGICAL METHODS FOR MODIFYING CELLULAR CARBON FLUX, naming Tom Beardslee as inventor, and designated by Attorney Docket No. VAL-3001-PC, which claims the benefit of U.S. provisional patent application no. 62/532,292, filed on July 13, 2017, entitled BIOLOGICAL METHODS FOR MODIFYING CELLULAR CARBON FLUX, naming Tom Beardslee as inventor, and designated by Attorney Docket No. VRD-3001-PV. This patent application also is related to U.S. provisional patent application no. 61/222,902 filed on July 2, 2009, entitled BIOLOGICAL METHODS FOR PREPARING ADIPIC ACID, naming Stephen Picataggio as inventor, and designated by Attorney Docket No. VRD-1001-PV. This patent application also is related to International patent application no. PCT/US2010/040837 filed on July 1, 2010, entitled BIOLOGICAL METHODS FOR PREPARING ADIPIC ACID, naming Stephen Picataggio and Tom Beardslee as inventors, and designated by Attorney Docket No. VRD-1001-PC. This patent application also is related to U.S. provisional patent application no. 61/430,097 filed on January 5, 2011, entitled BIOLOGICAL METHODS FOR PREPARING ADIPIC ACID, naming Stephen Picataggio and Tom Beardslee as inventors, and designated by Attorney Docket No. VRD-1001-PV2. This patent application also is related to U.S. provisional patent application no. 61/482,160 filed on May 3, 2011, entitled BIOLOGICAL METHODS FOR PREPARING ADIPIC ACID, naming Stephen Picataggio and Tom Beardslee as inventors, and designated by Attorney Docket No. VRD-1001-PV3. This patent application also is related to U.S. patent application no. 13/245,777 filed on September 26, 2011, entitled BIOLOGICAL METHODS FOR PREPARING ADIPIC ACID, naming Stephen Picataggio and Tom Beardslee as inventors, and designated by Attorney Docket No. VRD-1001-CT. This patent application also is related to U.S. patent application no. 13/245,780 filed on September 26, 2011, entitled BIOLOGICAL METHODS FOR PREPARING ADIPIC ACID, naming Stephen Picataggio and Tom Beardslee as inventors, and designated by Attorney Docket No. VRD-1001-UT. This patent application also is related to U.S. patent application no. 13/245,782 filed on September 26, 2011, entitled BIOLOGICAL METHODS FOR PREPARING ADIPIC ACID,

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1297 of 3077

naming Stephen Picataggio and Tom Beardslee as inventors, and designated by Attorney Docket No. VRD-1001-UT2. This patent application also is related to International patent application no. PCT/US2012/020230 filed on January 4, 2012, entitled BIOLOGICAL METHODS FOR PREPARING ADIPIC ACID, naming Stephen Picataggio and Tom Beardslee as inventors, and designated by Attorney Docket No. VRD-1001-PC2. This patent application also is related to International patent application no. PCT/US2012/056562 filed on September 21, 2012, entitled BIOLOGICAL METHODS FOR PREPARING ADIPIC ACID, naming Stephen Picataggio and Tom Beardslee as inventors, and designated by Attorney Docket No. VRD-1001-PC3. This patent application also is related to U.S. provisional patent application no. 62/136,350 filed on March 20, 2015, entitled BIOLOGICAL METHODS FOR PREPARING 3-HYDROXYPROPIONIC ACID, naming Eric Michael Knight as inventor, and designated by Attorney Docket No. VRD-2001-PV. This patent application also is related to International patent application no. PCT/US2016/023243 filed on March 18, 2016, entitled BIOLOGICAL METHODS FOR PREPARING 3-HYDROXYPROPIONIC ACID, naming Eric Michael Knight as inventor. This patent application also is related to U.S. provisional patent application no. 61/505,092 filed on July 6, 2011, entitled BIOLOGICAL METHODS FOR PREPARING SEBACIC ACID naming Stephen Picataggio and Tom Beardslee as inventors, and designated by Attorney Docket No. VRD-1005-PV. This patent application also is related to U.S. provisional patent application no. 61/523,216 filed August 12, 2011, entitled BIOLOGICAL METHODS FOR PREPARING DODECANEDIOIC ACID naming Stephen Picataggio and Tom Beardslee as inventors, and designated by Attorney Docket No. VRD-1006-PV. This patent application also is related to International patent application no. PCT/US2012/045615 filed on July 5, 2012, entitled BIOLOGICAL METHODS FOR PREPARING A FATTY DICARBOXYLIC ACID, naming Tom Beardslee, Stephen Picataggio, L. Dudley Eirich and Jose Miguel Laplaza as inventors, and designated by Attorney Docket No. VRD-1005-PC. This patent application also is related to International patent application no. PCT/US2012/045622 filed on July 5, 2012, entitled BIOLOGICAL METHODS FOR PREPARING A FATTY DICARBOXYLIC ACID, naming Tom Beardslee, Stephen Picataggio, Alex Hutagalung and Tom Fahland as inventors, and designated by Attorney Docket No. VRD-1006-PC. This patent application also is related to U.S. patent application no. 14/131,170 filed on April 14, 2014 entitled BIOLOGICAL METHODS FOR PREPARING A FATTY DICARBOXYLIC ACID, naming Tom Beardslee, Stephen Picataggio, L. Dudley Eirich and Jose Miguel Laplaza as inventors. This patent application also is related to U.S. patent application no. 14/131,174 filed on April 28, 2014, entitled BIOLOGICAL METHODS FOR PREPARING A FATTY DICARBOXYLIC ACID, naming Tom Beardslee,

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1298 of 3077

Stephen Picataggio, Alex Hutagalung and Tom Fahland as inventors. This patent application also is related to U.S. provisional patent application no. 61/739,656 filed December 19, 2012, entitled BIOLOGICAL METHODS FOR PREPARING A FATTY DICARBOXYLIC ACID naming Jose Laplaza, Tom Beardslee, Dudley Eirich and Stephen Picataggio as inventors, and designated by Attorney Docket No. VRD-1007-PV. This patent application also is related to U.S. provisional patent application no. 61/739,661 filed December 19, 2012, entitled BIOLOGICAL METHODS FOR PREPARING A FATTY DICARBOXYLIC ACID naming Tom Beardslee, Alex Hutagalung and Stephen Picataggio as inventors, and designated by Attorney Docket No. VRD-1008-PV. This patent application also is related to International patent application no. PCT/US2013/076664 filed on December 19, 2013, entitled BIOLOGICAL METHODS FOR PREPARING A FATTY DICARBOXYLIC ACID, naming Jose Laplaza, Tom Beardslee, Dudley Eirich and Stephen Picataggio as inventors. This patent application also is related to International patent application no. PCT/US2013/076739 filed on December 19, 2013, entitled BIOLOGICAL METHODS FOR PREPARING A FATTY DICARBOXYLIC ACID, naming Tom Beardslee, Alex Hutagalung and Stephen Picataggio as inventors. This patent application also is related to U.S. patent application no. 14/654,442 filed on June 19, 2015 entitled BIOLOGICAL METHODS FOR PREPARING A FATTY DICARBOXYLIC ACID, naming Jose Laplaza, Tom Beardslee, Dudley Eirich and Stephen Picataggio as inventors. This patent application also is related to U.S. patent application no. 14/654,458 filed on June 19, 2015, entitled BIOLOGICAL METHODS FOR PREPARING A FATTY DICARBOXYLIC ACID, naming Tom Beardslee, Alex Hutagalung and Stephen Picataggio as inventors. This patent application is also related to U.S. provisional patent application no. 62/011,500 filed on June 12, 2014 entitled PURIFICATION OF POLYCARBOXYLIC ACIDS, naming Jose Laplaza as inventor. This patent application is also related to U.S. patent application no. 14/738,600 filed on June 12, 2015 entitled PURIFICATION OF POLYCARBOXYLIC ACIDS, naming Jose Laplaza, William Andrew Evanko and Jason H. Radany as inventors. This patent application also is related to International patent application no. PCT/US2015/035634 filed June 12, 2015 entitled PURIFICATION OF POLYCARBOXYLIC ACIDS, naming Jose Laplaza as inventor. The entire content of each of the foregoing patent applications is incorporated herein by reference, including, without limitation, all text, tables and drawings.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1299 of 3077

Field

The technology relates in part to biological methods for producing terpenes and to engineered cells and microorganisms capable of such production.

Background

Cells and microorganisms employ various enzyme-driven biological pathways to support metabolism and growth. A cell synthesizes native proteins, including enzymes, in vivo based on the sequence of deoxyribonucleic acid (DNA) encoding the protein. DNA first is transcribed into a complementary ribonucleic acid (RNA) that contains a ribonucleotide sequence encoding the protein. RNA then directs translation of the encoded protein by interaction with various cellular components, such as ribosomes. When the resulting protein is an enzyme, it can participate as a biological catalyst in biochemical pathways involved in producing a variety of organic molecules by the cell or organism.

These pathways can be exploited for the harvesting of naturally produced organic molecules. The pathways also can be altered to increase production or to produce specific molecules that may be commercially valuable. Advances in recombinant molecular biology methodology allow researchers to isolate DNA from one cell or organism and insert it into another cell or organism, thus altering the cellular synthesis of enzymes or other proteins. Advances in recombinant molecular biology methodology also allow endogenous genes, carried in the genomic DNA of a cell or microorganism, to be increased in copy number, thus altering the cellular synthesis of enzymes or other proteins. Such genetic engineering can change the biological pathways within the host cell or organism, causing it to produce a desired product. Microorganic industrial production can minimize the use of caustic chemicals and the production of toxic byproducts, thus providing a “clean” source for certain compounds. The use of appropriate plant-derived feedstocks allows production of “green” compounds while further minimizing the need for and use of petroleum-derived compounds.

Summary

Provided herein in certain aspects are genetically modified microorganisms comprising one or more heterologous nucleic acids encoding one or more terpene biosynthesis polypeptides,

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1300 of 3077

where expression of at least one of the heterologous nucleic acids is regulated by a nucleic acid that provides for fatty acid or alkane induction of expression of the terpene biosynthesis polypeptide. Also provided herein in certain aspects are genetically modified microorganisms comprising one or more heterologous nucleic acids encoding one or more terpene biosynthesis polypeptides, and a genetic modification that alters the expression of a polypeptide providing for transport of acetyl-carnitine in the microorganisms. Also provided herein in certain aspects are methods for producing a terpene comprising contacting a genetically modified microorganism provided herein with a feedstock comprising a carbon source, and culturing the microorganism under conditions in which the terpenes are produced from the feedstock.

Also provided herein in certain aspects is a genetically modified *Candida viswanathii* yeast, comprising one or more heterologous nucleic acids encoding one or more terpene biosynthesis polypeptides. Also provided herein in certain aspects are methods for producing a terpene comprising contacting a genetically modified *Candida viswanathii* yeast provided herein with a feedstock comprising a carbon source, and culturing the microorganism under conditions in which the terpenes are produced from the feedstock.

Also provided in certain aspects is a genetically modified yeast comprising one or more heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase. Also provided in certain aspects is a genetically modified yeast comprising one or more heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase. Also provided in certain aspects is a genetically modified yeast comprising one or more heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase, and cytochrome p450 reductase.

Certain embodiments are described further in the following description, examples, claims and drawings.

Incorporation by Reference

All publications, patents and patent applications, GENBANK sequences (e.g., available at the World Wide Web Uniform Resource Locator (URL) ncbi.nlm.nih.gov of the National Center for

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1301 of 3077

Biotechnology Information (NCBI), sequences available through other databases, and websites and other published materials referred to throughout the entire disclosure herein, unless noted otherwise, are incorporated by reference in their entirety. Citation of any publications, patents and patent applications, GENBANK (and other database) sequences, websites and other published materials herein is not an admission that any of the foregoing is pertinent prior art, nor does it constitute any admission as to the contents or date of these publications or documents.

Brief Description of the Drawings

The drawings illustrate certain embodiments of the technology and are not limiting. For clarity and ease of illustration, the drawings are not made to scale and, in some instances, various aspects may be shown exaggerated or enlarged to facilitate an understanding of particular embodiments.

Fig. 1 is a diagrammatic representation of a cassette for the addition of a gene of interest (GOI) into a host non-functional ura3 locus using the single crossover integration method. The core of the cassette contains the GOI gene with a promoter (P_{XXX}) and terminator (T_{XXX}) for controlling transcription of the GOI gene. The *URA3* gene selectable marker is split with a 3' portion of the gene at one end of the cassette and a 5' portion positioned at the other end of the cassette. The segment of the expression cassette containing the gene of interest (GOI) is positioned between the *URA3* promoter (P_{URA3}) and terminator (T_{URA3}). Parts (A) and (B) of Fig. 1 show results of integration of one copy (A) and two copies (B) of the cassette. Integration of one cassette generates an added, functional GOI expression unit and may or may not provide for expression of a functional Ura3p, depending on the nature of the ura3 locus and the location of the split in the *URA3* selectable marker. Integration of two copies of the cassette generates a complete, functional *URA3* sequence by combining the 5' end of *URA3* from one copy of the cassette and the 3' end of *URA3* from the second copy of the cassette. Additional copies may also be integrated. Transformants are selected for by growth on uracil-free media. This integration method thus favors selection of transformants containing multiple copies of the GOI.

Fig. 2 is a diagrammatic illustration of an exemplary gene cassette for use in a double crossover homologous recombination integration referred to as "knock out" mutagenesis. Two slightly different cassettes are depicted for use in separately disrupting each of the two *FAT1* alleles in a diploid yeast. The two cassette-containing nucleic acid segments are referred to as "Deletion

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1302 of 3077

1" and "Deletion 2," respectively. Each cassette contains a *URA3* gene including a URA3 promoter (P_{URA3}) and terminator (T_{URA3}). Additionally, each cassette contains a repeat of the P_{URA3} sequence immediately downstream of the terminator sequence. The two separate deletion cassette-containing fragments differ in the sequences of the target gene that they contain on each side of the *URA3* cassette.

Fig. 3A, Fig. 3B, and Fig. 3C show diagrammatic illustrations of a "knock in" gene disruption method which disrupts one target gene ("GOI1") and also adds a desired gene of interest ("GOI2") at the disrupted locus. The basic URA3 disruption cassette is the same as that described in Fig. 2, except for an additional expression cassette immediately downstream of the second P_{URA3} repeat sequence. This expression cassette contains the gene of interest, GOI2, for adding to the endogenous GOI1 locus and includes a promoter (P_{XXX}) and terminator (T_{XXX}) for controlling transcription of GOI2. Immediately upstream of the first P_{URA3} sequence is a sequence of nucleotides of the GOI1 gene, and immediately downstream of the terminator (T_{XXX}) for GOI2 is another sequence of the GOI1 gene. These sequences are for use in integration of the cassettes into the GOI1 locus. Fig. 3B shows the locus after the integration of the knock-in cassette which incorporates the functional URA3 selection gene. To remove the *URA3* gene, transformants are grown in the presence of 5-FOA to facilitate a "loop-out" event that is driven by the direct repeat sequences on either side of the *URA3* gene (in this case P_{URA3}). The result of that event is shown in Fig. 3C which depicts the P_{URA3} sequence that remains followed by the functional GOI2 cassette.

Fig. 4 is a schematic illustration of the general, unmodified flow of carbon from a fatty acid carbon source in a wild-type eukaryotic cell, such as, for example, a yeast cell. "FA": fatty acid; "Co-A": coenzyme A; "PL": phospholipid; "TAG": triacylglyceride; FA-CoA": fatty acyl-CoA "*FAA1*" and "*FAT1*": acyl-CoA synthetase genes; "*PEX11*": peroxisomal biogenesis factor gene; "*PXA1*": peroxisomal transport protein gene; "β-Ox": β-oxidation; "Ac-CoA": acetyl-CoA; "*CAT2*": carnitine acetyltransferase gene; "Carn": carnitine; "Ac-Carn": acetyl-carnitine; "*CRC*1": mitochondrial acetyl-carnitine transport protein; "Cit": citrate; "TCA": tricarboxylic acid cycle; "Iso-Cit": isocitrate; "*ICL1*": isocitrate lyase gene; "Succ": succinate; "*MLS1*": malate synthase gene; "Glx": glyoxylate; "Mal": malate.

Fig. 5 is a schematic illustration of an engineered carbon flux pathway of a modified cell for use in producing a target molecule. The figure depicts cellular modifications in some embodiments

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1303 of 3077

of a eukaryotic (i.e., yeast in this example) platform system for developing particular target molecule production systems. The platform system contains an acetyl group carbon recycle loop that diverts acetyl moieties generated in the breakdown of fatty acids in peroxisomal β-oxidation("β-ox") into cytosolic fatty acid synthesis to regenerate a fatty acid that can be subjected to another cycle of peroxisomal β-oxidation. The recycle loop is depicted by the dark, solid reaction arrows beginning with extracellular fatty acid ("FA") internalization in the upper left corner of the figure. Free fatty acids that have entered the cell can undergo oxidation to dicarboxylic acids (DCA) through ω-oxidation ("ω-ox"). Multiple modifications introduced via genetic manipulation, as well as unmodified activities of the cell, are indicated as follows: acyl-CoA synthetase gene deletions shown as "faa1Δ" and "fat1Δ" and resulting disruption of cytosolic activation of fatty acids (indicated as a lightly shaded dotted line reaction arrow below the gene deletion symbols and extending from "FA" to "FA-CoA") and diminished entry of FA-CoA into lipid (triacylglycerides ("TAG") and phospholipids ("PL")) biosynthesis; endogenous, unmodified peroxisomal enzymes acyl-CoA synthetase ("FAA2") and thioesterase ("TES"); unmodified glyoxylate cycle ("GlyOx") showing endogenous isocitrate lyase enzyme ("ICL1"); unmodified endogenous peroxisomal carnitine acetyltransferase ("CAT2") for conversion of acetyl-CoA ("Ac-CoA") to acetyl-carnitine ("AC-Carn"); modified (indicated by diagonal hatch lines) cytosolic carnitine acetyltransferase ("$CAT2^{cyt}$") and acetyl-CoA carboxylase ("ACC1") enzymes; unmodified endogenous fatty acid synthase enzyme complex ("FAS"); modified and added cytosolic thioesterase enzyme ("TES^{cyt}") showing added activity as solid, dark reaction arrow extending from FA-CoA to FA (which represents the final segment of the recycle loop); modified (gene deletion) peroxisomal transport protein ("pxa1Δ") showing disrupted (lightly shaded dotted line reaction arrow) acyl-CoA ("FA-CoA") import activity; modified peroxisomal biogenesis factor ("PEX11") activity; modified (promoter replacement) mitochondrial acetyl-carnitine transport protein ("CRC1") showing diminished (lightly shaded dotted line reaction arrow) acetyl-carnitine import activity; modified mitochondrial carnitine acetyltransferase ("CAT2") activity showing decreased conversion (lightly shaded dotted line reaction arrow) of AC-Carn to AC-CoA; unmodified mitochondrial tricarboxylic acid cycle ("TCA"); lightly shaded dashed lines reflect unmodified cellular activities that are not part of the carbon recycle loop shown in dark, solid lines. The details of the modifications in this exemplary engineered platform system are provided in the Detailed Description that follows.

Fig. 6 is a schematic illustration of an engineered carbon flux pathway of a modified cell for use in producing a target molecule. The figure depicts cellular modifications in some embodiments

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1304 of 3077

of a eukaryotic (i.e., yeast in this example) platform system for developing particular target molecule production systems. The platform system is similar to that shown in Fig. 5 except for the following: modified (gene deletion) peroxisomal carnitine acetyltransferase ("cat2Δ") showing disrupted (lightly shaded dotted line reaction arrow) generation of peroxisomal acetyl-carnitine (AC-Carn shown with an "X" over it); modified and added peroxisomal acetyl-CoA hydrolase for converting acetyl-CoA (AC-CoA) to acetate ("Ac"); modified (promoter replacement) cytosolic acetyl-CoA synthetase ("ACS") activity. The details of the modifications in this exemplary engineered platform system are provided in the Detailed Description that follows.

Fig. 7 is a schematic illustration of an engineered carbon flux pathway of a modified cell for use in producing a target molecule. The figure depicts cellular modifications in some embodiments of a eukaryotic (i.e., yeast in this example) platform system for developing particular target molecule production systems. The carbon recycle loop in this platform system, depicted by the dark, solid reaction arrows, extends through mitochondrial metabolism and differs from that shown in Figs. 5 and 6. The mitochondrial acetyl-carnitine transporter ("CRC1") and carnitine acetytransferase ("CAT2") are unmodified in this exemplary platform system. A cytosolic ATP citrate lyase ("ACL1/2") activity is added to the system. The details of the modifications in this exemplary engineered platform system are provided in the Detailed Description that follows.

Fig. 8 is a schematic illustration of an engineered carbon flux pathway of a modified cell for use in producing a target molecule. The figure depicts cellular modifications in some embodiments of a eukaryotic (i.e., yeast in this example) platform system for for the enhanced production of malonyl-CoA and various target molecules that can be synthesized using malonyl-CoA as a precursor. The details of the modifications in this exemplary engineered platform system are provided in the Detailed Description that follows.

Fig. 9 is a schematic illustration of an example of an engineered production pathway for cell- or microbial-based synthesis of 3-hydroxypropionic acid ("3HP"). Added cytosolic malonyl-CoA reductase ("MCR") activity and modified 3-hydroxy-propionate-dehydrogenase ("HPD1") activities for 3HP synthesis are shown as well as modified (gene deleted) endogenous semialdehyde dehydrogenase ("ald6Δ") activity. The details of the modifications in this exemplary engineered platform system are provided in the Detailed Description that follows.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1305 of 3077

Fig. 10 is a schematic illustration of an example of an engineered production pathway for cell- or microbial-based synthesis of triacetic acid lactone ("TAL"). Added 2-pyrone synthase ("2PS") activity for TAL synthesis is shown. The details of the modifications in this exemplary engineered platform system are provided in the Detailed Description that follows.

Fig. 11 is a schematic illustration depicting cellular modifications in some embodiments of a eukaryotic (i.e., yeast in this example) platform system for the enhanced generation of acetyl-CoA and the production of a diverse array of target molecules (e.g., terpenes). In one aspect, Fig. 11 differs from Fig. 8 in that it shows an embodiment of the platform system in which target molecule production pathways extend from acetyl-CoA, instead of malonyl-CoA , as a precursor molecule. ("Mev" refers to the mevalonate pathway; "IPP" refers to isopentenyl diphosphate; "DMAPP" refers to dimethylallyl diphosphate.) The details of the modifications in this exemplary engineered platform system are provided in the Detailed Description that follows.

Fig. 12 is a restriction endonuclease site map of plasmid pAA061 showing the relative placement of the following nucleic acid sequences: *Candida* strain ATCC 20336 orotidine-5'-phosphate decarboxylase (*URA3*) gene promoter (Prom), open-reading frame and terminator (Term); β-lactamase (ampicillin-resistance) gene promoter (P(BLA)) and ORF (AP^r); and the *Escherichia coli* origin of replication (ORI). Also shown are the *Candida* strain ATCC 20336 phosphoglycerate kinase (*PGK*) gene promoter and terminator that were added to pAA061 to form pAA105.

Fig. 13 is a restriction endonuclease site map of plasmid pAA105 which was constructed by ligating the *Candida* strain ATCC 20336 phosphoglycerate kinase (*PGK*) gene promoter and terminator with the PstI/NdeI fragment of pAA061 (Fig. 12).

Fig. 14 is a restriction endonuclease site map of plasmid pAA219 which was constructed by inserting the *Candida* strain ATCC 20336 cytochrome P450 reductase (*CPRB*) ORF between the *PGK* gene promoter and terminator in pAA105 (Fig. 13).

Fig. 15 is a restriction endonuclease site map of a PstI/NdeI fragment of plasmid pAA073 which contains the *Candida* strain ATCC 20336 acyl-CoA oxidase (*POX4*) gene promoter and terminator with restriction sites between them for incorporating ORFs to be controlled by the inducible *POX4* promoter.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1306 of 3077

Fig. 16 is a restriction endonuclease site map of a fragment of plasmid pAA073. Plasmid pAA073 was constructed by ligating the PstI/NdeI fragment shown in Fig. 15 with the PstI/NdeI fragment of pAA061 (Fig. 12).

Fig. 17 is a restriction endonuclease site map of plasmid pAA153 which was constructed by inserting the *Candida* strain ATCC 20336 cytochrome P450 monooxygenase (*CYP52A14*) ORF between the *POX4* gene promoter and terminator in pAA073.

Fig. 18 is a diagrammatic representation of pAA153 (Fig. 17) linearized by endonuclease cutting of the plasmid at the *Cla*I site to yield a cassette for use in the addition of a *Candida* strain ATCC 20336 *CYP52A14* gene into a host non-functional ura3 locus using the single crossover integration method. The core of the cassette contains the *CYP52A14* gene with a *POX4* promoter and terminator for controlling transcription of the gene. Cutting of the plasmid at the *Cla*I site splits the *URA3* selectable marker and yields a linear DNA fragment with the *CYP52A14* gene expression cassette positioned between the *URA3* promoter (URA3 Prom) and terminator (URA3 Term).

Fig. 19 is a diagrammatic representation of plasmid pAA367 generated by (1) PCR amplification of two separate fragments of pAA153 (Fig. 17), one fragment containing a 3' URA3 sequence and the URA3 terminator and another fragment containing a *CYP52A14* gene expression cassette with a *POX4* promoter and terminator followed by the URA3 promoter and a 5' URA3 sequence, (2) joining of the two amplicons by overlap extension PCR to generate a single amplified fragment and (3) cloning of the single fragment into pCR-BluntII-TOPO.

Fig. 20 is a diagrammatic representation of a linear DNA expression cassette obtained by amplification from pAA367 (Fig. 19) that does not contain nucleic acid encoding an antibiotic selection marker (i.e., antibiotic-free).

Fig. 21A, Fig. 21B, and Fig. 21C show diagrammatic illustrations of a "knock out" gene disruption method which disrupts a target gene ("GOI"). Fig. 21A shows a double-crossover gene knock-out cassette for knocking out the function of a GOI containing a *URA3* selectable marker gene (including the gene promoter (P_{URA3}) and terminator (T_{URA3})) between 5' and 3' homologous sequences for the GOI. The *URA3* selectable marker also has DNA sequence

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1307 of 3077

direct repeats (P_{URA3}) at the beginning and at the end of the gene sequence. After transformation of the double-crossover gene knock-out cassette into a Ura$^-$ mutant, the *URA3* marker allows selection on SC-URA plates for colonies that have integrated the construct (Fig. 21B) disrupting the GOI and generating a Ura$^+$ phenotype. Subsequent growth of Ura$^+$ transformants on 5-fluoroorotic acid (5-FOA) yields Ura$^-$ cells resulting from removal of the *URA3* selectable marker from the genome by a second crossover homologous recombination between the DNA sequence direct repeats (P_{URA3}) (Fig. 21C). A DNA sequence direct repeat remains in the genome as a "scar" left behind at the gene knock out site. The *URA3* selection marker may now be used again for further genetic modifications.

Fig. 22 is a diagrammatic depiction of how the knock out gene disruption method illustrated in Figs. 21A-21C, which regenerates an auxotrophic (Ura$^-$) cell after the second homologous recombination event, enables the same *URA3*-based selection method to be used repeatedly on the same cell, for example, such as in the disruption of the second allele ("Deletion 2") of a gene of interest (GOI) following the disruption of the first allele.

Fig. 23 is a restriction endonuclease site map of plasmid pAA298 containing a double-crossover gene knock-out cassette for knocking out the function of a *FAT1* gene. As shown in the figure, the double-crossover gene knock-out cassette includes a *URA3* selectable marker gene (including the gene promoter and terminator) between 5' and 3' homologous sequences ("N-Fat1" and "C-Fat1," respectively) for the *FAT1* gene. The plasmid also contains elements from pCR-BluntII-TOPO.

Fig. 24 shows restriction endonuclease site maps of plasmids pAA1519 and pAA1520 each containing a double-crossover gene knock-out cassette for knocking out the function of a *CAT2* gene. Each *CAT2* gene deletion cassette includes a 5' *Candida viswanathii* strain ATCC 20336 *CAT2* DNA fragment ("Cv CAT2 5' homology"), a 3' *Candida viswanathii* strain ATCC 20336 *CAT2* DNA fragment ("Cv CAT2 3' homology") and a *Candida viswanathii* strain ATCC 20336 *URA3* gene fragment containing a *URA3* ORF ("Cv URA3"), *URA3* promoter and a *URA3* terminator followed by a repeat of the promoter. The gene knock-out cassettes are contained within pCR-BluntII-TOPO.

Fig. 25A and Fig. 25B show a comparison of the N-terminal (Fig. 25A) and C-terminal (Fig. 25B) amino acid sequences of an unmodified *Candida* strain ATCC 20336 carnitine acetyltransferase

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1308 of 3077

("Cv-CAT2 from pAA426") protein and of modified carnitine acetyltransferase proteins lacking one or both of the N-terminal mitochondrial targeting sequence (mts) and the C-terminal peroxisomal targeting sequence (pts). ("Cv-CAT2(-mts)" refers to the protein lacking only an N-terminal mitochondrial targeting sequence; "Cv-CAT2(-pts)" refers to the protein lacking only a C-terminal peroxisomal targeting sequence; "Cv-CAT2(-mts-pts)" refers to the protein lacking the N-terminal mitochondrial targeting sequence and the C-terminal peroxisomal targeting sequence.)

Fig. 26 is a map of plasmid pAA1164 containing all the elements of the pCR-BluntII-TOPO vector, two separate portions of a URA3 gene selectable marker originally cloned from *Candida* strain ATCC 20336, nucleic acid encoding a modified *Candida* strain ATCC 20336 Pox5p (i.e., Pox5(F98G)) and the *HDE* gene promoter and *POX4* gene terminator, both from *Candida* strain ATCC 20336. Also shown are the locations of sequences corresponding to oligonucleotides oAA4722 and oAA4723 which were used as primers in PCR amplification of the larger fragment sequence between these two sites. The amplified DNA fragment was used in the construction of pAA1610 (see details provided in the Examples herein). The "CvURA3" segment positioned following the 3' *URA3* fragment ("CvURA3 F23 to Stop") corresponds to the URA3 terminator, whereas the "CvURA3" segment positioned in front of the 5' *URA3* fragment ("CvURA3 M1 to L22") corresponds to the URA3 promoter.

Fig. 27 shows a comparison of the N-terminal amino acid sequences of an unmodified *Candida* strain ATCC 20336 cytosolic carnitine acetyltransferase ("Cv-Yat1p") protein and of modified Yat1p carnitine acetyltransferase proteins containing an added N-terminal mitochondrial targeting sequence (mts) in place of the initiating methionine of the unmodified Yat1p. ("Cv-Yat1p+CAT2mts" refers to the protein with an added N-terminal mitochondrial targeting sequence from the *Candida* strain ATCC 20336 mitochondrial carnitine acetyltransferase; "Cv-Yat1p+CIT1mts" refers to the protein with an added N-terminal mitochondrial targeting sequence from the *Candida* strain ATCC 20336 citrate synthase; "Cv-Yat1p+COX4mts" refers to the protein with an added N-terminal mitochondrial targeting sequence from the *Candida* strain ATCC 20336 cytochrome c oxidase.)

Fig. 28 is a restriction endonuclease site map of plasmid pAA245 which contains all the elements of the pCR-BluntII-TOPO vector, DNA encoding a *Candida* strain ATCC 20336 acetyl-CoA carboxylase ("CvACC1") enzyme and a 5' partial intron ("CvACC1 5' Intron partial") at the

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1309 of 3077

5' end of the ACC1 ORF. Also shown are the locations of sequences corresponding to oligonucleotides oAA0784 and oAA0785 which can be used as primers in PCR amplification of the Acc1-encoding DNA fragment from genomic DNA.

Fig. 29 is a restriction endonuclease site map of plasmid pAA326 which was generated by cutting pAA245 (Fig. 28) with BspQI and ligating the resulting *ACC1* gene fragment including 5' partial intron into BspQI-cut plasmid pAA105 (Fig.13) to put the gene under the control of the *PGK* promoter and terminator from *Candida* strain ATCC 20336.

Fig. 30 is a restriction endonuclease site map of plasmid pAA1634 generated by ligating a SpeI/XbaI fragment of pAA326 (Fig. 29) containing DNA encoding amino acids R643 to the STOP codon of Candida strain ATCC 20336 *ACC1* with SpeI/XbaI-digested pAA601 (Fig. 12). A series of site-directed mutagenesis reactions was performed on pAA1634 to introduce mutations into the truncated *ACC1* coding sequence as described in the Examples herein. Also shown in the figure is an illustration of two DNA fragments, one containing an *HDE* gene promoter and one containing DNA encoding amino acids M1 – S642 of the Acc1p (both from *Candida* strain ATCC 20336) that were subsequently ligated with pAA1634 to generate plasmids containing DNA encoding full-length mutant Acc1p.

Fig. 31 is a map of plasmid pAA2247 which was generated by ligating a SbfI/MluI fragment of plasmid pAA1908 containing DNA encoding a *Candida* strain ATCC 20336 Acc1p mutant (S1158A) surrounded by the *HDE* gene promoter ("Prom") and *PGK* gene terminator ("Term") from *Candida* strain ATCC 20336 with SbfI/MluI-digested plasmid pAA2153. Plasmid pAA2153 contains DNA encoding a *Candida* strain ATCC 20336 *URA3* selectable marker with a direct repeat of the T_{URA3} sequence located just upstream of the *URA3* gene promoter sequence (P_{URA3}) to yield T_{URA3}-P_{URA3}-*URA3*-T_{URA3}. The *URA3* selectable marker in pAA2153 is placed between genomic DNA sequence elements ("IGR5 5' homology" and "IGR5 3' homology") from *Candida* strain ATCC 20336 which are named IGR5. The IGR5 homology regions target integration of the intervening DNA into genomic DNA by homologous recombination. Also shown are the locations of DNA corresponding to primers oAA7259 and oAA7260 which can be used to amplify pAA2247 to generate a linear DNA for transformation of host cells.

Fig. 32 shows maps of plasmids pAA1613 and pAA1701 which contain double-crossover *Candida* strain ATCC 20336 *CRC1* gene knock-out cassettes that can be used to disrupt

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1310 of 3077

Candida CRC1 genes. The cassette in each plasmid contains a 5' *Candida* strain ATCC 20336 *CRC1* DNA fragment ("CRC1 5' homology"), a 3' *Candida* strain ATCC 20336 *CRC1* DNA fragment ("CRC1 3' homology") and a *Candida* strain ATCC 20336 *URA3* gene fragment containing a *URA3* promoter, *URA3* ORF, and a *URA3* terminator followed by a repeat of the *URA3* promoter. The cassettes in the two plasmids differ in the sizes of the *CRC1* gene homology regions and the orientation of the *URA3* selectable marker between the homology regions. The gene knock-out cassettes are contained within pCR-BluntII-TOPO. Also shown are the locations of DNA corresponding to primers oAA5511 and oAA5512 which can be used to PCR amplify linear cassettes from each plasmid to generate a linear DNA for transformation of host cells.

Fig. 33 is a map of plasmid pAA2214 which was generated by ligating a plasmid backbone containing the IGR5 homology regions and *URA3* selectable marker (with T_{URA3} repeat) amplified from plasmid pAA2247 (Fig. 31) with a 1,816-bp DNA fragment containing a *Candida* strain ATCC 20336 modified *CRC1* gene expression cassette. The expression cassette contains the glucose-6-phosphate isomerase (*G6PI*) promoter and *POX4* gene terminator from ATCC 20336 for controlling expression of the Crc1p. Also shown are the locations of DNA corresponding to primers oAA7259 and oAA7260 which can be used to PCR amplify a linear cassette from pAA2214 to generate a linear DNA for transformation of host cells.

Fig. 34 is a map of plasmid pAA2311 which contains a single-crossover cassette with nucleic acid encoding a *Candida* strain ATCC 20336 *CRC1* gene protein linked to a *Candida* strain ATCC 20336 *G6PI* low-expression promoter. The *CRC1* expression elements, P_{G6PI}-*CRC1*-T_{POX4}, were obtained as a DNA fragment amplified from plasmid pAA2214 (Fig. 33). This fragment was ligated with a fragment amplified from pAA1164 (Fig. 26) containing all the elements of the pCR-BluntII-TOPO vector, two separate portions of a URA3 selectable marker and a *POX4* gene terminator to yield pAA2311. Also shown are the locations of sequences corresponding to oligonucleotides oAA2206 and oAA2209 which can be used as primers in PCR amplification of a 3,307-bp linear, antibiotic-free DNA fragment for use in transforming host cells for expression of the Crc1p.

Fig. 35 is a map of plasmid pAA879 which contains a double-crossover knock-in cassette with nucleic acid encoding a *Candida* strain ATCC 20336 cytochrome P450 reductase (*CPRB*) gene protein. The plasmid contains all the elements of the pCR-BluntII-TOPO vector, two separate

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1311 of 3077

portions ("FAA1 5' homology" and "CvFAA1 3' homology") of a *Candida* strain ATCC 20336 *FAA1* gene, elements for the expression of CprB protein ("POX4 Promoter," "CvCPRB" and "POX4 term") and a *URA3* selectable marker with P_{URA3} repeat (URA3 Prom-URA3-URA3 Term-URA3 Prom). Also shown are the locations of sequences corresponding to oligonucleotides oAA3557 and oAA3564 which can be used as primers in PCR amplification of a linear DNA fragment for use in transforming host cells for disruption of the *FAA1* gene and expression of CPRB protein.

Fig. 36 is a map of plasmid pAA208 which includes two separate portions ("POX4 5' homology" and "POX4 3' homology") of a *Candida* strain ATCC 20336 *POX4* gene and a *Candida* strain ATCC 20336 *URA3* gene selectable marker with P_{URA3} repeat (URA3 Prom-URA3-URA3 Term-URA3 Prom).

Fig. 37 is a map of plasmid pAA850 which contains a double-crossover knock-in cassette with nucleic acid encoding a *Candida* strain ATCC 20336 *PEX11* gene protein. The plasmid contains all the elements of the pCR-BluntII-TOPO vector, two separate portions ("POX4 5' homology" and "POX4 3' homology") of a *Candida* strain ATCC 20336 *POX4* gene, elements for the expression of Pex11 protein ("POX4 Prom," "PEX11" and "POX4 term") and a *URA3* selectable marker with P_{URA3} repeat (URA3 Prom-URA3-URA3 Term-URA3 Prom). Also shown are the locations of sequences corresponding to oligonucleotides oAA3355 and oAA3357 which can be used as primers in PCR amplification of a linear DNA fragment for use in transforming host cells for disruption of the *POX4* gene and expression of Pex11 protein.

Fig. 38 shows a partial amino acid sequence of *Candida* strain ATCC 20336 Pox5p acyl-CoA oxidase and the results of analysis of the amino acid sequence using HotSpot Wizard (a software program tool for identifying sites for engineering of substrate specificity and/or activity of enzymes using a combination of structural, functional and sequence analysis). HotSpot Wizard identified several amino acid positions, or "hotspots," of Pox5p to mutate, with each position given a score from 1 (cold) to 9 (hot). The different HotSpot residues identified are highlighted in the figure and shaded according to the score assigned to the residue.

Fig. 39 is an illustration of an overlap extension PCR-based method for generating nucleic acids encoding mutants (F98G and W429F) of a *Candida* strain ATCC 20336 Pox5 acyl-CoA oxidase. The oligonucleotides ("Oligos") used in the PCR amplifications are listed in the table shown in

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1312 of 3077

the figure. Oligos B and C contain the desired point mutations. Sequences for each of the oligonucleotides are provided in the Examples herein.

Fig. 40 shows maps of plasmids pAA1117 and pAA1155 which contain *Candida* strain ATCC 20336 *PXA1* gene knock-out cassettes that can be used to disrupt alleles of *PXA1* genes. Both plasmids contain all the elements of the pCR-BluntII-TOPO vector, two separate portions ("PXA1 5' homology" and "PXA1 3' homology") of a *Candida* strain ATCC 20336 *PXA1* gene, and a *URA3* selectable marker with P_{URA3} repeat (URA3 Prom-URA3-URA3 Term-URA3 Prom). Also shown are the locations of sequences corresponding to oligonucleotides oAA2679/oAA2684 and oAA2914/oAA2919 which can be used as primers in PCR amplification of pAA1117 and pAA1155, respectively, to obtain a linear DNA fragment for use in transforming host cells for disruption of the *PXA1* gene. Plasmids pAA1117 and pAA1155 differ in the sizes of *PXA1* gene 5' and 3' homology sequences and the orientation of the *URA3* selectable marker nucleic acid sequence contained in the plasmids.

Fig. 41 is a flow diagram showing the parent-daughter relationship for exemplary engineered yeast strains that can be used in generating cells and organisms for use in target molecule platform and production systems described herein. Strains in bold type are Crc^- strains.

Fig. 42 shows a restriction endonuclease site map of plasmid pAA276 which contains a *Candida* strain ATCC 20336 *FAA1* gene knock-out cassette that can be used to disrupt alleles of *FAA1* genes. The plasmid contains all the elements of the pCR-BluntII-TOPO vector, two separate portions ("FAA1 N terminal" and "FAA1 C terminal") of a *Candida* strain ATCC 20336 *FAA1* gene, and a *URA3* selectable marker with URA3 Promoter repeat (Promoter URA3 -URA3 - Terminator URA3 – Promoter URA3).

Fig. 43 shows a restriction endonuclease site map of plasmid pAA918 which contains a *Candida* strain ATCC 20336 *POX5* gene knock-out cassette that can be used to disrupt alleles of *POX5* genes. The plasmid contains all the elements of the pCR-BluntII-TOPO vector, two separate portions ("5' FR POX5" and "3' FR POX5") of a *Candida* strain ATCC 20336 *POX5* gene, and a *URA3* selectable marker with URA3 terminator repeat (Tura3 – Pura3 - URA3 -Tura3).

Fig. 44 shows photographs of the agar plates obtained from spot growth assays of wild-type *Candida* strain ATCC 20336 ("CAT2/CAT2") and mutant strains as follows: *cat2-Δ1::*P_{URA3}*/cat2-*

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1313 of 3077

Δ2::URA3 (sAA4594) and a Cat2- *Candida* strain (*cat2-Δ1::P_{URA3}/cat2-Δ2::P_{URA3}*) that had been transformed with either pAA1610 or pAA1876. The upper and lower photographs on the left side of the figure are control agar plates containing synthetic complete media with dextrose minus uracil ("SCD-URA"), and the upper and lower photographs on the right side of the figure are plates containing yeast nitrogen base without amino acids, plus phosphate and 2% oleic acid ("YNBP + 2% oleic acid"). Each row of "spots" corresponds to serial dilutions of cells of the strain designated to the right of each row (increasing dilutions from right-to-left for the control agar plates and from left-to-right for the plates containing YNBP + 2% oleic acid).

Fig. 45 shows photographs of the agar plates obtained from spot growth assays of wild-type *Candida* strain ATCC 20336 ("CAT2/CAT2") and mutant strains as follows: *cat2-Δ1::P_{URA3}/cat2-Δ2::URA3* (sAA4594) and a Cat2- *Candida* strain (*cat2-Δ1::P_{URA3}/cat2-Δ2::P_{URA3}*) that had been transformed with one or more of plasmids pAA1610, pAA1967, pAA1968 and pAA1969. The photograph on the left side of the figure is of control agar plates containing synthetic complete media with dextrose minus uracil ("SCD-URA"), and the photograph on the right side of the figure is of plates containing yeast nitrogen base without amino acids, plus phosphate and 2% oleic acid ("YNBP + 2% oleic acid"). Each row of "spots" corresponds to serial dilutions of cells of the strain designated to the right of each row (increasing dilutions from right-to-left for the control agar plates and from left-to-right for the plates containing YNBP + 2% oleic acid).

Fig. 46 shows photographs of the agar plates obtained from spot growth assays of wild-type *Candida* strain ATCC 20336 ("CAT2/CAT2") and mutant strains as follows: *cat2-Δ1::P_{URA3}/cat2-Δ2::URA3* (sAA4594) and a Cat2- *Candida* strain (*cat2-Δ1::P_{URA3}/cat2-Δ2::P_{URA3}*) that had been transformed with pAA1846, pAA1875 and one or more of plasmids pAA1967, pAA1968, pAA1969 and pAA1847. The photograph on the left side of the figure is of control agar plates containing synthetic complete media with dextrose minus uracil ("SCD-URA"), and the photograph on the right side of the figure is of plates containing yeast nitrogen base without amino acids, plus phosphate and 2% oleic acid ("YNBP + 2% oleic acid"). Each row of "spots" corresponds to serial dilutions of cells of the strain designated to the right of each row (increasing dilutions from right-to-left for the control agar plates and from left-to-right for the plates containing YNBP + 2% oleic acid).

Fig. 47 shows photographs of the agar plates obtained from spot growth assays of wild-type *Candida* strain ATCC 20336 and mutant strains as follows: *ura3/ura3 pox4a::ura3/pox4b::ura3*

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1314 of 3077

faa1:: P_{URA3}/*faa1::*P_{URA3} *fat1-Δ1::*P_{URA3}/*fat1-Δ2::URA3* (sAA875), *ura3/ura3 pox4a::ura3/pox4b::ura3 faa1::*P_{URA3}/*faa1::*P_{URA3} *fat1-Δ1::*P_{URA3}/*fat1-Δ2::*P_{URA3} *crc1-Δ1::URA3/CRC1* (sAA4057), *ura3/ura3 pox4a::ura3/pox4b::ura3 faa1::*P_{URA3}/*faa1::*P_{URA3} *fat1-Δ1::*P_{URA3}/*fat1-Δ2::*P_{URA3} *crc1-Δ1::*P_{URA3}/*crc1-Δ2::URA3* (sAA4281), *ura3/ura3 crc1-Δ1::URA3/CRC1* (sAA4368), and *ura3/ura3 crc1-Δ1::*P_{URA3}/*crc1-Δ2::URA3* (sAA9398). Also shown are the agar plates obtained from spot growth assays of strains sAA5916, sAA5917 and sAA5918 generated by transforming strain sAA4377 (*ura3/ura3 pox4a::ura3/pox4b::ura3 faa1::*P_{URA3}/*faa1::*P_{URA3} *fat1-Δ1::*P_{URA3}/*fat1-Δ2::*P_{URA3} *crc1-Δ1::*P_{URA3}/*crc1-Δ2::*P_{URA3}) with a double-crossover integration cassette containing DNA encoding a *Candida* strain ATCC 20336 Crc1p linked to the *G6PI*, *PIC2* or *SUL2* promoter, respectively. The photographs on the left side of the figure are of control agar plates containing synthetic complete media with dextrose minus uracil ("SCD-URA"), and the photographs on the right side of the figure are of plates containing yeast nitrogen base without amino acids, plus phosphate and 2% oleic acid ("YNBP + 2% oleic acid"). Each row of "spots" corresponds to serial dilutions of cells of the strain designated to the right of each row (increasing dilutions from right-to-left for the control agar plates and from left-to-right for the plates containing YNBP + 2% oleic acid).

Fig. 48A shows GC-MS of valencene production from sAA7449 with its library match shown in Fig. 48B.

Fig. 49A shows GC-MS of valencene production from sAA7453 with its library match shown in Fig. 49B.

Fig. 50 shows the yeast upper mevalonate pathway. The enzymes acetyl-CoA C-acetyltransferase (*ERG10*), HMG-CoA synthase (*ERG13*), and HMG-CoA reductase (*HMG1*), are native to yeast described herein. The bracketed number under each enzyme refers to the IUBMB enzyme nomenclature for the enzyme.

Fig. 51 shows the yeast lower mevalonate pathway. The enzymes Mevalonate kinase (*ERG12*), Phosphomevalonate kinase (*ERG8*), Diphosphomevalonate decarboxylase (*MVD1*; *ERG19*), Isopentenyl diphosphate delta isomerase (*IDI1*), are native to yeast described herein. The bracketed number under each enzyme refers to the IUBMB enzyme nomenclature for the enzyme.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1315 of 3077

Fig. 52 shows a biosynthetic pathway to ergosterol or geranylgeranyl diphosphate. *ERG20*, *ERG9*, and *BTS1* are native to yeast described herein.

Fig. 53 shows carotenoid production from farnesyl pyrophosphate. *CrtE*, *CrtB*, *CrtI, CrtY, CrtZ, and CrtW* are heterologous genes.

Fig. 54 shows an alternative route to lutein, canthaxanthin, and astaxanthin from β-carotene by the cytochrome P450 hydroxylase *CrtS*. *CrtS* is a heterologous enzyme which requires the activity of *CrtR*, a heterologous cytochrome p450 reductase.

Fig. 55 shows a route from farnesyl pyrophosphate to valencene by TPS1, a heterologous valencene synthase.

Fig. 56 show an absorption spectra of extracted carotenoids from sAA001, sAA7443, sAA7444, and sAA7445. Strains sAA7443, aAA7444 and sAA7445 are generated from sAA001 plus CsCrtE, CsCrtB and CsCrtI each under a glyceraldehyde-3-phosphate dehydrogenase (GPD) promoter. Lycopene absorption maxima are expected at 443, 472, and 502 nm.

Fig. 57 shows an HPLC chromatogram of lycopene production in sAA001 and sAA8503.

Detailed Description

There are multiple cellular metabolic pathways that utilize carbon-containing molecules for varying purposes, such as, for example, biomass production, energy generation and growth. Different metabolic pathways can occur in distinct areas of the cell. For example, in eukaryotic microorganisms metabolic processes such as glycolysis, the pentose phosphate pathway and gluconeogenesis occur in the cytoplasm, whereas β-oxidation, the tricarboxylic acid (TCA) cycle and glyoxylate cycle are carried out, in whole or in part, in cellular organelles. The different metabolic pathways can be differentially utilized to maintain the basic functions required for survival of a microorganism under a variety of conditions (e.g., varying carbon source, nutrient, and oxygen availabilities). These features contribute to the ability of microorganisms to readily adapt to a variety of environmental conditions. The adaptability of microorganisms facilitates manipulation of microbial metabolic processes for the production of commercially useful materials (e.g., proteins, lipids, and organic acids).

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1316 of 3077

There are a number of molecules that are valuable as final products, and/or as raw materials in generating a desired product, that incorporate carbon atoms resulting from catabolic and anabolic carbon-metabolism pathways in microorganisms. A desired molecule that can be produced in cells and microorganisms is referred to herein as a "target" molecule or product. Some of these molecules are generated in wild-type microorganisms, whereas others that are not produced in a native microorganism can be generated through modification of a wild-type organism. In either case, the goal in the development of cellular and microbial production systems is to maximize yield and efficiency and minimize loss for process economy. Biological cell- or organism-based systems for production of carbon-containing molecules, such as, for example, polymers of substituted or unsubstituted hydrocarbons, may not be optimally efficient due to loss of carbon atoms that are transferred into other, non-target-producing, metabolic processes in the cell. In order for a bioproduction system to be cost-efficient and economically viable, it generally should meet certain metrics of titer, rate and yield with minimal by-product formation.

In order to minimize carbon loss and increase process efficiency of biological cell-based production systems, provided herein are modified cells and organisms (e.g., microorganisms) in which the flux of cellular carbon has been altered relative to an unmodified cell or organism. In some aspects, the cells or organisms are modified to redirect carbon from entering one or more growth and/or energy production metabolic pathways so that it is available for use in other inherent and/or engineered production processes. Alteration of carbon flux facilitates engineering of the cells or organisms for enhanced production of desired target molecules, including, for example, organic acids, terpenes and precursor molecules that can be used in the production of industrial chemicals. As such, modified cells and microorganisms provided herein are useful as platform systems that can be used for enhanced production of many different desired target molecules (e.g., terpenes) either singly or multiply in co-production microbial systems. Also provided herein are methods of modifying cellular carbon flux and methods of generating cells or microorganisms in which carbon flux has been optimized for production of target molecules, e.g., carbon-containing compounds, terpenes. Further provided herein are methods of producing target molecules (e.g., terpenes) using cell-based or microbial biosynthesis systems, including, for example, modified cells or microorganisms in which the flux of cellular carbon has been altered relative to an unmodified microorganism. Also provided are

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1317 of 3077

compositions, including, but not limited to, nucleic acids and chemical media and combinations, that can be used in the methods provided herein.

Terpenes

Provided herein are methods for producing terpenes and engineered cells and microorganisms capable of producing terpenes. Terpenes are compounds generally made from isoprene units $(C_5H_8)_n$. Isoprene, also known as 2-methyl-1,3-butadiene or 2-Methylbuta-1,3-diene, is an organic compound having the formula $CH_2=C(CH_3)-CH=CH_2$. Isoprene units may be assembled in various combinations to produce thousands of possible terpene compounds. Terpenes may include hemiterpenes (one isoprene unit), monoterpenes (two isoprene units), sesquiterpenes (three isoprene units), diterpenes (four isoprene units), sesterterpenes (five isoprene units), triterpenes (six isoprene units), sesquarterpenes (seven isoprene units), tetraterpenes (eight isoprene units), and polyterpenes (many isoprene units; more than eight isoprene units). Hemiterpenes may include, for example, isoprene, prenol, isovaleric acid, benzoate, and prenyl acetate. Monoterpenes may include, for example, α-ocimene, β-ocimene, α-myrcene, β-myrcene, geraniol, citronellal, citronellol, linalool, citral A, halomon, S-limonene, R-limonene, phellandrene, α-terpinene, menthol, S-carvone, R-carvone, safranal, terpineol, thymol, carvacrol, umbellulone, piperitone, pulegone, rose oxide, lactisole, sabinene, camphene, eucalyptol, thujene, thujone, pinene, nepetalactone, ascaridole, borneol, verbenone, camphor, ethyl fenchol, and cyclosantene. Sesquiterpenes may include, for example, farnesene, farnesol, nerolidol, zingiberene, humulene, bisabolane, bisabolene, elemol, caryophyllene, guaiol, vetivazulene, cadinene, caryophyllene, cuparane, laurene, laurane, oppositane, mutisianthol, thapsane, lepidozanes, chiloscyphane, pinguisanes, herbertanes, botrydial, ngaione, longifolene, copaene, patchoulol, norpatchoulenol, santalol, cortisol, progesterone, oestrogen, testosterone, and 11-ketotestosterone. Diterpenes may include, for example, phytol, cembrene A, retinol, retinal, labdane, abietic acid, ferruginol, tetrahydrocannabinol, forskolin, cafestol, and gibberellins. Sesterterpenes may include, for example, geranylfarnesol and amyrin. Triterpenes may include, for example, squalene, sapelenins, tangshenoside I, cholecalciferol, dihydrotachysterol, lanosterol, cholesterol, cycloartenol, lovastatin, cucurbitacin, 20-hydroxyecdysone, campesterol, stigmasterol, ergosterol, bile acids, betulinic acid, withaferin A, ginsenosides, eleutherosides, astragalosides, bacoside A, araloside A, ziziphin, and gymnemic acids. Tetraterpenes may include, for example, phytoene, phytofluene, lycopene, torulene, γ-carotene, δ-carotene, citranaxanthin, rubixanthin, α-carotene, β-carotene, astaxanthin,

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1318 of 3077

cryptoxanthins, canthaxanthin, violaxanthin, antheraxanthin, zeaxanthin, lutein, diadinoxanthin, diatoxanthin, neoxanthin, fucoxanthins and flavoxanthins. Polyterpenes may include, for example, polypropylene, rubber and latex.

Terpenes may include molecules containing only carbon (C) and hydrogen (H) atoms and also may include molecules containing carbon (C) and hydrogen (H) atoms in addition to other atoms (e.g. oxygen (O)). Terpene molecules containing atoms in addition to carbon (C) and hydrogen (H) may be referred to as terpenoids. In certain instances, compounds referred to as terpenoids contain only carbon (C) and hydrogen (H). Use of the term terpene herein includes terpenoids. Terpene structures may be acyclic (no carbon rings), monocyclic (one carbon ring), bicyclic (two carbon rings), tricyclic (three carbon rings), tetracyclic (four carbon rings), pentacyclic (five carbon rings), hexacyclic (six carbon rings), heptacyclic (seven carbon rings), or octacyclic (eight carbon rings). Terpene structures may contain more than eight carbon rings.

Terpenes may include carotenoids. Carotenoids (also referred to as tetraterpenoids) generally are pigments that provide red, orange, and yellow pigmentation to certain organisms, and include certain 40-carbon molecules. Over 700 carotenoids are known, which include, for example, carotenes (e.g., γ-carotene, β-carotene, α-carotene, ε-carotene, δ-carotene), xanthophylls (e.g., lutein, zeaxanthin, neoxanthin, violaxanthin, flavoxanthin, and α- and β-cryptoxanthin), torulene, lycopene, lutein epoxide, hydroxyechinenone, phoenicoxanthin, astaxanthin, fucoxanthin, 19′-butanoyloxyfucoxanthin, 19’-hexanoyloxyfucoxanthin, diatoxanthin, diadinoxanthin 19-hexanoyloxyparacentrone 3-acetate, gyroxanthin, canthaxanthin, citranaxanthin, and apo-caroten-ester. About 40 to 50 carotenoids can be found in food, and about 6 carotenoids can be found in human serum. Carotenoids may be synthesized de novo by organisms including, for example, archae, bacteria, fungi, algae, plants, aphids & spider mites. Carotenoids have various uses including, for example, antioxidants, color attractants, protection of plants from sun damage, precursors to vitamins, inflammatory disease protection, eye health, bone growth, and immune function.

Provided herein are microorganisms genetically modified to produce terpenes. For example, a genetically modified microorganism may comprise one or more heterologous nucleic acids encoding one or more terpene biosynthesis polypeptides. Terpene biosynthesis polypeptides may include any polypeptide capable of producing a terpene molecule or a terpene precursor. Terpene biosynthesis polypeptides may include one or more enzymes. For example, terpene

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1319 of 3077

biosynthesis polypeptides may include one or more of terpene synthase, phytoene synthase, geranylgeranyl diphosphate synthase, phytoene desaturase, lycopene cyclase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase, zeaxanthin glucosyltransferase, valencene synthase, and cytochrome p450 reductase. In some embodiments, a genetically modified microorganism comprises heterologous nucleic acids encoding one or more of phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase. In some embodiments, a genetically modified microorganism comprises heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase. In some embodiments, a genetically modified microorganism comprises heterologous nucleic acids encoding one or more of geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase. In some embodiments, a genetically modified microorganism comprises heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase. In some embodiments, a genetically modified microorganism comprises heterologous nucleic acids encoding one or more of geranylgeranyl diphosphate synthase, phytoene desaturase, lycopene cyclase, and phytoene synthase. In some embodiments, a genetically modified microorganism comprises heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase, lycopene cyclase, and phytoene synthase. In some embodiments, a genetically modified microorganism comprises a heterologous nucleic acid encoding valencene synthase.

In some embodiments, a genetically modified microorganism comprises heterologous nucleic acids encoding one or more of geranylgeranyl diphosphate synthase, phytoene desaturase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase, and cytochrome p450 reductase. In some embodiments, a genetically modified microorganism comprises heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase, and cytochrome p450 reductase. In some embodiments, a genetically modified microorganism comprises heterologous nucleic acids encoding one or more of geranylgeranyl diphosphate synthase, phytoene desaturase, lycopene cyclase, phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase, and cytochrome p450 reductase. In some embodiments, a genetically modified microorganism comprises heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase, lycopene cyclase, phytoene

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1320 of 3077

synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase, and cytochrome p450 reductase.

Terpene biosynthesis polypeptides may include any polypeptide (e.g., enzyme) capable of producing a terpene precursor. In certain embodiments, terpene biosynthesis polypeptides include one or more polypeptides (e.g., enzymes) in a mevalonate pathway (e.g., upper mevalonate pathway, lower mevalonate pathway). Typically, a mevalonate pathway begins with cytoplasmic acetyl-CoA, and the end product of a mevalonate pathway is isopentenyl diphosphate (IPP), which can serve as a monomer unit for terpene production. Polypeptides (e.g., enzymes) in a mevalonate pathway may include, for example, one or more of acetyl-CoA C-acetyltransferase, HMG-CoA synthase, HMG-CoA reductase, mevalonate kinase, phosphomevalonate kinase, diphosphomevalonate decarboxylase, isopentyl diphosphate delta isomerase, dimethylallyltranstransferase, and farnesyl diphosphate synthetase. In some embodiments, amounts and/or activities of one or more proteins chosen from acetyl-CoA C-acetyltransferase, HMG-CoA synthase, HMG-CoA reductase, mevalonate kinase, phosphomevalonate kinase, diphosphomevalonate decarboxylase, isopentyl diphosphate delta isomerase, dimethylallyltranstransferase, and farnesyl diphosphate synthetase are increased in a genetically modified microorganism provided herein. In some embodiments, a genetically modified microorganism comprises heterologous nucleic acids encoding one or more of acetyl-CoA C-acetyltransferase, HMG-CoA synthase, HMG-CoA reductase, mevalonate kinase, phosphomevalonate kinase, diphosphomevalonate decarboxylase, isopentyl diphosphate delta isomerase, dimethylallyltranstransferase, and farnesyl diphosphate synthetase.

In some embodiments, a heterologous nucleic acid encoding a terpene biosynthesis polypeptide is endogenously expressed in a microorganism that is different from the microorganism genetically modified to produce terpenes. In some embodiments, a heterologous nucleic acid encoding a terpene biosynthesis polypeptide is endogenously expressed in a microorganism that is of a different species, genus, family, order, class, phylum or kingdom than the microorganism genetically modified to produce terpenes. For example, a heterologous nucleic acid encoding a terpene biosynthesis polypeptide may be endogenously expressed in a microorganism chosen from *Cronobacter* spp, *Callitropsis* spp, *Xanthophyllomyces* spp, *Agrobacterium* spp, and *Pantoea* spp.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1321 of 3077

A heterologous nucleic acid encoding a terpene biosynthesis polypeptide may include an endogenous nucleic acid linked to a promoter or other regulatory element that is not normally functionally linked to the endogenous nucleic acid sequence. Promoters may include, for example, a mutated form of an endogenous promoter, an endogenous promoter replaced with a modified version of the endogenous promoter, and an endogenous promoter replaced with a promoter located elsewhere in the same genome or from a different genome. A modification may be performed by directed mutagenesis, and the mutated promoter may be referred to as "heterologous" because it is no longer native. In certain instances, a modification may be performed by replacing an endogenous promoter with a modified promoter, the latter of which would be considered heterologous because the modified promoter is not normally functionally linked to an endogenous nucleic acid.

In some embodiments, a heterologous nucleic acid encoding a terpene biosynthesis polypeptide is regulated according to an inducible system. For example, a heterologous nucleic acid encoding a terpene biosynthesis polypeptide may be regulated by a nucleic acid that provides for fatty acid induction of expression of the terpene biosynthesis polypeptide. The fatty acid that can induce expression of a terpene biosynthesis polypeptide may be a saturated fatty acid or an unsaturated fatty acid. For example, a fatty acid may be chosen from one or more of oleic acid, palmitoleic acid, erucic acid, linoleic acid, palmitic acid, caproic acid, enanthic acid, caprylic acid pelargonic acid, capric acid, undecylic acid, lauric acid, myristic acid, pentadecanoic acid, margaric acid, stearic acid arachidic acid, behenic acid, tridecylic acid, and linolenic acid.

In some embodiments, a nucleic acid that provides for fatty acid induction of expression of a terpene biosynthesis polypeptide is a promoter or a portion of a promoter. In some embodiments, a nucleic acid that provides for fatty acid induction of expression of a terpene biosynthesis polypeptide comprises a fatty acid response element. In some embodiments, a nucleic acid that provides for fatty acid induction of expression of a terpene biosynthesis polypeptide is a promoter, or a portion of a promoter, comprising a fatty acid response element. In some embodiments, a fatty acid response element comprises an oleic acid response element. In some embodiments, a nucleic acid that provides for fatty acid induction of expression of a terpene biosynthesis polypeptide comprises a promoter region of a heterologous gene. In some embodiments, a nucleic acid that provides for fatty acid induction of expression of a terpene biosynthesis polypeptide comprises a promoter region of an endogenous gene. For example, a promoter region may be a promoter region of a gene

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1322 of 3077

encoding hydratase-dehydrogenase-epimerase (HDE), acyl co-A oxidase (POX; e.g., POX4), acyl co-A thiolase (POT), peroxin (PEX), peroxisomal protein POX18, or peroxisomal adenine nucleotide transporter protein (ANT1). In some embodiments, a nucleic acid that provides for fatty acid induction of expression of a terpene biosynthesis polypeptide comprises a promoter region of a gene encoding hydratase-dehydrogenase-epimerase (HDE).

In some embodiments, a nucleic acid that provides for fatty acid induction of expression of a terpene biosynthesis polypeptide comprises a promoter region of a gene from *Candida*. For example, a promoter region may be a promoter region of a gene encoding *Candida* hydratase-dehydrogenase-epimerase (HDE), *Candida* acyl co-A oxidase (POX; e.g., POX4), *Candida* acyl co-A thiolase (POT), *Candida* peroxin (PEX), peroxisomal protein POX18, or *Candida* peroxisomal adenine nucleotide transporter protein (ANT1). In some embodiments, a nucleic acid that provides for fatty acid induction of expression of a terpene biosynthesis polypeptide comprises a promoter region of a gene encoding *Candida* hydratase-dehydrogenase-epimerase (HDE).

In some embodiments, a heterologous nucleic acid encoding a terpene biosynthesis polypeptide is regulated by a nucleic acid that provides for alkane induction of expression of the terpene biosynthesis polypeptide. An alkane may be chosen from one or more of hexane, heptane, nonane, decane, undecane, dodecane, tridecane, tetradecane, pentadecane, hexadecane, heptadecane, and octadecane, for example. In some embodiments, a nucleic acid that provides for alkane induction of expression of a terpene biosynthesis polypeptide is a promoter or a portion of a promoter. In some embodiments, a nucleic acid that provides for alkane induction of expression of a terpene biosynthesis polypeptide comprises an alkane response element. In some embodiments, a nucleic acid that provides for alkane induction of expression of a terpene biosynthesis polypeptide is a promoter, or a portion of a promoter, comprising an alkane response element. In some embodiments, an alkane response element comprises an alkane response element 1 (ARE1) sequence or an ARE1-like sequence as described herein. In some embodiments, a nucleic acid that provides for alkane induction of expression of a terpene biosynthesis polypeptide comprises a promoter region, or a portion of a promoter region, of a heterologous gene.

In some embodiments, a heterologous nucleic acid encoding a terpene biosynthesis polypeptide is regulated by a nucleic acid that provides for glucose induction of expression of the terpene

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1323 of 3077

biosynthesis polypeptide. In some embodiments, a nucleic acid that provides for glucose induction of expression of a terpene biosynthesis polypeptide comprises a promoter region of a gene encoding glyceraldehyde-3-phosphate dehydrogenase (GPD). In some embodiments, a nucleic acid that provides for glucose induction of expression of a terpene biosynthesis polypeptide comprises a promoter region of a gene from *Candida*. In some embodiments, a nucleic acid that provides for glucose induction of expression of a terpene biosynthesis polypeptide comprises a promoter region of a gene encoding *Candida* glyceraldehyde-3-phosphate dehydrogenase (GPD).

Modification of an acetyl-CoA C-acetyltransferase activity

An initial step in the upper mevalonate pathway is the conversion of 2 acetyl-CoA molecules to acetoacetyl-CoA and CoA. An acetyl-CoA C-acetyltransferase enzyme (also referred to as ERG10, acetoacetyl-CoA thiolase, acetyl-CoA acetyltransferase, ergosterol biosynthesis protein 10, beta-acetoacetyl coenzyme A thiolase, 2-methylacetoacetyl-CoA thiolase, 3-oxothiolase, acetyl coenzyme A thiolase, acetyl-CoA:N-acetyltransferase, thiolase II, LPB3, TSM0115; EC 2.3.1.9) is a cytosolic and/or peroxisomal enzyme that can transfer an acetyl group from one acetyl-CoA molecule to another, forming acetoacetyl-CoA. In some embodiments of the cells, organisms, compositions and methods provided herein, the amount and/or activity of an acetyl-CoA C-acetyltransferase in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase the amount and/or activity of an acetyl-CoA C-acetyltransferase, may be modified to decrease the amount and/or activity of an acetyl-CoA C-acetyltransferase, or may be modified to alternately increase and decrease the amount and/or activity of an acetyl-CoA C-acetyltransferase depending, for example, on the cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

In certain aspects, the amount and/or activity of an acetyl-CoA C-acetyltransferase in a cell is increased. Increasing the amount and/or activity of an acetyl-CoA C-acetyltransferase may be particularly beneficial in embodiments in which the flux of carbons from fatty acids is directed toward a particular target product pathway (e.g., mevalonate production, isopentenyl diphosphate production, geranyl diphosphate production, farnesyl diphosphate production, geranylgeranyl diphosphate production, terpene production, carotenoid production, beta carotene production, lycopene production, astaxanthin production) and away from other cellular metabolic pathways not involved in target molecule production.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1324 of 3077

In certain embodiments, acetyl-CoA C-acetyltransferase activity is unchanged in a host or engineered cell or organism. In some embodiments, the amount and/or activity of a host acetyl-CoA C-acetyltransferase can be increased by increasing the number of copies of a nucleic acid encoding an acetyl-CoA C-acetyltransferase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the nucleic acid); by increasing the activity of a promoter that regulates transcription of a nucleic acid encoding an acetyl-CoA C-acetyltransferase; or by increasing the number of copies of a nucleic acid encoding an acetyl-CoA C-acetyltransferase and increasing the activity of a promoter that regulates transcription of a nucleic acid encoding an acetyl-CoA C-acetyltransferase. In some embodiments, an acetyl-CoA C-acetyltransferase is endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, the amount of acetyl-CoA C-acetyltransferase protein expressed in a cell can be increased by introducing heterologous nucleic acid encoding an acetyl-CoA C-acetyltransferase into a cell or microorganism. For example, introduction of heterologous nucleic acid encoding an acetyl-CoA C-acetyltransferase can result in increased copy number of such nucleic acids and/or provide for modification of the cellular location in which the protein is expressed.

In some embodiments, an acetyl-CoA C-acetyltransferase enzyme can be a fungal or bacterial protein. In a particular embodiment, the acetyl-CoA C-acetyltransferase enzyme can be a *Candida* (e.g., *C. tropicalis*, *C. viswanathii*, *C. maltosa*, *C. cloacae*), *Yarrowia* (e.g., *Y. lipolytica*), *Fusarium* (e.g., *F. oxysporum*), *Bacillus* (e.g., *B. megaterium*, *B. subtilis*), *Chronobacter* (e.g., *C. sakazakii*), *Xanthophyllomyces* (e.g., *X. dendrorhous*), or *Pantoea* (e.g., *P. ananatis*) protein. An example of a *Candida viswanathii* nucleotide sequence (SEQ ID NO: 371) encoding a polypeptide (*ERG10*; SEQ ID NO: 337) having an acetyl-CoA C-acetyltransferase activity is provided herein. Additional nonlimiting examples of nucleotide sequences encoding polypeptides having acetyl-CoA C-acetyltransferase activity include: *Candida tropicalis CTPACTB* gene for acetoacetyl-CoA thiolase A (Genbank accession no. D13471.1), *Candida tropicalis CTPACTA* gene for acetoacetyl-CoA thiolase A (Genbank accession no. D13470.1), and *Candida tropicalis MYA-3404* acetyl-CoA acetyltransferase IB (NCBI Reference Sequence: XM_002547232.1).

Presence, absence or amount of acetyl-CoA C-acetyltransferase activity can be detected by any suitable method known in the art. For example, detection can be performed by using an

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1325 of 3077

acetyltransferase activity kit (e.g., ENZO, ADI-907-026). Nucleic acid sequences encoding native and/or modified acetyl-CoA C-acetyltransferase sequences also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cell or organism exhibits increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism.

The promoter used for regulating transcription of a nucleic acid (e.g., a heterologous nucleic acid) encoding an acetyl-CoA C-acetyltransferase can be modified. For example, the amount of an acetyl-CoA C-acetyltransferase protein expressed in a particular cellular location may be increased by including in the nucleic acid (e.g., heterologous nucleic acid) a strong heterologous promoter and/or a promoter that provides for a different pattern of expression in the cell or microorganism. An example of one such heterologous promoter is a *Candida* oleate-induced peroxisomal protein (*POX18*) gene promoter. Another example of one such heterologous promoter is a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene promoter. The nucleotide sequences of a *Candida viswanathii* HDE gene promoter and a *Candida* POX18 gene promoter are provided herein as are examples of additional fatty acid-inducible promoters. In some embodiments, the promoter is modified by replacing the endogenous promoter with a promoter comprising an alkane response element (ARE1). Non-limiting examples of assays suitable for assessing induction of acetyl-CoA C-acetyltransferase expression by a carbon source or feedstock include RT-PCR or qRT-PCR after the host cell or microorganism has been exposed to the chosen carbon source or feedstock for varying amounts of time.

Decreasing acetyl-CoA C-acetyltransferase activity in a cell can be accomplished by modifying the amount of acetyl-CoA C-acetyltransferase protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous acetyl-CoA C-acetyltransferase gene in a cell or organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type acetyl-CoA C-acetyltransferase such that the encoded modified or substituted acetyl-CoA C-acetyltransferase protein has a reduced enzyme activity.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1326 of 3077

Modification of an HMG-CoA synthase activity

A further step in the upper mevalonate pathway is the reaction in which acetyl-CoA condenses with acetoacetyl-CoA to form 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA). An HMG-CoA synthase enzyme (also referred to as ERG13, hydroxymethylglutaryl-CoA synthase, (S)-3-hydroxy-3-methylglutaryl-CoA acetoacetyl-CoA-lyase, (CoA-acetylating), 3-hydroxy-3-methylglutaryl CoA synthetase, 3-hydroxy-3-methylglutaryl coenzyme A synthase, 3-hydroxy-3-methylglutaryl coenzyme A synthetase, 3-hydroxy-3-methylglutaryl-CoA synthase, 3-hydroxy-3-methylglutaryl-coenzyme A synthase, beta-hydroxy-beta-methylglutaryl-CoA synthase, acetoacetyl coenzyme A transacetase, hydroxymethylglutaryl coenzyme A synthase, and hydroxymethylglutaryl coenzyme A-condensing enzyme; EC 2.3.3.10) contains a catalytic cysteine residue that acts as a nucleophile in the acetylation of the enzyme by acetyl-CoA (first substrate) to produce an acetyl-enzyme thioester, releasing the reduced coenzyme A. A subsequent nucleophilic attack on acetoacetyl-CoA (second substrate) forms HMG-CoA. In some embodiments of the cells, organisms, compositions and methods provided herein, the amount and/or activity of an HMG-CoA synthase in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase the amount and/or activity of an HMG-CoA synthase, may be modified to decrease the amount and/or activity of an HMG-CoA synthase, or may be modified to alternately increase and decrease the amount and/or activity of an HMG-CoA synthase depending, for example, on the cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

In certain aspects, the amount and/or activity of an HMG-CoA synthase in a cell is increased. Increasing the amount and/or activity of an HMG-CoA synthase may be particularly beneficial in embodiments in which the flux of carbons from fatty acids is directed toward a particular target product pathway (e.g., mevalonate production, isopentenyl diphosphate production, geranyl diphosphate production, farnesyl diphosphate production, geranylgeranyl diphosphate production, terpene production, carotenoid production, beta carotene production, lycopene production, astaxanthin production) and away from other cellular metabolic pathways not involved in target molecule production.

In certain embodiments, HMG-CoA synthase activity is unchanged in a host or engineered cell or organism. In some embodiments, the amount and/or activity of a host HMG-CoA synthase can be increased by increasing the number of copies of a nucleic acid encoding an HMG-CoA

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1327 of 3077

synthase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the nucleic acid); by increasing the activity of a promoter that regulates transcription of a nucleic acid encoding an HMG-CoA synthase; or by increasing the number of copies of a nucleic acid encoding an HMG-CoA synthase and increasing the activity of a promoter that regulates transcription of a nucleic acid encoding an HMG-CoA synthase. In some embodiments, an HMG-CoA synthase is endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, the amount of HMG-CoA synthase protein expressed in a cell can be increased by introducing heterologous nucleic acid encoding an HMG-CoA synthase into a cell or microorganism. For example, introduction of heterologous nucleic acid encoding an HMG-CoA synthase can result in increased copy number of such nucleic acids and/or provide for modification of the cellular location in which the protein is expressed.

In some embodiments, an HMG-CoA synthase enzyme can be a fungal or bacterial protein. In a particular embodiment, the HMG-CoA synthase enzyme can be a *Candida* (e.g., *C. tropicalis*, *C. viswanathii*, *C. maltosa*, *C. cloacae*), *Yarrowia* (e.g., *Y. lipolytica*), *Fusarium* (e.g., *F. oxysporum*), *Bacillus* (e.g., *B. megaterium*, *B. subtilis*), *Chronobacter* (e.g., *C. sakazakii*), *Xanthophyllomyces* (e.g., *X. dendrorhous*), or *Pantoea* (e.g., *P. ananatis*) protein. An example of a *Candida viswanathii* nucleotide sequence (SEQ ID NO: 372) encoding a polypeptide (*ERG13*; SEQ ID NO: 338) having an HMG-CoA synthase is provided herein. Additional nonlimiting examples of nucleotide sequences encoding polypeptides having HMG-CoA synthase activity include: *Candida tropicalis MYA-3404* hydroxymethylglutaryl-CoA synthase (NCBI Reference Sequence: XM_002546412.1), and *Candida tanzawaensis NRRL Y-17324* hydroxymethylglutaryl-CoA synthase (NCBI Reference Sequence: XM_020207398.1).

Presence, absence or amount of HMG-CoA synthase activity can be detected by any suitable method known in the art. For example, detection can be performed by using a visible wavelength spectrophotometric assay for HMG-CoA synthase (see e.g., Skaff et al. (2010) Anal. Biochem. 396(1):96-102). Nucleic acid sequences encoding native and/or modified HMG-CoA synthase sequences also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cell or organism exhibits increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1328 of 3077

The promoter used for regulating transcription of a nucleic acid (e.g., a heterologous nucleic acid) encoding an HMG-CoA synthase can be modified. For example, the amount of an HMG-CoA synthase protein expressed in a particular cellular location may be increased by including in the nucleic acid (e.g., heterologous nucleic acid) a strong heterologous promoter and/or a promoter that provides for a different pattern of expression in the cell or microorganism. An example of one such heterologous promoter is a *Candida* oleate-induced peroxisomal protein (*POX18*) gene promoter. Another example of one such heterologous promoter is a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene promoter. The nucleotide sequences of a *Candida viswanathii* HDE gene promoter and a *Candida* POX18 gene promoter are provided herein as are examples of additional fatty acid-inducible promoters. In some embodiments, the promoter is modified by replacing the endogenous promoter with a promoter comprising an alkane response element (ARE1). Non-limiting examples of assays suitable for assessing induction of HMG-CoA synthase expression by a carbon source or feedstock include RT-PCR or qRT-PCR after the host cell or microorganism has been exposed to the chosen carbon source or feedstock for varying amounts of time.

Decreasing HMG-CoA synthase activity in a cell can be accomplished by modifying the amount of HMG-CoA synthase protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous HMG-CoA synthase gene in a cell or organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type HMG-CoA synthase such that the encoded modified or substituted HMG-CoA synthase protein has a reduced enzyme activity.

Modification of an HMG-CoA reductase activity

A further (rate controlling) step in the upper mevalonate pathway is the conversion of HMG-CoA to mevalonate by HMG-CoA reductase. HMG-CoA reductase (EC 1.1.1.34) may also be referred to as HMG1, 3-hydroxy-3-methyl-glutaryl-coenzyme A reductase, HMGCR, LDLCQ3, 3-hydroxy-3-methylglutaryl-CoA reductase, and hydroxymethylglutaryl-CoA reductase. In some embodiments of the cells, organisms, compositions and methods provided herein, the amount and/or activity of an HMG-CoA reductase in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase the amount and/or activity of an HMG-CoA reductase, may be modified to decrease the amount and/or activity of an HMG-CoA reductase, or may be modified to alternately increase and decrease the amount and/or activity of an HMG-

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1329 of 3077

CoA reductase depending, for example, on the cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

In certain aspects, the amount and/or activity of an HMG-CoA reductase in a cell is increased. Increasing the amount and/or activity of an HMG-CoA reductase may be particularly beneficial in embodiments in which the flux of carbons from fatty acids is directed toward a particular target product pathway (e.g., mevalonate production, isopentenyl diphosphate production, geranyl diphosphate production, farnesyl diphosphate production, geranylgeranyl diphosphate production, terpene production, carotenoid production, beta carotene production, lycopene production, astaxanthin production) and away from other cellular metabolic pathways not involved in target molecule production.

In certain embodiments, HMG-CoA reductase activity is unchanged in a host or engineered cell or organism. In some embodiments, the amount and/or activity of a host HMG-CoA reductase can be increased by increasing the number of copies of a nucleic acid encoding an HMG-CoA reductase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the nucleic acid); by increasing the activity of a promoter that regulates transcription of a nucleic acid encoding an HMG-CoA reductase; or by increasing the number of copies of a nucleic acid encoding an HMG-CoA reductase and increasing the activity of a promoter that regulates transcription of a nucleic acid encoding an HMG-CoA reductase. In some embodiments, an HMG-CoA reductase is endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, the amount of HMG-CoA reductase protein expressed in a cell can be increased by introducing heterologous nucleic acid encoding an HMG-CoA reductase into a cell or microorganism. For example, introduction of heterologous nucleic acid encoding an HMG-CoA reductase can result in increased copy number of such nucleic acids and/or provide for modification of the cellular location in which the protein is expressed.

In some embodiments, an HMG-CoA reductase enzyme can be a fungal or bacterial protein. In a particular embodiment, the HMG-CoA reductase enzyme can be a *Candida* (e.g., *C. tropicalis*, *C. viswanathii*, *C. maltosa*, *C. cloacae*), *Yarrowia* (e.g., *Y. lipolytica*), *Fusarium* (e.g., *F. oxysporum*), *Bacillus* (e.g., *B. megaterium*, *B. subtilis*), *Chronobacter* (e.g., *C. sakazakii*), *Xanthophyllomyces* (e.g., *X. dendrorhous*), or *Pantoea* (e.g., *P. ananatis*) protein. An example of a *Candida viswanathii* nucleotide sequence (SEQ ID NO: 373) encoding a polypeptide (*HMG1*; SEQ ID NO: 339) having an HMG-CoA reductase activity is provided herein. An

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1330 of 3077

additional nonlimiting example of a nucleotide sequence encoding a polypeptide having HMG-CoA reductase activity includes: *Candida tropicalis MYA-3404* 3-hydroxy-3-methylglutaryl-coenzyme A reductase (NCBI Reference Sequence: XM_002550004.1).

Presence, absence or amount of HMG-CoA reductase activity can be detected by any suitable method known in the art. For example, detection can be performed by using an HMG-CoA Reductase Assay Kit (e.g., SIGMA-ALDRICH, CS1090 Sigma); and/or a Colorimetric HMG-CoA Reductase Activity Assay Kit (ABCAM, ab204701). Nucleic acid sequences encoding native and/or modified HMG-CoA reductase sequences also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cell or organism exhibits increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism.

The promoter used for regulating transcription of a nucleic acid (e.g., a heterologous nucleic acid) encoding an HMG-CoA reductase can be modified. For example, the amount of an HMG-CoA reductase protein expressed in a particular cellular location may be increased by including in the nucleic acid (e.g., heterologous nucleic acid) a strong heterologous promoter and/or a promoter that provides for a different pattern of expression in the cell or microorganism. An example of one such heterologous promoter is a *Candida* oleate-induced peroxisomal protein (*POX18*) gene promoter. Another example of one such heterologous promoter is a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene promoter. The nucleotide sequences of a *Candida viswanathii* HDE gene promoter and a *Candida* POX18 gene promoter are provided herein as are examples of additional fatty acid-inducible promoters. In some embodiments, the promoter is modified by replacing the endogenous promoter with a promoter comprising an alkane response element (ARE1). Non-limiting examples of assays suitable for assessing induction of HMG-CoA reductase expression by a carbon source or feedstock include RT-PCR or qRT-PCR after the host cell or microorganism has been exposed to the chosen carbon source or feedstock for varying amounts of time.

Decreasing HMG-CoA reductase activity in a cell can be accomplished by modifying the amount of HMG-CoA reductase protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous HMG-CoA reductase gene in a cell or organism with a weaker

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1331 of 3077

heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type HMG-CoA reductase such that the encoded modified or substituted HMG-CoA reductase protein has a reduced enzyme activity.

Modification of a mevalonate kinase activity

The initial step in the lower mevalonate pathway is the conversion of mevalonate to phosphomevalonate by mevalonate kinase. Mevalonate kinase (EC 2.7.1.36) may also be referred to as ERG12, MVK, LRBP, MK, MVLK, and POROK3. In some embodiments of the cells, organisms, compositions and methods provided herein, the amount and/or activity of a mevalonate kinase in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase the amount and/or activity of a mevalonate kinase, may be modified to decrease the amount and/or activity of a mevalonate kinase, or may be modified to alternately increase and decrease the amount and/or activity of a mevalonate kinase depending, for example, on the cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

In certain aspects, the amount and/or activity of a mevalonate kinase in a cell is increased. Increasing the amount and/or activity of a mevalonate kinase may be particularly beneficial in embodiments in which the flux of carbons from fatty acids is directed toward a particular target product pathway (e.g., isopentenyl diphosphate production, geranyl diphosphate production, farnesyl diphosphate production, geranylgeranyl diphosphate production, terpene production, carotenoid production, beta carotene production, lycopene production, astaxanthin production) and away from other cellular metabolic pathways not involved in target molecule production.

In certain embodiments, mevalonate kinase activity is unchanged in a host or engineered cell or organism. In some embodiments, the amount and/or activity of a host mevalonate kinase can be increased by increasing the number of copies of a nucleic acid encoding a mevalonate kinase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the nucleic acid); by increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a mevalonate kinase; or by increasing the number of copies of a nucleic acid encoding a mevalonate kinase and increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a mevalonate kinase. In some embodiments, a mevalonate kinase is endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1332 of 3077

and methods provided herein, the amount of mevalonate kinase protein expressed in a cell can be increased by introducing heterologous nucleic acid encoding a mevalonate kinase into a cell or microorganism. For example, introduction of heterologous nucleic acid encoding a mevalonate kinase can result in increased copy number of such nucleic acids and/or provide for modification of the cellular location in which the protein is expressed.

In some embodiments, a mevalonate kinase enzyme can be a fungal or bacterial protein. In a particular embodiment, the mevalonate kinase enzyme can be a *Candida* (e.g., *C. tropicalis*, *C. viswanathii*, *C. maltosa*, *C. cloacae*), *Yarrowia* (e.g., *Y. lipolytica*), *Fusarium* (e.g., *F. oxysporum*), *Bacillus* (e.g., *B. megaterium*, *B. subtilis*), *Chronobacter* (e.g., *C. sakazakii*), *Xanthophyllomyces* (e.g., *X. dendrorhous*), or *Pantoea* (e.g., *P. ananatis*) protein. An example of a *Candida viswanathii* nucleotide sequence (SEQ ID NO: 374) encoding a polypeptide (*ERG12*; SEQ ID NO: 340) having a mevalonate kinase activity is provided herein.

Presence, absence or amount of mevalonate kinase activity can be detected by any suitable method known in the art. For example, detection can be performed by using a mevalonate kinase assay (see e.g., Green et al. (1970) Anal. Biochem. 38(1):130-138; Gibson et al. (1989) Enzyme 41(1):47-55). Nucleic acid sequences encoding native and/or modified mevalonate kinase sequences also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cell or organism exhibits increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism.

The promoter used for regulating transcription of a nucleic acid (e.g., a heterologous nucleic acid) encoding a mevalonate kinase can be modified. For example, the amount of a mevalonate kinase protein expressed in a particular cellular location may be increased by including in the nucleic acid (e.g., heterologous nucleic acid) a strong heterologous promoter and/or a promoter that provides for a different pattern of expression in the cell or microorganism. An example of one such heterologous promoter is a *Candida* oleate-induced peroxisomal protein (*POX18*) gene promoter. Another example of one such heterologous promoter is a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene promoter. The nucleotide sequences of a *Candida viswanathii* HDE gene promoter and a *Candida* POX18 gene promoter are provided herein as are examples of additional fatty acid-inducible promoters. In some

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1333 of 3077

embodiments, the promoter is modified by replacing the endogenous promoter with a promoter comprising an alkane response element (ARE1). Non-limiting examples of assays suitable for assessing induction of mevalonate kinase expression by a carbon source or feedstock include RT-PCR or qRT-PCR after the host cell or microorganism has been exposed to the chosen carbon source or feedstock for varying amounts of time.

Decreasing mevalonate kinase activity in a cell can be accomplished by modifying the amount of mevalonate kinase protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous mevalonate kinase gene in a cell or organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type mevalonate kinase such that the encoded modified or substituted mevalonate kinase protein has a reduced enzyme activity.

Modification of a phosphomevalonate kinase activity

A further step in the lower mevalonate pathway is the conversion of phosphomevalonate to diphosphomevalonate (mevalonate-5-phosphate to mevalonate-5-pyrophosphate; 5-phosphomevalonate to 5-diphosphomevalonate) by phosphomevalonate kinase. Phosphomevalonate kinase (ERG8; EC 2.7.4.2) may also be referred to as 5-phosphomevalonate phosphotransferase and PMVK. In some embodiments of the cells, organisms, compositions and methods provided herein, the amount and/or activity of a phosphomevalonate kinase in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase the amount and/or activity of a phosphomevalonate kinase, may be modified to decrease the amount and/or activity of a phosphomevalonate kinase, or may be modified to alternately increase and decrease the amount and/or activity of a phosphomevalonate kinase depending, for example, on the cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

In certain aspects, the amount and/or activity of a phosphomevalonate kinase in a cell is increased. Increasing the amount and/or activity of a phosphomevalonate kinase may be particularly beneficial in embodiments in which the flux of carbons from fatty acids is directed toward a particular target product pathway (e.g., isopentenyl diphosphate production, geranyl diphosphate production, farnesyl diphosphate production, geranylgeranyl diphosphate production, terpene production, carotenoid production, beta carotene production, lycopene

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1334 of 3077

production, astaxanthin production) and away from other cellular metabolic pathways not involved in target molecule production.

In certain embodiments, phosphomevalonate kinase activity is unchanged in a host or engineered cell or organism. In some embodiments, the amount and/or activity of a host phosphomevalonate kinase can be increased by increasing the number of copies of a nucleic acid encoding a phosphomevalonate kinase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the nucleic acid); by increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a phosphomevalonate kinase; or by increasing the number of copies of a nucleic acid encoding a phosphomevalonate kinase and increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a phosphomevalonate kinase. In some embodiments, a phosphomevalonate kinase is endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, the amount of phosphomevalonate kinase protein expressed in a cell can be increased by introducing heterologous nucleic acid encoding a phosphomevalonate kinase into a cell or microorganism. For example, introduction of heterologous nucleic acid encoding an phosphomevalonate kinase can result in increased copy number of such nucleic acids and/or provide for modification of the cellular location in which the protein is expressed.

In some embodiments, a phosphomevalonate kinase enzyme can be a fungal or bacterial protein. In a particular embodiment, the phosphomevalonate kinase enzyme can be a *Candida* (e.g., *C. tropicalis*, *C. viswanathii*, *C. maltosa*, *C. cloacae*), *Yarrowia* (e.g., *Y. lipolytica*), *Fusarium* (e.g., *F. oxysporum*), *Bacillus* (e.g., *B. megaterium*, *B. subtilis*), *Chronobacter* (e.g., *C. sakazakii*), *Xanthophyllomyces* (e.g., *X. dendrorhous*), or *Pantoea* (e.g., *P. ananatis*) protein. An example of a *Candida viswanathii* nucleotide sequence (SEQ ID NO: 375) encoding a polypeptide (*ERG8*; SEQ ID NO: 341) having a phosphomevalonate kinase activity is provided herein.

Presence, absence or amount of phosphomevalonate kinase activity can be detected by any suitable method known in the art. For example, detection can be performed by using a microplate assay for phosphomevalonate kinase activity (see e.g., Schulte et al. (1999) Anal. Biochem. 269(2):245-54). Nucleic acid sequences encoding native and/or modified phosphomevalonate kinase sequences also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1335 of 3077

thereof), or quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cell or organism exhibits increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism.

The promoter used for regulating transcription of a nucleic acid (e.g., a heterologous nucleic acid) encoding a phosphomevalonate kinase can be modified. For example, the amount of a phosphomevalonate kinase protein expressed in a particular cellular location may be increased by including in the nucleic acid (e.g., heterologous nucleic acid) a strong heterologous promoter and/or a promoter that provides for a different pattern of expression in the cell or microorganism. An example of one such heterologous promoter is a *Candida* oleate-induced peroxisomal protein (*POX18*) gene promoter. Another example of one such heterologous promoter is a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene promoter. The nucleotide sequences of a *Candida viswanathii* HDE gene promoter and a *Candida* POX18 gene promoter are provided herein as are examples of additional fatty acid-inducible promoters. In some embodiments, the promoter is modified by replacing the endogenous promoter with a promoter comprising an alkane response element (ARE1). Non-limiting examples of assays suitable for assessing induction of phosphomevalonate kinase expression by a carbon source or feedstock include RT-PCR or qRT-PCR after the host cell or microorganism has been exposed to the chosen carbon source or feedstock for varying amounts of time.

Decreasing phosphomevalonate kinase activity in a cell can be accomplished by modifying the amount of phosphomevalonate kinase protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous phosphomevalonate kinase gene in a cell or organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type phosphomevalonate kinase such that the encoded modified or substituted phosphomevalonate kinase protein has a reduced enzyme activity.

Modification of a diphosphomevalonate decarboxylase activity

A further step in the lower mevalonate pathway is the conversion of mevalonate 5-diphosphate (diphosphomevalonate; mevalonate-5-pyrophosphate; 5-diphosphomevalonate) to isopentenyl diphosphate. A diphosphomevalonate decarboxylase enzyme (also referred to as MVD1,

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1336 of 3077

ERG19, pyrophosphomevalonate decarboxylase, mevalonate-5-pyrophosphate decarboxylase, pyrophosphomevalonic acid decarboxylase, 5-pyrophosphomevalonate decarboxylase, mevalonate 5-diphosphate decarboxylase, and ATP:(R)-5-diphosphomevalonate carboxy-lyase (dehydrating); EC 4.1.1.33) can convert mevalonate 5-diphosphate to isopentenyl diphosphate through ATP dependent decarboxylation. In some embodiments of the cells, organisms, compositions and methods provided herein, the amount and/or activity of a diphosphomevalonate decarboxylase in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase the amount and/or activity of a diphosphomevalonate decarboxylase, may be modified to decrease the amount and/or activity of a diphosphomevalonate decarboxylase, or may be modified to alternately increase and decrease the amount and/or activity of a diphosphomevalonate decarboxylase depending, for example, on the cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

In certain aspects, the amount and/or activity of a diphosphomevalonate decarboxylase in a cell is increased. Increasing the amount and/or activity of a diphosphomevalonate decarboxylase may be particularly beneficial in embodiments in which the flux of carbons from fatty acids is directed toward a particular target product pathway (e.g., isopentenyl diphosphate production, geranyl diphosphate production, farnesyl diphosphate production, geranylgeranyl diphosphate production, terpene production, carotenoid production, beta carotene production, lycopene production, astaxanthin production) and away from other cellular metabolic pathways not involved in target molecule production.

In certain embodiments, diphosphomevalonate decarboxylase activity is unchanged in a host or engineered cell or organism. In some embodiments, the amount and/or activity of a host diphosphomevalonate decarboxylase can be increased by increasing the number of copies of a nucleic acid encoding a diphosphomevalonate decarboxylase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the nucleic acid); by increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a diphosphomevalonate decarboxylase; or by increasing the number of copies of a nucleic acid encoding a diphosphomevalonate decarboxylase and increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a diphosphomevalonate decarboxylase. In some embodiments, a diphosphomevalonate decarboxylase is endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, the amount of

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1337 of 3077

diphosphomevalonate decarboxylase protein expressed in a cell can be increased by introducing heterologous nucleic acid encoding a diphosphomevalonate decarboxylase into a cell or microorganism. For example, introduction of heterologous nucleic acid encoding a diphosphomevalonate decarboxylase can result in increased copy number of such nucleic acids and/or provide for modification of the cellular location in which the protein is expressed.

In some embodiments, a diphosphomevalonate decarboxylase enzyme can be a fungal or bacterial protein. In a particular embodiment, the diphosphomevalonate decarboxylase enzyme can be a *Candida* (e.g., *C. tropicalis*, *C. viswanathii*, *C. maltosa*, *C. cloacae*), *Yarrowia* (e.g., *Y. lipolytica*), *Fusarium* (e.g., *F. oxysporum*), *Bacillus* (e.g., *B. megaterium*, *B. subtilis*), *Chronobacter* (e.g., *C. sakazakii*), *Xanthophyllomyces* (e.g., *X. dendrorhous*), or *Pantoea* (e.g., *P. ananatis*) protein. An example of a *Candida viswanathii* nucleotide sequence (SEQ ID NO: 376) encoding a polypeptide (*MVD1* (*ERG19)*; SEQ ID NO: 342) having a diphosphomevalonate decarboxylase activity is provided herein. Additional nonlimiting examples of nucleotide sequences encoding polypeptides having diphosphomevalonate decarboxylase activity include: *Candida tropicalis MYA-3404* diphosphomevalonate decarboxylase (NCBI Reference Sequence: XM_002549937.1) and *Candida tanzawaensis NRRL Y-17324* Diphosphomevalonate decarboxylase (NCBI Reference Sequence: XM_020210701.1).

Presence, absence or amount of diphosphomevalonate decarboxylase activity can be detected by any suitable method known in the art. Nucleic acid sequences encoding native and/or modified diphosphomevalonate decarboxylase sequences also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cell or organism exhibits increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism.

The promoter used for regulating transcription of a nucleic acid (e.g., a heterologous nucleic acid) encoding a diphosphomevalonate decarboxylase can be modified. For example, the amount of a diphosphomevalonate decarboxylase protein expressed in a particular cellular location may be increased by including in the nucleic acid (e.g., heterologous nucleic acid) a strong heterologous promoter and/or a promoter that provides for a different pattern of

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1338 of 3077

expression in the cell or microorganism. An example of one such heterologous promoter is a *Candida* oleate-induced peroxisomal protein (*POX18*) gene promoter. Another example of one such heterologous promoter is a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene promoter. The nucleotide sequences of a *Candida viswanathii* HDE gene promoter and a *Candida* POX18 gene promoter are provided herein as are examples of additional fatty acid-inducible promoters. In some embodiments, the promoter is modified by replacing the endogenous promoter with a promoter comprising an alkane response element (ARE1). Non-limiting examples of assays suitable for assessing induction of diphosphomevalonate decarboxylase expression by a carbon source or feedstock include RT-PCR or qRT-PCR after the host cell or microorganism has been exposed to the chosen carbon source or feedstock for varying amounts of time.

Decreasing diphosphomevalonate decarboxylase activity in a cell can be accomplished by modifying the amount of diphosphomevalonate decarboxylase protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous diphosphomevalonate decarboxylase gene in a cell or organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type diphosphomevalonate decarboxylase such that the encoded modified or substituted diphosphomevalonate decarboxylase protein has a reduced enzyme activity.

Modification of an isopentenyl diphosphate delta isomerase activity

A further step in the lower mevalonate pathway is the conversion of isopentenyl diphosphate (isopentenyl pyrophosphate, IPP) to dimethylallyl diphosphate (dimethylallyl pyrophosphate, DMAPP). An isopentenyl diphosphate delta isomerase enzyme (also referred to as IDI1, isopentenyl pyrophosphate isomerase, IPP isomerase; EC 5.3.3.2) can catalyze the isomerization of isopentenyl diphosphate to dimethylallyl diphosphate by an antarafacial transposition of hydrogen. In some embodiments of the cells, organisms, compositions and methods provided herein, the amount and/or activity of an isopentenyl diphosphate delta isomerase in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase the amount and/or activity of an isopentenyl diphosphate delta isomerase, may be modified to decrease the amount and/or activity of an isopentenyl diphosphate delta isomerase, or may be modified to alternately increase and decrease the amount and/or activity of an isopentenyl diphosphate delta isomerase depending, for example, on the cellular

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1339 of 3077

location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

In certain aspects, the amount and/or activity of an isopentenyl diphosphate delta isomerase in a cell is increased. Increasing the amount and/or activity of an isopentenyl diphosphate delta isomerase may be particularly beneficial in embodiments in which the flux of carbons from fatty acids is directed toward a particular target product pathway (e.g., geranyl diphosphate production, farnesyl diphosphate production, geranylgeranyl diphosphate production, terpene production, carotenoid production, beta carotene production, lycopene production, astaxanthin production) and away from other cellular metabolic pathways not involved in target molecule production.

In certain embodiments, isopentenyl diphosphate delta isomerase activity is unchanged in a host or engineered cell or organism. In some embodiments, the amount and/or activity of a host isopentenyl diphosphate delta isomerase can be increased by increasing the number of copies of a nucleic acid encoding an isopentenyl diphosphate delta isomerase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the nucleic acid); by increasing the activity of a promoter that regulates transcription of a nucleic acid encoding an isopentenyl diphosphate delta isomerase; or by increasing the number of copies of a nucleic acid encoding an isopentenyl diphosphate delta isomerase and increasing the activity of a promoter that regulates transcription of a nucleic acid encoding an isopentenyl diphosphate delta isomerase. In some embodiments, an isopentenyl diphosphate delta isomerase is endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, the amount of isopentenyl diphosphate delta isomerase protein expressed in a cell can be increased by introducing heterologous nucleic acid encoding an isopentenyl diphosphate delta isomerase into a cell or microorganism. For example, introduction of heterologous nucleic acid encoding an isopentenyl diphosphate delta isomerase can result in increased copy number of such nucleic acids and/or provide for modification of the cellular location in which the protein is expressed.

In some embodiments, an isopentenyl diphosphate delta isomerase enzyme can be a fungal or bacterial protein. In a particular embodiment, the isopentenyl diphosphate delta isomerase enzyme can be a *Candida* (e.g., *C. tropicalis*, *C. viswanathii*, *C. maltosa*, *C. cloacae*), *Yarrowia* (e.g., *Y. lipolytica*), *Fusarium* (e.g., *F. oxysporum*), *Bacillus* (e.g., *B. megaterium*, *B. subtilis*),

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1340 of 3077

Chronobacter (e.g., *C. sakazakii*), *Xanthophyllomyces* (e.g., *X. dendrorhous*), or *Pantoea* (e.g., *P. ananatis*) protein. An example of a *Candida viswanathii* nucleotide sequence (SEQ ID NO: 377) encoding a polypeptide (*IDI1*; SEQ ID NO: 343) having an isopentenyl diphosphate delta isomerase activity is provided herein. An additional nonlimiting example of a nucleotide sequence encoding a polypeptide having isopentenyl diphosphate delta isomerase activity is a *Candida tropicalis MYA-3404* isopentenyl-diphosphate delta-isomerase (NCBI Reference Sequence: XM_002545339.1).

Presence, absence or amount of isopentenyl diphosphate delta isomerase activity can be detected by any suitable method known in the art. For example, detection can be performed by using an isopentenyl diphosphate delta isomerase activity assay (see e.g., Diaz et al. (2012) Insect Biochem. Mol. Bio. 42(10):751-757). Nucleic acid sequences encoding native and/or modified isopentenyl diphosphate delta isomerase sequences also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cell or organism exhibits increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism.

The promoter used for regulating transcription of a nucleic acid (e.g., a heterologous nucleic acid) encoding an isopentenyl diphosphate delta isomerase can be modified. For example, the amount of an isopentenyl diphosphate delta isomerase protein expressed in a particular cellular location may be increased by including in the nucleic acid (e.g., heterologous nucleic acid) a strong heterologous promoter and/or a promoter that provides for a different pattern of expression in the cell or microorganism. An example of one such heterologous promoter is a *Candida* oleate-induced peroxisomal protein (*POX18*) gene promoter. Another example of one such heterologous promoter is a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene promoter. The nucleotide sequences of a *Candida viswanathii* HDE gene promoter and a *Candida* POX18 gene promoter are provided herein as are examples of additional fatty acid-inducible promoters. In some embodiments, the promoter is modified by replacing the endogenous promoter with a promoter comprising an alkane response element (ARE1). Non-limiting examples of assays suitable for assessing induction of isopentenyl diphosphate delta isomerase expression by a carbon source or feedstock include RT-PCR or qRT-PCR after the

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1341 of 3077

host cell or microorganism has been exposed to the chosen carbon source or feedstock for varying amounts of time.

Decreasing isopentenyl diphosphate delta isomerase activity in a cell can be accomplished by modifying the amount of isopentenyl diphosphate delta isomerase protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous isopentenyl diphosphate delta isomerase gene in a cell or organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type isopentenyl diphosphate delta isomerase such that the encoded modified or substituted isopentenyl diphosphate delta isomerase protein has a reduced enzyme activity.

Modification of a terpene synthase activity

Provided herein are one or more modifications to one or more terpene synthase activities. Terpene synthases may include one or more enzymes in a family of enzymes involved in terpene biosynthesis, such as, for example, terpene synthases known in the art and described in Yamada et al. (2015) PNAS 112(3):857-862. In some embodiments of the cells, organisms, compositions and methods provided herein, the amount and/or activity of a terpene synthase in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase the amount and/or activity of a terpene synthase, may be modified to decrease the amount and/or activity of a terpene synthase, or may be modified to alternately increase and decrease the amount and/or activity of a terpene synthase depending, for example, on the cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

In certain aspects, the amount and/or activity of a terpene synthase in a cell is increased. Increasing the amount and/or activity of a terpene synthase may be particularly beneficial in embodiments in which the flux of carbons from fatty acids is directed toward a particular target product pathway (e.g., mevalonate production, isopentenyl diphosphate production, geranyl diphosphate production, farnesyl diphosphate production, geranylgeranyl diphosphate production, terpene production, carotenoid production, beta carotene production, lycopene production, astaxanthin production) and away from other cellular metabolic pathways not involved in target molecule production.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1342 of 3077

In certain embodiments, terpene synthase activity is unchanged in a host or engineered cell or organism. In some embodiments, the amount and/or activity of a host terpene synthase can be increased by increasing the number of copies of a nucleic acid encoding a terpene synthase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the nucleic acid); by increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a terpene synthase; or by increasing the number of copies of a nucleic acid a terpene synthase and increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a terpene synthase. In some embodiments, a terpene synthase is endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, the amount of terpene synthase protein expressed in a cell can be increased by introducing heterologous nucleic acid encoding a terpene synthase into a cell or microorganism. For example, introduction of heterologous nucleic acid encoding a terpene synthase can result in increased copy number of such nucleic acids and/or provide for modification of the cellular location in which the protein is expressed. In some embodiments, a terpene synthase is not endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, a terpene synthase protein can be expressed in a cell by introducing heterologous nucleic acid encoding a terpene synthase into a cell or microorganism.

In some embodiments, a terpene synthase enzyme can be a fungal or bacterial protein. In a particular embodiment, the terpene synthase enzyme can be a *Candida* (e.g., *C. tropicalis*, *C. viswanathii*, *C. maltosa*, *C. cloacae*), *Yarrowia* (e.g., *Y. lipolytica*), *Fusarium* (e.g., *F. oxysporum*), *Bacillus* (e.g., *B. megaterium*, *B. subtilis*), *Chronobacter* (e.g., *C. sakazakii*), *Xanthophyllomyces* (e.g., *X. dendrorhous*), or *Pantoea* (e.g., *P. ananatis*) protein.

Presence, absence or amount of terpene synthase activity can be detected by any suitable method known in the art. For example, detection can be performed by using an appropriate terpene synthase activity assay. Nucleic acid sequences encoding native and/or modified terpene synthase sequences also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cell or organism exhibits increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1343 of 3077

The promoter used for regulating transcription of a nucleic acid (e.g., a heterologous nucleic acid) encoding a terpene synthase can be modified. For example, the amount of a terpene synthase protein expressed in a particular cellular location may be increased by including in the nucleic acid (e.g., heterologous nucleic acid) a strong heterologous promoter and/or a promoter that provides for a different pattern of expression in the cell or microorganism. An example of one such heterologous promoter is a *Candida* oleate-induced peroxisomal protein (*POX18*) gene promoter. Another example of one such heterologous promoter is a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene promoter. The nucleotide sequences of a *Candida viswanathii* HDE gene promoter and a *Candida* POX18 gene promoter are provided herein as are examples of additional fatty acid-inducible promoters. In some embodiments, the promoter is modified by replacing the endogenous promoter with a promoter comprising an alkane response element (ARE1). Non-limiting examples of assays suitable for assessing induction of terpene synthase expression by a carbon source or feedstock include RT-PCR or qRT-PCR after the host cell or microorganism has been exposed to the chosen carbon source or feedstock for varying amounts of time.

Decreasing terpene synthase activity in a cell can be accomplished by modifying the amount of terpene synthase protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous terpene synthase gene in a cell or organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type terpene synthase such that the encoded modified or substituted terpene synthase protein has a reduced enzyme activity.

Modification of a dimethylallyltranstransferase activity

One component of terpene biosynthesis is the conversion of dimethylallyl diphosphate (dimethylallyl pyrophosphate) and isopentenyl diphosphate (isopentenyl pyrophosphate) into farnesyl diphosphate (farnesyl pyrophosphate). A dimethylallyltranstransferase enzyme (EC 2.5.1.1) can convert dimethylallyl diphosphate and isopentenyl diphosphate into farnesyl diphosphate. Dimethylallyltranstransferase may also be referred to as ERG20, farnesyl pyrophosphate synthase, farnesyl diphosphate synthase, or farnesyl diphosphate synthetase (EC 2.5.1.10). Geranyl diphosphate (geranyl pyrophosphate) often is created in an intermediate step. In some embodiments of the cells, organisms, compositions and methods provided herein, the amount and/or activity of a dimethylallyltranstransferase in a cell is modified. For

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1344 of 3077

example, in some aspects, a cell or microorganism may be modified to increase the amount and/or activity of a dimethylallyltranstransferase, may be modified to decrease the amount and/or activity of a dimethylallyltranstransferase, or may be modified to alternately increase and decrease the amount and/or activity of a dimethylallyltranstransferase depending, for example, on the cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

In certain aspects, the amount and/or activity of a dimethylallyltranstransferase in a cell is increased. Increasing the amount and/or activity of a dimethylallyltranstransferase may be particularly beneficial in embodiments in which the flux of carbons from fatty acids is directed toward a particular target product pathway (e.g., geranyl diphosphate production, farnesyl diphosphate production, geranylgeranyl diphosphate production, terpene production, carotenoid production, beta carotene production, lycopene production, astaxanthin production) and away from other cellular metabolic pathways not involved in target molecule production.

In certain embodiments, dimethylallyltranstransferase activity is unchanged in a host or engineered cell or organism. In some embodiments, the amount and/or activity of a host dimethylallyltranstransferase can be increased by increasing the number of copies of a nucleic acid encoding a dimethylallyltranstransferase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the nucleic acid); by increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a dimethylallyltranstransferase; or by increasing the number of copies of a nucleic acid encoding a dimethylallyltranstransferase and increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a dimethylallyltranstransferase. In some embodiments, a dimethylallyltranstransferase is endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, the amount of dimethylallyltranstransferase protein expressed in a cell can be increased by introducing heterologous nucleic acid encoding a dimethylallyltranstransferase into a cell or microorganism. For example, introduction of heterologous nucleic acid encoding a dimethylallyltranstransferase can result in increased copy number of such nucleic acids and/or provide for modification of the cellular location in which the protein is expressed. In some embodiments, a dimethylallyltranstransferase is not endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, a dimethylallyltranstransferase protein can be expressed in a cell by

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1345 of 3077

introducing heterologous nucleic acid encoding a dimethylallyltranstransferase into a cell or microorganism.

In some embodiments, a dimethylallyltranstransferase enzyme can be a fungal or bacterial protein. In a particular embodiment, the dimethylallyltranstransferase enzyme can be a *Candida* (e.g., *C. tropicalis*, *C. viswanathii*, *C. maltosa*, *C. cloacae*), *Yarrowia* (e.g., *Y. lipolytica*), *Fusarium* (e.g., *F. oxysporum*), *Bacillus* (e.g., *B. megaterium*, *B. subtilis*), *Chronobacter* (e.g., *C. sakazakii*), *Xanthophyllomyces* (e.g., *X. dendrorhous*), or *Pantoea* (e.g., *P. ananatis*) protein. An example of a *Candida viswanathii* nucleotide sequence (SEQ ID NO: 378) encoding a polypeptide (*ERG20*; SEQ ID NO: 344) having a dimethylallyltranstransferase activity is provided herein. An additional nonlimiting example of a nucleotide sequence encoding a polypeptide having dimethylallyltranstransferase activity is *Candida tropicalis MYA-3404* farnesyl pyrophosphate synthetase (NCBI Reference Sequence: XM_002547254.1).

Presence, absence or amount of dimethylallyltranstransferase activity can be detected by any suitable method known in the art. For example, detection can be performed by using a dimethylallyltranstransferase ELISA kit (e.g., MYBIOSOURCE, MBS943684). Nucleic acid sequences encoding native and/or modified dimethylallyltranstransferase sequences also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cell or organism exhibits increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism.

The promoter used for regulating transcription of a nucleic acid (e.g., a heterologous nucleic acid) encoding a dimethylallyltranstransferase can be modified. For example, the amount of a dimethylallyltranstransferase protein expressed in a particular cellular location may be increased by including in the nucleic acid (e.g., heterologous nucleic acid) a strong heterologous promoter and/or a promoter that provides for a different pattern of expression in the cell or microorganism. An example of one such heterologous promoter is a *Candida* oleate-induced peroxisomal protein (*POX18*) gene promoter. Another example of one such heterologous promoter is a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene promoter. The nucleotide sequences of a *Candida viswanathii* HDE gene promoter and a *Candida* POX18 gene promoter are provided herein as are examples of additional fatty acid-inducible promoters. In some

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1346 of 3077

embodiments, the promoter is modified by replacing the endogenous promoter with a promoter comprising an alkane response element (ARE1). Non-limiting examples of assays suitable for assessing induction of dimethylallyltranstransferase expression by a carbon source or feedstock include RT-PCR or qRT-PCR after the host cell or microorganism has been exposed to the chosen carbon source or feedstock for varying amounts of time.

Decreasing dimethylallyltranstransferase activity in a cell can be accomplished by modifying the amount of dimethylallyltranstransferase protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous dimethylallyltranstransferase gene in a cell or organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type dimethylallyltranstransferase such that the encoded modified or substituted dimethylallyltranstransferase protein has a reduced enzyme activity.

Modification of a geranylgeranyl diphosphate synthase activity

Another component of terpene biosynthesis is the synthesis of geranylgeranyl diphosphate from farnesyl diphosphate and isopentenyl diphosphate. A geranylgeranyl diphosphate synthase enzyme (also referred to as BTS1, CrtE, GGPS1, GGPPS, GGPPS1, geranylgeranyl diphosphate synthase 1; EC 2.5.1.29) can catalyze the synthesis of geranylgeranyl diphosphate from farnesyl diphosphate and isopentenyl diphosphate. In some embodiments of the cells, organisms, compositions and methods provided herein, the amount and/or activity of a geranylgeranyl diphosphate synthase in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase the amount and/or activity of a geranylgeranyl diphosphate synthase, may be modified to decrease the amount and/or activity of geranylgeranyl diphosphate synthase, or may be modified to alternately increase and decrease the amount and/or activity of an geranylgeranyl diphosphate synthase depending, for example, on the cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

In certain aspects, the amount and/or activity of a geranylgeranyl diphosphate synthase in a cell is increased. Increasing the amount and/or activity of a geranylgeranyl diphosphate synthase may be particularly beneficial in embodiments in which the flux of carbons from fatty acids is directed toward a particular target product pathway (e.g., mevalonate production, isopentenyl

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1347 of 3077

diphosphate production, geranyl diphosphate production, farnesyl diphosphate production, geranylgeranyl diphosphate production, terpene production, carotenoid production, beta carotene production, lycopene production, astaxanthin production) and away from other cellular metabolic pathways not involved in target molecule production.

In certain embodiments, geranylgeranyl diphosphate synthase activity is unchanged in a host or engineered cell or organism. In some embodiments, the amount and/or activity of a host geranylgeranyl diphosphate synthase can be increased by increasing the number of copies of a nucleic acid encoding a geranylgeranyl diphosphate synthase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the nucleic acid); by increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a geranylgeranyl diphosphate synthase; or by increasing the number of copies of a nucleic acid encoding a geranylgeranyl diphosphate synthase and increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a geranylgeranyl diphosphate synthase. In some embodiments, a geranylgeranyl diphosphate synthase is endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, the amount of geranylgeranyl diphosphate synthase protein expressed in a cell can be increased by introducing heterologous nucleic acid encoding a geranylgeranyl diphosphate synthase into a cell or microorganism. For example, introduction of heterologous nucleic acid encoding a geranylgeranyl diphosphate synthase can result in increased copy number of such nucleic acids and/or provide for modification of the cellular location in which the protein is expressed. In some embodiments, a geranylgeranyl diphosphate synthase is not endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, a geranylgeranyl diphosphate synthase protein can be expressed in a cell by introducing heterologous nucleic acid encoding a geranylgeranyl diphosphate synthase into a cell or microorganism.

In some embodiments, a geranylgeranyl diphosphate synthase enzyme can be a fungal or bacterial protein. In a particular embodiment, the geranylgeranyl diphosphate synthase enzyme can be a *Candida* (e.g., *C. tropicalis*, *C. viswanathii*, *C. maltosa*, *C. cloacae*), *Yarrowia* (e.g., *Y. lipolytica*), *Fusarium* (e.g., *F. oxysporum*), *Bacillus* (e.g., *B. megaterium*, *B. subtilis*), *Chronobacter* (e.g., *C. sakazakii*), *Xanthophyllomyces* (e.g., *X. dendrorhous*), or *Pantoea* (e.g., *P. ananatis*) protein. An example of a *Candida viswanathii* nucleotide sequence (SEQ ID NO: 379) encoding a polypeptide (*BTS1*; SEQ ID NO: 345) having a geranylgeranyl diphosphate synthase activity is provided herein. Additional nonlimiting examples of nucleotide sequences

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1348 of 3077

encoding polypeptides having geranylgeranyl diphosphate synthase activity are provided herein which include: *Chronobacter sakazakii CrtE* gene (SEQ ID NO:357), which encodes *CsCrtE* (SEQ ID NO:326); *Xanthophyllomyces dendrorhous CrtE* gene (SEQ ID NO: 360, which encodes *XdCrtE* (SEQ ID NO:329); and *Pantoea ananatis CrtE* gene (SEQ ID NO: 363), which encodes *PaCrtE* (SEQ ID NO:332).

Presence, absence or amount of geranylgeranyl diphosphate synthase activity can be detected by any suitable method known in the art. For example, detection can be performed by using a geranylgeranyl diphosphate synthase ELISA kit (e.g., MYBIOSOURCE, MBS929545). Nucleic acid sequences encoding native and/or modified geranylgeranyl diphosphate synthase sequences also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cell or organism exhibits increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism.

The promoter used for regulating transcription of a nucleic acid (e.g., a heterologous nucleic acid) encoding a geranylgeranyl diphosphate synthase can be modified. For example, the amount of a geranylgeranyl diphosphate synthase protein expressed in a particular cellular location may be increased by including in the nucleic acid (e.g., heterologous nucleic acid) a strong heterologous promoter and/or a promoter that provides for a different pattern of expression in the cell or microorganism. An example of one such heterologous promoter is a *Candida* oleate-induced peroxisomal protein (*POX18*) gene promoter. Another example of one such heterologous promoter is a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene promoter. The nucleotide sequences of a *Candida viswanathii* HDE gene promoter and a *Candida* POX18 gene promoter are provided herein as are examples of additional fatty acid-inducible promoters. In some embodiments, the promoter is modified by replacing the endogenous promoter with a promoter comprising an alkane response element (ARE1). Non-limiting examples of assays suitable for assessing induction of geranylgeranyl diphosphate synthase expression by a carbon source or feedstock include RT-PCR or qRT-PCR after the host cell or microorganism has been exposed to the chosen carbon source or feedstock for varying amounts of time.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1349 of 3077

Decreasing geranylgeranyl diphosphate synthase activity in a cell can be accomplished by modifying the amount of geranylgeranyl diphosphate synthase protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous geranylgeranyl diphosphate synthase gene in a cell or organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type geranylgeranyl diphosphate synthase such that the encoded modified or substituted geranylgeranyl diphosphate synthase protein has a reduced enzyme activity.

Modification of a phytoene synthase activity

Another component of terpene biosynthesis is the conversion of geranylgeranyl diphosphate (geranylgeranyl pyrophosphate) to phytoene. A phytoene synthase enzyme (also referred to as CrtB, prephytoene-diphosphate synthase, 15-cis-phytoene synthase, PSase, geranylgeranyl-diphosphate geranylgeranyltransferase; EC 2.5.1.32) is a transferase enzyme that can catalyze the conversion of geranylgeranyl diphosphate to phytoene. In some embodiments of the cells, organisms, compositions and methods provided herein, the amount and/or activity of a phytoene synthase in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase the amount and/or activity of a phytoene synthase, may be modified to decrease the amount and/or activity of a phytoene synthase, or may be modified to alternately increase and decrease the amount and/or activity of a phytoene synthase depending, for example, on the cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

In certain aspects, the amount and/or activity of a phytoene synthase in a cell is increased. Increasing the amount and/or activity of a phytoene synthase may be particularly beneficial in embodiments in which the flux of carbons from fatty acids is directed toward a particular target product pathway (e.g., terpene production, carotenoid production, beta carotene production, lycopene production, astaxanthin production) and away from other cellular metabolic pathways not involved in target molecule production.

In certain embodiments, phytoene synthase activity is unchanged in a host or engineered cell or organism. In some embodiments, the amount and/or activity of a host phytoene synthase can be increased by increasing the number of copies of a nucleic acid encoding a phytoene synthase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the nucleic acid); by

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1350 of 3077

increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a phytoene synthase; or by increasing the number of copies of a nucleic acid encoding a phytoene synthase and increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a phytoene synthase. In some embodiments, a phytoene synthase is endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, the amount of phytoene synthase protein expressed in a cell can be increased by introducing heterologous nucleic acid encoding a phytoene synthase into a cell or microorganism. For example, introduction of heterologous nucleic acid encoding a phytoene synthase can result in increased copy number of such nucleic acids and/or provide for modification of the cellular location in which the protein is expressed. In some embodiments, a phytoene synthase is not endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, a phytoene synthase protein can be expressed in a cell by introducing heterologous nucleic acid encoding a phytoene synthase into a cell or microorganism.

In some embodiments, a phytoene synthase enzyme can be a fungal or bacterial protein. In a particular embodiment, the phytoene synthase enzyme can be a *Candida* (e.g., *C. tropicalis*, *C. viswanathii*, *C. maltosa*, *C. cloacae*), *Yarrowia* (e.g., *Y. lipolytica*), *Fusarium* (e.g., *F. oxysporum*), *Bacillus* (e.g., *B. megaterium*, *B. subtilis*), *Chronobacter* (e.g., *C. sakazakii*), *Xanthophyllomyces* (e.g., *X. dendrorhous*), or *Pantoea* (e.g., *P. ananatis*) protein. An example of a *Chronobacter sakazakii* nucleotide sequence (SEQ ID NO: 359) encoding a polypeptide (*CsCrtB*; SEQ ID NO: 328) having a phytoene synthase activity is provided herein. An example of a *Pantoea ananatis* nucleotide sequence (SEQ ID NO: 365) encoding a polypeptide (*PaCrtB*; SEQ ID NO: 334) having a phytoene synthase activity is provided herein.

Presence, absence or amount of phytoene synthase activity can be detected by any suitable method known in the art. For example, detection can be performed by using phytoene synthase activity assay (see e.g., Lopez-Emparan et al. (2014) PLoS One 9(12):e114878; Schledz et al. (1996) The Plant Journal 10(5):781-792). Nucleic acid sequences encoding native and/or modified phytoene synthase sequences also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cell or

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1351 of 3077

organism exhibits increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism.

The promoter used for regulating transcription of a nucleic acid (e.g., a heterologous nucleic acid) encoding a phytoene synthase can be modified. For example, the amount of a phytoene synthase protein expressed in a particular cellular location may be increased by including in the nucleic acid (e.g., heterologous nucleic acid) a strong heterologous promoter and/or a promoter that provides for a different pattern of expression in the cell or microorganism. An example of one such heterologous promoter is a *Candida* oleate-induced peroxisomal protein (*POX18*) gene promoter. Another example of one such heterologous promoter is a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene promoter. The nucleotide sequences of a *Candida viswanathii* HDE gene promoter and a *Candida* POX18 gene promoter are provided herein as are examples of additional fatty acid-inducible promoters. In some embodiments, the promoter is modified by replacing the endogenous promoter with a promoter comprising an alkane response element (ARE1). Non-limiting examples of assays suitable for assessing induction of phytoene synthase expression by a carbon source or feedstock include RT-PCR or qRT-PCR after the host cell or microorganism has been exposed to the chosen carbon source or feedstock for varying amounts of time.

Decreasing phytoene synthase activity in a cell can be accomplished by modifying the amount of phytoene synthase protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous phytoene synthase gene in a cell or organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type phytoene synthase such that the encoded modified or substituted phytoene synthase protein has a reduced enzyme activity.

Modification of a phytoene desaturase activity

Another component of terpene biosynthesis is the conversion of phytoene to lycopene. A phytoene desaturase enzyme can catalyze the conversion of phytoene to lycopene. In some embodiments, a phytoene desaturase is a 3,4-didehydrolycopene-forming phytoene desaturase (also referred to as 5-step phytoene desaturase, five-step phytoene desaturase, AI-1, 15-cis-phytoene:acceptor oxidoreductase (3,4-didehydrolycopene-forming); EC 1.3.99.30). In some embodiments, a phytoene desaturase is a neurosporene-forming phytoene desaturase (also

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1352 of 3077

referred to as CrtI, 3-step phytoene desaturase, three-step phytoene desaturase, 15-cis-phytoene:acceptor oxidoreductase (neurosporene-forming); EC 1.3.99.28). In some embodiments of the cells, organisms, compositions and methods provided herein, the amount and/or activity of a phytoene desaturase in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase the amount and/or activity of a phytoene desaturase, may be modified to decrease the amount and/or activity of a phytoene desaturase, or may be modified to alternately increase and decrease the amount and/or activity of a phytoene desaturase depending, for example, on the cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

In certain aspects, the amount and/or activity of a phytoene desaturase in a cell is increased. Increasing the amount and/or activity of a phytoene desaturase may be particularly beneficial in embodiments in which the flux of carbons from fatty acids is directed toward a particular target product pathway (e.g., terpene production, carotenoid production, beta carotene production, lycopene production, astaxanthin production) and away from other cellular metabolic pathways not involved in target molecule production.

In certain embodiments, phytoene desaturase activity is unchanged in a host or engineered cell or organism. In some embodiments, the amount and/or activity of a host phytoene desaturase can be increased by increasing the number of copies of a nucleic acid encoding a phytoene desaturase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the nucleic acid); by increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a phytoene desaturase; or by increasing the number of copies of a nucleic acid encoding a phytoene desaturase and increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a phytoene desaturase. In some embodiments, a phytoene desaturase is endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, the amount of phytoene desaturase protein expressed in a cell can be increased by introducing heterologous nucleic acid encoding a phytoene desaturase into a cell or microorganism. For example, introduction of heterologous nucleic acid encoding a phytoene desaturase can result in increased copy number of such nucleic acids and/or provide for modification of the cellular location in which the protein is expressed. In some embodiments, a phytoene desaturase is not endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, a phytoene desaturase protein can

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1353 of 3077

be expressed in a cell by introducing heterologous nucleic acid encoding a phytoene desaturase into a cell or microorganism.

In some embodiments, a phytoene desaturase enzyme can be a fungal or bacterial protein. In a particular embodiment, the phytoene desaturase enzyme can be a *Candida* (e.g., *C. tropicalis*, *C. viswanathii*, *C. maltosa*, *C. cloacae*), *Yarrowia* (e.g., *Y. lipolytica*), *Fusarium* (e.g., *F. oxysporum*), *Bacillus* (e.g., *B. megaterium*, *B. subtilis*), *Chronobacter* (e.g., *C. sakazakii*), *Xanthophyllomyces* (e.g., *X. dendrorhous*), or *Pantoea* (e.g., *P. ananatis*) protein. An example of a *Chronobacter sakazakii* nucleotide sequence (SEQ ID NO: 358) encoding a polypeptide (*CsCrtI*; SEQ ID NO: 327) having a phytoene desaturase activity is provided herein. An example of a *Xanthophyllomyces dendrorhous* nucleotide sequence (SEQ ID NO: 361) encoding a polypeptide (*XdCrtI*; SEQ ID NO: 330) having a phytoene desaturase activity is provided herein. An example of a *Pantoea ananatis* nucleotide sequence (SEQ ID NO: 364) encoding a polypeptide (*PaCrtI*; SEQ ID NO: 333) having a phytoene desaturase activity is provided herein.

Presence, absence or amount of phytoene desaturase activity can be detected by any suitable method known in the art. For example, detection can be performed by using a phytoene desaturase activity assay (see e.g., Xu et al. (2007) Microbiology 153:1642-52). Nucleic acid sequences encoding native and/or modified phytoene desaturase sequences also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cell or organism exhibits increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism.

The promoter used for regulating transcription of a nucleic acid (e.g., a heterologous nucleic acid) encoding a phytoene desaturase can be modified. For example, the amount of a phytoene desaturase protein expressed in a particular cellular location may be increased by including in the nucleic acid (e.g., heterologous nucleic acid) a strong heterologous promoter and/or a promoter that provides for a different pattern of expression in the cell or microorganism. An example of one such heterologous promoter is a *Candida* oleate-induced peroxisomal protein (*POX18*) gene promoter. Another example of one such heterologous promoter is a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene promoter. The nucleotide

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1354 of 3077

sequences of a *Candida viswanathii* HDE gene promoter and a *Candida* POX18 gene promoter are provided herein as are examples of additional fatty acid-inducible promoters. In some embodiments, the promoter is modified by replacing the endogenous promoter with a promoter comprising an alkane response element (ARE1). Non-limiting examples of assays suitable for assessing induction of phytoene desaturase expression by a carbon source or feedstock include RT-PCR or qRT-PCR after the host cell or microorganism has been exposed to the chosen carbon source or feedstock for varying amounts of time.

Decreasing phytoene desaturase activity in a cell can be accomplished by modifying the amount of phytoene desaturase protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous phytoene desaturase gene in a cell or organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type phytoene desaturase such that the encoded modified or substituted phytoene desaturase protein has a reduced enzyme activity.

Modification of a lycopene cyclase activity and/or modification of a bifunctional lycopene cyclase/phytoene synthase activity

Another component of terpene biosynthesis is the conversion of lycopene to beta carotene. A lycopene cyclase enzyme (also referred to as lycopene beta-cyclase, CrtY, CrtL (beta-ionone end group producing), CrtL (eta-ionone end group producing) and CrtL (capsanthin/capsorubin synthase); EC 5.5.1.19) can catalyze the conversion of lycopene to beta carotene. The cyclization of lycopene is typically the final step in carotenoid biosynthesis and may proceed via one of two pathways: the formation of a beta ring by beta-cyclase, or an epsilon ring by epsilon-cyclase. Epsilon-cyclase adds one ring, forming a monocyclic delta-carotene, and beta-cyclase introduces a ring at both ends of lycopene to form a bicyclic beta-carotene. In some embodiments, a lycopene cyclase is a bifunctional lycopene cyclase/phytoene synthase (also referred to as CrtYB), which includes a lycopene beta-cyclase domain (EC 5.5.1.19) and a phytoene synthase domain (EC 2.5.1.32). A bifunctional lycopene cyclase can catalyze the reaction from geranylgeranyl diphosphate to phytoene (phytoene synthase) and the reaction from lycopene to beta-carotene via an intermediate gamma-carotene (lycopene cyclase). A bifunctional enzyme containing lycopene cyclase activity and phytoene synthase activity may be referred to herein as a bifunctional lycopene cyclase/phytoene synthase or a bifunctional lycopene cyclase.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1355 of 3077

In some embodiments of the cells, organisms, compositions and methods provided herein, the amount and/or activity of a lycopene cyclase in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase the amount and/or activity of a lycopene cyclase, may be modified to decrease the amount and/or activity of a lycopene cyclase, or may be modified to alternately increase and decrease the amount and/or activity of a lycopene cyclase depending, for example, on the cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

In some embodiments of the cells, organisms, compositions and methods provided herein, the amount and/or activity of a bifunctional lycopene cyclase/phytoene synthase in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase the amount and/or activity of a bifunctional lycopene cyclase/phytoene synthase, may be modified to decrease the amount and/or activity of a bifunctional lycopene cyclase/phytoene synthase, or may be modified to alternately increase and decrease the amount and/or activity of a bifunctional lycopene cyclase/phytoene synthase depending, for example, on the cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

In certain aspects, the amount and/or activity of a lycopene cyclase in a cell is increased. Increasing the amount and/or activity of a lycopene cyclase may be particularly beneficial in embodiments in which the flux of carbons from fatty acids is directed toward a particular target product pathway (e.g., terpene production, carotenoid production, beta carotene production, astaxanthin production) and away from other cellular metabolic pathways not involved in target molecule production.

In certain aspects, the amount and/or activity of a bifunctional lycopene cyclase/phytoene synthase in a cell is increased. Increasing the amount and/or activity of a bifunctional lycopene cyclase/phytoene synthase may be particularly beneficial in embodiments in which the flux of carbons from fatty acids is directed toward a particular target product pathway (e.g., terpene production, carotenoid production, beta carotene production, astaxanthin production) and away from other cellular metabolic pathways not involved in target molecule production.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1356 of 3077

In certain embodiments, lycopene cyclase activity is unchanged in a host or engineered cell or organism. In some embodiments, the amount and/or activity of a host lycopene cyclase can be increased by increasing the number of copies of a nucleic acid encoding a lycopene cyclase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the nucleic acid); by increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a lycopene cyclase; or by increasing the number of copies of a nucleic acid encoding a lycopene cyclase and increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a lycopene cyclase. In some embodiments, a lycopene cyclase is endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, the amount of lycopene cyclase protein expressed in a cell can be increased by introducing heterologous nucleic acid encoding a lycopene cyclase into a cell or microorganism. For example, introduction of heterologous nucleic acid encoding a lycopene cyclase can result in increased copy number of such nucleic acids and/or provide for modification of the cellular location in which the protein is expressed. In some embodiments, a lycopene cyclase is not endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, a lycopene cyclase protein can be expressed in a cell by introducing heterologous nucleic acid encoding a lycopene cyclase into a cell or microorganism.

In certain embodiments, bifunctional lycopene cyclase/phytoene synthase activity is unchanged in a host or engineered cell or organism. In some embodiments, the amount and/or activity of a host bifunctional lycopene cyclase/phytoene synthase can be increased by increasing the number of copies of a nucleic acid encoding a bifunctional lycopene cyclase/phytoene synthase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the nucleic acid); by increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a bifunctional lycopene cyclase/phytoene synthase; or by increasing the number of copies of a nucleic acid encoding a bifunctional lycopene cyclase/phytoene synthase and increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a bifunctional lycopene cyclase/phytoene synthase. In some embodiments, a bifunctional lycopene cyclase/phytoene synthase is endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, the amount of bifunctional lycopene cyclase/phytoene synthase protein expressed in a cell can be increased by introducing heterologous nucleic acid encoding a bifunctional lycopene cyclase/phytoene synthase into a cell or microorganism. For example, introduction of heterologous nucleic acid encoding a bifunctional lycopene cyclase/phytoene synthase can result in increased copy number of such nucleic acids and/or

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1357 of 3077

provide for modification of the cellular location in which the protein is expressed. In some embodiments, a bifunctional lycopene cyclase/phytoene synthase is not endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, a bifunctional lycopene cyclase/phytoene synthase protein can be expressed in a cell by introducing heterologous nucleic acid encoding a bifunctional lycopene cyclase/phytoene synthase into a cell or microorganism.

In some embodiments, a lycopene cyclase enzyme and/or a bifunctional lycopene cyclase/phytoene synthase enzyme can be a fungal or bacterial protein. In a particular embodiment, the lycopene cyclase enzyme and/or the bifunctional lycopene cyclase/phytoene synthase enzyme can be a *Candida* (e.g., *C. tropicalis*, *C. viswanathii*, *C. maltosa*, *C. cloacae*), *Yarrowia* (e.g., *Y. lipolytica*), *Fusarium* (e.g., *F. oxysporum*), *Bacillus* (e.g., *B. megaterium*, *B. subtilis*), *Chronobacter* (e.g., *C. sakazakii*), *Xanthophyllomyces* (e.g., *X. dendrorhous*), or *Pantoea* (e.g., *P. ananatis*) protein. An example of a *Xanthophyllomyces dendrorhous* nucleotide sequence (SEQ ID NO: 362) encoding a polypeptide (*XdCrtYB*; SEQ ID NO: 331) having a bifunctional lycopene cyclase/phytoene synthase activity is provided herein.

Presence, absence or amount of lycopene cyclase activity and/or bifunctional lycopene cyclase/phytoene synthase activity can be detected by any suitable method known in the art. For example, detection can be performed by using a lycopene cyclase activity assay (see e.g., Yu et al. (2010) J. Biol. Chem. 285(16):12109-12120). Nucleic acid sequences encoding native and/or modified lycopene cyclase sequences and/or bifunctional lycopene cyclase/phytoene synthase sequences also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cell or organism exhibits increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism.

The promoter used for regulating transcription of a nucleic acid (e.g., a heterologous nucleic acid) encoding a lycopene cyclase and/or a bifunctional lycopene cyclase/phytoene synthase can be modified. For example, the amount of a lycopene cyclase protein and/or a bifunctional lycopene cyclase/phytoene synthase protein expressed in a particular cellular location may be increased by including in the nucleic acid (e.g., heterologous nucleic acid) a strong heterologous promoter and/or a promoter that provides for a different pattern of expression in the cell or

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1358 of 3077

microorganism. An example of one such heterologous promoter is a *Candida* oleate-induced peroxisomal protein (*POX18*) gene promoter. Another example of one such heterologous promoter is a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene promoter. The nucleotide sequences of a *Candida viswanathii* HDE gene promoter and a *Candida* POX18 gene promoter are provided herein as are examples of additional fatty acid-inducible promoters. In some embodiments, the promoter is modified by replacing the endogenous promoter with a promoter comprising an alkane response element (ARE1). Non-limiting examples of assays suitable for assessing induction of lycopene cyclase expression and/or a bifunctional lycopene cyclase/phytoene synthase expression by a carbon source or feedstock include RT-PCR or qRT-PCR after the host cell or microorganism has been exposed to the chosen carbon source or feedstock for varying amounts of time.

Decreasing lycopene cyclase activity in a cell can be accomplished by modifying the amount of lycopene cyclase protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous lycopene cyclase gene in a cell or organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type lycopene cyclase such that the encoded modified or substituted lycopene cyclase protein has a reduced enzyme activity.

Decreasing bifunctional lycopene cyclase/phytoene synthase activity in a cell can be accomplished by modifying the amount of bifunctional lycopene cyclase/phytoene synthase protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous bifunctional lycopene cyclase/phytoene synthase gene in a cell or organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type bifunctional lycopene cyclase/phytoene synthase such that the encoded modified or substituted bifunctional lycopene cyclase/phytoene synthase protein has a reduced enzyme activity.

Modification of a beta carotene ketolase activity

Additional components of terpene biosynthesis include, for example, astaxanthin biosysnthesis (e.g., the conversion of beta carotene to astaxanthin), canthaxanthin biosynthesis, and zeaxanthin biosynthesis. Canthaxanthin, for example, may be produced using a beta carotene ketolase enzyme (also referred to as CrtW, beta carotene monoketolase, beta-carotene

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1359 of 3077

oxygenase; EC 1.3.5.B4). Astaxanthin, for example, may be produced using a beta carotene ketolase enzyme in conjunction with a beta carotene hydroxylase (discussed below). In some embodiments of the cells, organisms, compositions and methods provided herein, the amount and/or activity of a beta carotene ketolase in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase the amount and/or activity of a beta carotene ketolase, may be modified to decrease the amount and/or activity of a beta carotene ketolase, or may be modified to alternately increase and decrease the amount and/or activity of a beta carotene ketolase depending, for example, on the cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

In certain aspects, the amount and/or activity of a beta carotene ketolase in a cell is increased. Increasing the amount and/or activity of a beta carotene ketolase may be particularly beneficial in embodiments in which the flux of carbons from fatty acids is directed toward a particular target product pathway (e.g., terpene production, carotenoid production, astaxanthin production) and away from other cellular metabolic pathways not involved in target molecule production.

In certain embodiments beta carotene ketolase activity is unchanged in a host or engineered cell or organism. In some embodiments, the amount and/or activity of a host beta carotene ketolase can be increased by increasing the number of copies of a nucleic acid encoding an beta carotene ketolase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the nucleic acid); by increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a beta carotene ketolase; or by increasing the number of copies of a nucleic acid encoding a beta carotene ketolase and increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a beta carotene ketolase. In some embodiments, a beta carotene ketolase is endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, the amount of beta carotene ketolase protein expressed in a cell can be increased by introducing heterologous nucleic acid encoding a beta carotene ketolase into a cell or microorganism. For example, introduction of heterologous nucleic acid encoding a beta carotene ketolase can result in increased copy number of such nucleic acids and/or provide for modification of the cellular location in which the protein is expressed. In some embodiments, a beta carotene ketolase is not endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, a beta carotene ketolase protein can be expressed in a cell by introducing heterologous nucleic acid encoding a beta carotene ketolase into a cell or microorganism.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1360 of 3077

In some embodiments, a beta carotene ketolase enzyme can be a fungal or bacterial protein. In a particular embodiment, the beta carotene ketolase enzyme can be a *Candida* (e.g., *C. tropicalis*, *C. viswanathii*, *C. maltosa*, *C. cloacae*), *Yarrowia* (e.g., *Y. lipolytica*), *Fusarium* (e.g., *F. oxysporum*), *Bacillus* (e.g., *B. megaterium*, *B. subtilis*), *Chronobacter* (e.g., *C. sakazakii*), *Xanthophyllomyces* (e.g., *X. dendrorhous*), *Pantoea* (e.g., *P. ananatis*), or *Agrobacterium* (e.g., *A. aurantiacum*) protein. An example of an *Agrobacterium aurantiacum* nucleotide sequence (SEQ ID NO: 369) encoding a polypeptide (*AaCrtW*; SEQ ID NO: 351) having a beta carotene ketolase activity is provided herein.

Presence, absence or amount of beta carotene ketolase activity can be detected by any suitable method known in the art. For example, detection can be performed by using an appropriate beta carotene ketolase activity assay. Nucleic acid sequences encoding native and/or modified beta carotene ketolase sequences also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cell or organism exhibits increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism.

The promoter used for regulating transcription of a nucleic acid (e.g., a heterologous nucleic acid) encoding a beta carotene ketolase can be modified. For example, the amount of a beta carotene ketolase protein expressed in a particular cellular location may be increased by including in the nucleic acid (e.g., heterologous nucleic acid) a strong heterologous promoter and/or a promoter that provides for a different pattern of expression in the cell or microorganism. An example of one such heterologous promoter is a *Candida* oleate-induced peroxisomal protein (*POX18*) gene promoter. Another example of one such heterologous promoter is a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene promoter. The nucleotide sequences of a *Candida viswanathii* HDE gene promoter and a *Candida* POX18 gene promoter are provided herein as are examples of additional fatty acid-inducible promoters. In some embodiments, the promoter is modified by replacing the endogenous promoter with a promoter comprising an alkane response element (ARE1). Non-limiting examples of assays suitable for assessing induction of beta carotene ketolase expression by a carbon source or feedstock

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1361 of 3077

include RT-PCR or qRT-PCR after the host cell or microorganism has been exposed to the chosen carbon source or feedstock for varying amounts of time.

Decreasing beta carotene ketolase activity in a cell can be accomplished by modifying the amount of beta carotene ketolase protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous beta carotene ketolase gene in a cell or organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type beta carotene ketolase such that the encoded modified or substituted beta carotene ketolase protein has a reduced enzyme activity.

Modification of a beta carotene hydroxylase activity

In certain instances, astaxanthin and/or zeaxanthin may be produced using a beta carotene hydroxylase enzyme (also referred to as CrtZ, beta-carotene 3-hydroxylase, beta-carotene 3,3'-monooxygenase, beta-carotene, NADH:oxygen 3-oxidoreductase; EC 1.14.13.129). In some embodiments of the cells, organisms, compositions and methods provided herein, the amount and/or activity of a beta carotene hydroxylase in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase the amount and/or activity of a beta carotene hydroxylase, may be modified to decrease the amount and/or activity of a beta carotene hydroxylase, or may be modified to alternately increase and decrease the amount and/or activity of a beta carotene hydroxylase depending, for example, on the cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

In certain aspects, the amount and/or activity of a beta carotene hydroxylase in a cell is increased. Increasing the amount and/or activity of a beta carotene hydroxylase may be particularly beneficial in embodiments in which the flux of carbons from fatty acids is directed toward a particular target product pathway (e.g., terpene production, carotenoid production, astaxanthin production) and away from other cellular metabolic pathways not involved in target molecule production.

In certain embodiments, beta carotene hydroxylase activity is unchanged in a host or engineered cell or organism. In some embodiments, the amount and/or activity of a host beta carotene hydroxylase can be increased by increasing the number of copies of a nucleic acid encoding a beta carotene hydroxylase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1362 of 3077

copies of the nucleic acid); by increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a beta carotene hydroxylase; or by increasing the number of copies of a nucleic acid encoding a beta carotene hydroxylase and increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a beta carotene hydroxylase. In some embodiments, a beta carotene hydroxylase is endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, the amount of beta carotene hydroxylase protein expressed in a cell can be increased by introducing heterologous nucleic acid encoding a beta carotene hydroxylase into a cell or microorganism. For example, introduction of heterologous nucleic acid encoding a beta carotene hydroxylase can result in increased copy number of such nucleic acids and/or provide for modification of the cellular location in which the protein is expressed. In some embodiments, a beta carotene hydroxylase is not endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, a beta carotene hydroxylase protein can be expressed in a cell by introducing heterologous nucleic acid encoding a beta carotene hydroxylase into a cell or microorganism.

In some embodiments, a beta carotene hydroxylase enzyme can be a fungal or bacterial protein. In a particular embodiment, the beta carotene hydroxylase enzyme can be a *Candida* (e.g., *C. tropicalis*, *C. viswanathii*, *C. maltosa*, *C. cloacae*), *Yarrowia* (e.g., *Y. lipolytica*), *Fusarium* (e.g., *F. oxysporum*), *Bacillus* (e.g., *B. megaterium*, *B. subtilis*), *Chronobacter* (e.g., *C. sakazakii*), *Xanthophyllomyces* (e.g., *X. dendrorhous*), *Pantoea* (e.g., *P. ananatis*) or *Agrobacterium* (e.g., *A. aurantiacum*) protein. An example of an *Agrobacterium aurantiacum* nucleotide sequence (SEQ ID NO: 368) encoding a polypeptide (*AaCrtZ*; SEQ ID NO: 350) having a beta carotene hydroxylase activity is provided herein.

Presence, absence or amount of beta carotene hydroxylase activity can be detected by any suitable method known in the art. For example, detection can be performed by using an appropriate beta carotene hydroxylase activity assay. Nucleic acid sequences encoding native and/or modified beta carotene hydroxylase sequences also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cell or organism exhibits increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1363 of 3077

The promoter used for regulating transcription of a nucleic acid (e.g., a heterologous nucleic acid) encoding a beta carotene hydroxylase can be modified. For example, the amount of a beta carotene hydroxylase protein expressed in a particular cellular location may be increased by including in the nucleic acid (e.g., heterologous nucleic acid) a strong heterologous promoter and/or a promoter that provides for a different pattern of expression in the cell or microorganism. An example of one such heterologous promoter is a *Candida* oleate-induced peroxisomal protein (*POX18*) gene promoter. Another example of one such heterologous promoter is a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene promoter. The nucleotide sequences of a *Candida viswanathii* HDE gene promoter and a *Candida* POX18 gene promoter are provided herein as are examples of additional fatty acid-inducible promoters. In some embodiments, the promoter is modified by replacing the endogenous promoter with a promoter comprising an alkane response element (ARE1). Non-limiting examples of assays suitable for assessing induction of beta carotene hydroxylase expression by a carbon source or feedstock include RT-PCR or qRT-PCR after the host cell or microorganism has been exposed to the chosen carbon source or feedstock for varying amounts of time.

Decreasing beta carotene hydroxylase activity in a cell can be accomplished by modifying the amount of beta carotene hydroxylase protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous beta carotene hydroxylase gene in a cell or organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type beta carotene hydroxylase such that the encoded modified or substituted beta carotene hydroxylase protein has a reduced enzyme activity.

Modification of an astaxanthin synthase activity

In certain instances, astaxanthin may be produced using an astaxanthin synthase enzyme (also referred to as CrtS, astaxanthin synthetase, cytochrome-450 hydroxylase, cytochrome-P450 hydroxylase/astaxanthin synthase). In some embodiments of the cells, organisms, compositions and methods provided herein, the amount and/or activity of an astaxanthin synthase in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase the amount and/or activity of an astaxanthin synthase, may be modified to decrease the amount and/or activity of an astaxanthin synthase, or may be modified to

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1364 of 3077

alternately increase and decrease the amount and/or activity of an astaxanthin synthase depending, for example, on the cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

In certain aspects, the amount and/or activity of an astaxanthin synthase in a cell is increased. Increasing the amount and/or activity of an astaxanthin synthase may be particularly beneficial in embodiments in which the flux of carbons from fatty acids is directed toward a particular target product pathway (e.g., terpene production, carotenoid production, astaxanthin production) and away from other cellular metabolic pathways not involved in target molecule production.

In certain embodiments, astaxanthin synthase activity is unchanged in a host or engineered cell or organism. In some embodiments, the amount and/or activity of a host astaxanthin synthase can be increased by increasing the number of copies of a nucleic acid encoding an astaxanthin synthase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the nucleic acid); by increasing the activity of a promoter that regulates transcription of a nucleic acid encoding an astaxanthin synthase; or by increasing the number of copies of a nucleic acid encoding an astaxanthin synthase and increasing the activity of a promoter that regulates transcription of a nucleic acid encoding an astaxanthin synthase. In some embodiments, an astaxanthin synthase is endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, the amount of astaxanthin synthase protein expressed in a cell can be increased by introducing heterologous nucleic acid encoding an astaxanthin synthase into a cell or microorganism. For example, introduction of heterologous nucleic acid encoding an astaxanthin synthase can result in increased copy number of such nucleic acids and/or provide for modification of the cellular location in which the protein is expressed. In some embodiments, a astaxanthin synthase is not endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, a astaxanthin synthase protein can be expressed in a cell by introducing heterologous nucleic acid encoding a astaxanthin synthase into a cell or microorganism.

In some embodiments, an astaxanthin synthase enzyme can be a fungal or bacterial protein. In a particular embodiment, the astaxanthin synthase enzyme can be a *Candida* (e.g., *C. tropicalis*, *C. viswanathii*, *C. maltosa*, *C. cloacae*), *Yarrowia* (e.g., *Y. lipolytica*), *Fusarium* (e.g., *F. oxysporum*), *Bacillus* (e.g., *B. megaterium*, *B. subtilis*), *Chronobacter* (e.g., *C. sakazakii*), *Xanthophyllomyces* (e.g., *X. dendrorhous*), or *Pantoea* (e.g., *P. ananatis*) protein. An example

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1365 of 3077

of a *Xanthophyllomyces dendrorhous* nucleotide sequence (SEQ ID NO: 367) encoding a polypeptide (*XdCrtS*; SEQ ID NO: 336) having an astaxanthin synthase activity is provided herein.

Presence, absence or amount astaxanthin synthase activity can be detected by any suitable method known in the art. For example, detection can be performed by using an appropriate astaxanthin synthase activity assay. Nucleic acid sequences encoding native and/or modified astaxanthin synthase sequences also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cell or organism exhibits increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism.

The promoter used for regulating transcription of a nucleic acid (e.g., a heterologous nucleic acid) encoding an astaxanthin synthase can be modified. For example, the amount of an astaxanthin synthase protein expressed in a particular cellular location may be increased by including in the nucleic acid (e.g., heterologous nucleic acid) a strong heterologous promoter and/or a promoter that provides for a different pattern of expression in the cell or microorganism. An example of one such heterologous promoter is a *Candida* oleate-induced peroxisomal protein (*POX18*) gene promoter. Another example of one such heterologous promoter is a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene promoter. The nucleotide sequences of a *Candida viswanathii* HDE gene promoter and a *Candida* POX18 gene promoter are provided herein as are examples of additional fatty acid-inducible promoters. In some embodiments, the promoter is modified by replacing the endogenous promoter with a promoter comprising an alkane response element (ARE1). Non-limiting examples of assays suitable for assessing induction of astaxanthin synthase expression by a carbon source or feedstock include RT-PCR or qRT-PCR after the host cell or microorganism has been exposed to the chosen carbon source or feedstock for varying amounts of time.

Decreasing astaxanthin synthase activity in a cell can be accomplished by modifying the amount of astaxanthin synthase protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous astaxanthin synthase gene in a cell or organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1366 of 3077

modifying a gene encoding a wild-type astaxanthin synthase such that the encoded modified or substituted astaxanthin synthase protein has a reduced enzyme activity.

Modification of a zeaxanthin glucosyltransferase activity

Another component of terpene biosynthesis is the conversion of zeaxanthin to zeaxanthin diglucoside. A zeaxanthin glucosyltransferase enzyme (also referred to as CrtX, UDP-glucose:zeaxanthin beta-D-glucosyltransferase; EC 2.4.1.276) can catalyze the conversion of zeaxanthin to zeaxanthin diglucoside. In some embodiments of the cells, organisms, compositions and methods provided herein, the amount and/or activity of a zeaxanthin glucosyltransferase in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase the amount and/or activity of a zeaxanthin glucosyltransferase, may be modified to decrease the amount and/or activity of a zeaxanthin glucosyltransferase, or may be modified to alternately increase and decrease the amount and/or activity of a zeaxanthin glucosyltransferase depending, for example, on the cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

In certain aspects, the amount and/or activity of a zeaxanthin glucosyltransferase in a cell is increased. Increasing the amount and/or activity of a zeaxanthin glucosyltransferase may be particularly beneficial in embodiments in which the flux of carbons from fatty acids is directed toward a particular target product pathway (e.g., terpene production, carotenoid production, xanthophyll production, zeaxanthin diglucoside production), and away from other cellular metabolic pathways not involved in target molecule production.

In certain embodiments, zeaxanthin glucosyltransferase activity is unchanged in a host or engineered cell or organism. In some embodiments, the amount and/or activity of a host zeaxanthin glucosyltransferase can be increased by increasing the number of copies of a nucleic acid encoding a zeaxanthin glucosyltransferase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the nucleic acid); by increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a zeaxanthin glucosyltransferase; or by increasing the number of copies of a nucleic acid encoding a zeaxanthin glucosyltransferase and increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a zeaxanthin glucosyltransferase. In some embodiments, a zeaxanthin glucosyltransferase is endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1367 of 3077

provided herein, the amount of zeaxanthin glucosyltransferase protein expressed in a cell can be increased by introducing heterologous nucleic acid encoding a zeaxanthin glucosyltransferase into a cell or microorganism. For example, introduction of heterologous nucleic acid encoding a zeaxanthin glucosyltransferase can result in increased copy number of such nucleic acids and/or provide for modification of the cellular location in which the protein is expressed. In some embodiments, a zeaxanthin glucosyltransferase is not endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, a zeaxanthin glucosyltransferase protein can be expressed in a cell by introducing heterologous nucleic acid encoding a zeaxanthin glucosyltransferase into a cell or microorganism.

In some embodiments, a zeaxanthin glucosyltransferase enzyme can be a fungal or bacterial protein. In a particular embodiment, the zeaxanthin glucosyltransferase enzyme can be a *Candida* (e.g., *C. tropicalis*, *C. viswanathii*, *C. maltosa*, *C. cloacae*), *Yarrowia* (e.g., *Y. lipolytica*), *Fusarium* (e.g., *F. oxysporum*), *Bacillus* (e.g., *B. megaterium*, *B. subtilis*), *Chronobacter* (e.g., *C. sakazakii*), *Xanthophyllomyces* (e.g., *X. dendrorhous*), or *Pantoea* (e.g., *P. ananatis*) protein.

Presence, absence or amount of zeaxanthin glucosyltransferase activity can be detected by any suitable method known in the art. For example, detection can be performed by using an appropriate zeaxanthin glucosyltransferase activity assay. Nucleic acid sequences encoding native and/or modified zeaxanthin glucosyltransferase sequences also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cell or organism exhibits increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism.

The promoter used for regulating transcription of a nucleic acid (e.g., a heterologous nucleic acid) encoding a zeaxanthin glucosyltransferase can be modified. For example, the amount of a zeaxanthin glucosyltransferase protein expressed in a particular cellular location may be increased by including in the nucleic acid (e.g., heterologous nucleic acid) a strong heterologous promoter and/or a promoter that provides for a different pattern of expression in the cell or microorganism. An example of one such heterologous promoter is a *Candida* oleate-induced peroxisomal protein (*POX18*) gene promoter. Another example of one such heterologous

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1368 of 3077

promoter is a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene promoter. The nucleotide sequences of a *Candida viswanathii* HDE gene promoter and a *Candida* POX18 gene promoter are provided herein as are examples of additional fatty acid-inducible promoters. In some embodiments, the promoter is modified by replacing the endogenous promoter with a promoter comprising an alkane response element (ARE1). Non-limiting examples of assays suitable for assessing induction of zeaxanthin glucosyltransferase expression by a carbon source or feedstock include RT-PCR or qRT-PCR after the host cell or microorganism has been exposed to the chosen carbon source or feedstock for varying amounts of time.

Decreasing zeaxanthin glucosyltransferase activity in a cell can be accomplished by modifying the amount of zeaxanthin glucosyltransferase protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous zeaxanthin glucosyltransferase gene in a cell or organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type zeaxanthin glucosyltransferase such that the encoded modified or substituted zeaxanthin glucosyltransferase protein has a reduced enzyme activity.

Modification of a valencene synthase activity

Another component of terpene biosynthesis is the synthesis of valencene. A valencene synthase enzyme (also referred to as TPS1, (2E,6E)-farnesyl-diphosphate diphosphate-lyase (valencene-forming); EC 4.2.3.73) can catalyze the conversion of farnesyl diphosphate (farnesyl pyrophosphate) to valencene. In some embodiments of the cells, organisms, compositions and methods provided herein, the amount and/or activity of a valencene synthase in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase the amount and/or activity of a valencene synthase, may be modified to decrease the amount and/or activity of a valencene synthase, or may be modified to alternately increase and decrease the amount and/or activity of a valencene synthase depending, for example, on the cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

In certain aspects, the amount and/or activity of a valencene synthase in a cell is increased. Increasing the amount and/or activity of a valencene synthase may be particularly beneficial in embodiments in which the flux of carbons from fatty acids is directed toward a particular target

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1369 of 3077

product pathway (e.g., terpene production, valencene production, nootkatone production) and away from other cellular metabolic pathways not involved in target molecule production.

In certain embodiments, valencene synthase activity is unchanged in a host or engineered cell or organism. In some embodiments, the amount and/or activity of a host valencene synthase can be increased by increasing the number of copies of a nucleic acid encoding a valencene synthase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the nucleic acid); by increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a valencene synthase; or by increasing the number of copies of a nucleic acid encoding a valencene synthase and increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a valencene synthase. In some embodiments, a valencene synthase is endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, the amount valencene synthase protein expressed in a cell can be increased by introducing heterologous nucleic acid encoding a valencene synthase into a cell or microorganism. For example, introduction of heterologous nucleic acid encoding a valencene synthase can result in increased copy number of such nucleic acids and/or provide for modification of the cellular location in which the protein is expressed. In some embodiments, a valencene synthase is not endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, a valencene synthase protein can be expressed in a cell by introducing heterologous nucleic acid encoding a valencene synthase into a cell or microorganism.

In some embodiments, a valencene synthase enzyme can be a fungal or bacterial protein. In a particular embodiment, the valencene synthase enzyme can be a *Candida* (e.g., *C. tropicalis*, *C. viswanathii*, *C. maltosa*, *C. cloacae*), *Yarrowia* (e.g., *Y. lipolytica*), *Fusarium* (e.g., *F. oxysporum*), *Bacillus* (e.g., *B. megaterium*, *B. subtilis*), *Chronobacter* (e.g., *C. sakazakii*), *Xanthophyllomyces* (e.g., *X. dendrorhous*), *Pantoea* (e.g., *P. ananatis*), or *Callitropsis* (e.g., *C. nootkatensis*) protein. An example of a *Callitropsis nootkatensis* nucleotide sequence (SEQ ID NO: 370) encoding a polypeptide (*TPS1*; SEQ ID NO: 352) having a valencene synthase activity is provided herein.

Presence, absence or amount of valencene synthase activity can be detected by any suitable method known in the art. For example, detection can be performed by using an appropriate valencene synthase activity assay. Nucleic acid sequences encoding native and/or modified

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1370 of 3077

valencene synthase sequences also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cell or organism exhibits increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism.

The promoter used for regulating transcription of a nucleic acid (e.g., a heterologous nucleic acid) encoding a valencene synthase can be modified. For example, the amount of a valencene synthase protein expressed in a particular cellular location may be increased by including in the nucleic acid (e.g., heterologous nucleic acid) a strong heterologous promoter and/or a promoter that provides for a different pattern of expression in the cell or microorganism. An example of one such heterologous promoter is a *Candida* oleate-induced peroxisomal protein (*POX18*) gene promoter. Another example of one such heterologous promoter is a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene promoter. The nucleotide sequences of a *Candida viswanathii* HDE gene promoter and a *Candida* POX18 gene promoter are provided herein as are examples of additional fatty acid-inducible promoters. In some embodiments, the promoter is modified by replacing the endogenous promoter with a promoter comprising an alkane response element (ARE1). Non-limiting examples of assays suitable for assessing induction of valencene synthase expression by a carbon source or feedstock include RT-PCR or qRT-PCR after the host cell or microorganism has been exposed to the chosen carbon source or feedstock for varying amounts of time.

Decreasing valencene synthase activity in a cell can be accomplished by modifying the amount of valencene synthase protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous valencene synthase gene in a cell or organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type valencene synthase such that the encoded modified or substituted valencene synthase protein has a reduced enzyme activity.

Modification of RAS2 activity

RAS2 is a guanine nucleotide-binding protein that can be activated by binding GTP (e.g., in the presence of glucose). In some embodiments of the microorganisms, compositions and methods

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1371 of 3077

provided herein, the amount and/or activity of RAS2 in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase RAS2 and/or RAS2 activity, may be modified to decrease RAS2 and/or RAS2 activity, or may be modified to alternately increase and decrease RAS2 and/or RAS2 activity depending, for example, on the cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

In certain aspects, the amount and/or activity of a RAS2 in a cell is decreased. Reducing or eliminating the amount and/or activity of a RAS2 may be particularly beneficial in embodiments in which the flux of carbons from fatty acids is directed toward a particular target product pathway and away from other cellular metabolic pathways involving activated fatty acids. In certain embodiments, an endogenous microbial gene encoding RAS2 (e.g., yeast *RAS2* gene) can be disrupted or deleted in a host microorganism to reduce or eliminate RAS2 activity in the host relative to a microorganism in which the gene has not been modified. Methods for decreasing the amount and/or activity of RAS2 in a cell include, but are not limited to, modifying the amount of RAS2 protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous RAS2 gene in a cell or organism with a weaker heterologous promoter, deleting or disrupting one or both copies of an endogenous gene, and/or replacing or modifying a gene encoding a wild-type RAS2 such that the encoded modified or substituted RAS2 protein has a reduced enzyme activity. For example, expression of a host RAS2 activity can be decreased or eliminated by disruption (e.g., knockout, insertion mutagenesis, the like and combinations thereof) of a host gene encoding the protein, or by decreasing the activity of the promoter (e.g., through addition of repressor sequences to the promoter or 5'UTR or replacing the promoter) that controls transcription of a RAS2 gene using recombinant molecular biology techniques known in the art and/or described herein. In one embodiment, a diploid yeast, such as, for example, a *Candida* yeast, when used as a host microorganism can be subjected to genetic modification in which one of the two alleles of a RAS2 gene is disrupted or deleted. In so doing, a single allele of the gene remains for a reduced amount of RAS2 expression in the microorganism and a reduced amount of the protein in the cell.

One method for disrupting an endogenous RAS2 gene is by recombinantly inserting a heterologous nucleic acid (e.g., a nucleotide sequence encoding a selectable marker such as an enzyme that restores an auxotrophic host organism to prototrophy) into the endogenous gene, thereby generating an engineered organism deficient in RAS2 activity. This can be done, for example, through homologous recombination in which a heterologous nucleic acid

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1372 of 3077

containing sequences of an endogenous RAS2 gene and a disrupting sequence (e.g., a knock-out gene cassette such as described herein) is introduced into a host cell or microorganism. Nucleic acids encoding a RAS2 can be obtained from a number of sources, including, for example, yeast cells. Genomic DNA from cell sources can be amplified using oligonucleotide primers based on the nucleotide sequence of a RAS2 encoding gene, including examples provided herein.

Presence, absence or amount of RAS2 activity can be detected by any suitable method known in the art. For example, detection can be performed by using an appropriate RAS2 activity assay (e.g., enzymatic assays, PCR based assays (e.g., qPCR, RT-PCR), immunological detection methods (e.g., antibodies specific for RAS2), the like and combinations thereof). Nucleic acid sequences encoding native and/or modified RAS2 sequences also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cell or organism exhibits increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism.

Carbon-containing products

Certain organic acids and polyketides are chemical intermediates in manufacturing processes used to make polyamides, polyurethanes and plasticizers, all of which have wide applications in producing items such as antiseptics, carpets, elastomers, food packaging, lubricants, top-grade coatings, hot-melt coating and adhesives, painting materials, corrosion inhibitor, surfactant, engineering plastics and can also be used as a starting material in the manufacture of fragrances. Some large-scale synthetic processes for making organic acids and polyketides include the use of noxious chemicals and/or solvents, some require high temperatures, and all require significant energy input. In addition, some of the processes emit toxic byproducts (such as nitrous oxide) and give rise to environmental concerns. Furthermore, chemical synthesis and extraction of desirable chemical compounds, such as terpenes, for example, from natural sources yields low product levels and is often not economically feasible. Provided herein are methods for producing organic acid and other organic chemical intermediate target molecules using biological systems provided herein. Such production systems may have significantly less

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1373 of 3077

environmental impact and could be economically competitive with synthetic manufacturing systems.

Organic Acids

Examples of organic acid target molecules that can be produced using compositions and methods provided herein include, but are not limited to, fatty acids, diacids and β-hydroxy acids (e.g., hydroxyalkanoate monomers) and salts and esters thereof. Fatty acids generally tend to be aliphatic acids of varying carbon chain lengths. Naturally occurring fatty acids in biological systems generally contain an even number of carbon atoms, typically between about 12 to about 24, or about 14 to about 24, and most commonly, 16 or 18 carbon atoms. Based on the number of carbons in a fatty acid carbon chain, it can be categorized as a short-, medium- or long-chain fatty acid. Generally, short-chain fatty acids have a chain length of about 2 to about 6 carbon atoms, medium-chain fatty acids have a chain length of about 8 to about 10 carbon atoms, long-chain fatty acids have a chain length of about 12 to about 20 carbon atoms and very long-chain length fatty acids have a chain length of about 22 or about 24 or more carbon atoms. The carbon atom bonds in the alkyl chain may all be single bonds (i.e., a saturated fatty acid) or may contain one or more double bonds (i.e., an unsaturated fatty acid). Unsaturated fatty acids having one double bond are also referred to as monoenoic; unsaturated fatty acids having two or more double bonds in the carbon chain are also referred to as polyenoic and polyunsaturated (PUFA). The carbon chain in a fatty acid may also be substituted with hydroxyl, methyl, or other groups in place of a hydrogen. Carboxylic acids, such as fatty acids, can partially dissociate in aqueous media and exist as undissociated, uncharged molecules and as a dissociated, anionic form.

Fatty acids containing one carboxyl group can also be referred to as monocarboxylic fatty acids. A fatty acid containing two carboxyl groups (e.g., α,ω-dicarboxylic acids) is a fatty dicarboxylic acid, also referred to herein as a diacid. An example of a diacid is adipic acid (hexanedioic acid) which contains six carbon atoms. A diacid sometimes is a C4 to a C24 diacid (i.e., a diacid containing 4 carbons to 24 carbons) and sometimes is a C8, C10, C12, C14, C16, C18, or C20 diacid. Diacids can contain an even as well as an odd number (e.g., C5, C7, C9, C11, C13, C15, C17, C19, C21 or C23) of carbons. A hydrocarbon portion of a diacid sometimes is fully saturated and sometimes a diacid includes one or more unsaturations (e.g., double bonds).

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1374 of 3077

Non-limiting examples of diacids include octadecanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid and other organic intermediates using biological systems. Non-limiting examples of fatty dicarboxylic acids include adipic acid (hexanedioic acid, 1,4-butanedicarboxylic acid), suberic acid (i.e., octanedioic acid, 1,8-octanedioic acid, octanedioic acid, octane-1,8-dioic acid, 1,6-hexanedicarboxylic acid, capryllic diacids), azelaic acid, sebacic acid (i.e., 1,10-decanedioic acid, decanedioic acid, decane-1,10-dioic acid, 1,8-octanedicarboxylic acid, capric diacid), undecanedioc acid, dodecanedioic acid (i.e., DDDA, 1,12-dodecanedioic acid, dodecanedioic acid, dodecane-1,12-dioic acid, 1,10-decanedicarboxylic acid, decamethylenedicaboxylic acid, 1,10-dicarboxydecane, lauric diacid), tetradecanedioic acid (i.e., TDDA, 1,14-tetradecanedioic acid, tetradecanedioic acid, tetradecane-1,14-dioic acid, 1,12-dodecanedicarboxylic acid, myristic diacid), thapsic acid (i.e., hexadecanedioic acid, 1,16-hexadecanedioic acid, hexadecanedioic acid, hexadecane-1,16-dioic acid, 1,14-tetradecanedicarboxylic acid, palmitic diacid), cis-9-hexadecenedioic acid (i.e., palmitoleic diacids), octanedioic acid (i.e., 1,18-octadecanedioic acid, octadecanedioic acid, octadecane-1,18-dioic acid, 1,16-hexadecanedicarboxylic acid, stearic diacid), cis-9-octadecenedioic acid (i.e., oleic diacids), cis-9,12-octadecenedioic acid (i.e., linoleic diacids), cis-9,12,15-octadecenedioic acid (i.e., linolenic diacids), arachidic diacid (i.e., eicosanoic diacid, icosanoic diacid), 11-eicosenoic diacid (i.e., cis-11-eicosenedioic acid), 13-eicosenoic diacids (i.e., cis-13-eicosenedioic acid), arachidonic diacid (i.e., cis-5,8,11,14-eicosatetraenedioic acid) and salts and esters of fatty acids, including, for example, any of the foregoing diacids.

Adipic acid and suberic acid are 6- and 8-carbon dicarboxylic acids, respectively, that are chemical intermediates in manufacturing processes used to make certain polyamides, polyurethanes and plasticizers. Azelaic acid, a 9-carbon dicarboxylic acid, is also used in therapeutic compositions due to its antibacterial and keratolytic activities. Sebacic acid, a 10-carbon dicarboxylic acid, is also used in cosmetics and candles and as an intermediate in producing aromatics and antiseptics. Dodecandioic acid (DDDA), a 12-carbon dicarboxylic acid, is widely used in forming polyamides, such as nylon. Some large-scale industrial processes for making adipic acid include (i) liquid phase oxidation of ketone alcohol oil (KA oil); (ii) air oxidation/hydration of cyclohexane with boric acid to make cyclohexanol, followed by oxidation with nitric acid; and (iii) hydrocyanation of butadiene to a pentenenitrile mixture, followed by hydroisomerization of adiponitrile, followed by hydrogenation. Suberic acid can be synthetically manufactured by oxidation of cyclooctene with ozone oxygen or ozone H_2O_2. Methods of

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1375 of 3077

chemical synthesis of sebacic acid include alkaline cleavage of ricinoleic acid and electrolytic dimerization of monomethyl adipate. DDDA is synthetically produced from butadiene in a multistep chemical process. These energy-requiring processes involve the use and/or production of toxic chemicals and/or solvents.

3-hydroxypropionic acid (3-HP or 3HP, used interchangeably herein, which collectively refers to 3-hydroxypropionic acid, a 3-hydroxypropionate salt or ester thereof, or mixtures thereof in any proportion) is a platform chemical that can be converted into a variety of valuable products such as poly(hydroxypropionate), 1,3-propanediol, ethyl 3-ethoxypropionate (EEP), malonic acid and acrylic acid. For example, 3-HP can be dehydrated to produce acrylic acid, which in turn can be esterified to produce methyl acrylate or aminated to produce acrylamide. Acrylamide can further be converted by dehydration to acrylonitrile, acrylonitrile can be condensed to produce adiponitrile and adiponitrile can be hydrolysed to produce hexamethylenediamine (HMDA). In addition, polymerized acrylic acid (with itself or with other monomers such as acrylamide, acrylonitrile, vinyl, styrene, or butadiene) can produce a variety of homopolymers and copolymers that are used in the manufacture of various plastics, coatings, adhesives, elastomers, latex applications, emulsions, leather finishings, and paper coating, as well as floor polishes and paints. Acrylic acid also can be used as a chemical intermediate for the production of acrylic esters such as ethyl acrylate, butyl acrylate, methyl acrylate, and 2-ethyl hexyl acrylate and superabsorbent polymers (glacial acrylic acid).

Polyketides

Polyketides are secondary metabolites polymerized from short-chain carboxylic acid units (e.g., acetate, proprionate, malonate and butyrate). Many polyketide-derived molecules are valuable pharmaceuticals such as antibiotics, antitumor agents and cholesterol-lowering drugs. Non-limiting examples of polyketides include triacetic acid lactone (TAL or 4-hydroxy-6-methyl-2-pyrone, used interchangeably herein) and 6-methylsalicylic acid (6-MSA or 2-hydroxy-6-methylbenzoic acid). TAL can be converted into end products such as sorbic acid and 1,3-pentadiene, and can serve as a precursor in the synthesis of compounds (e.g, phloroglucinol and resorcinol) used in production of resin and adhesive formulations.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1376 of 3077

Terpenes

Terpenes are made up of units of isoprene (C_5H_8) (isoprene, methylbuta-1,3-diene, hemiterpene are used interchangeably herein) that can be joined together in a variety of different combinations to generate thousands of terpene compounds. Terpenes can be categorized according to the number of isoprene units contained in the molecule: monoterpenes (2 isoprene units), sesquiterpenes (3 isoprene units), diterpenes (4 isoprene units), sesterterpenes (5 isoprene units), triterpenes (6 isoprene units), sesquaterpenes (7 isoprene units), tetraterpenes (8 isoprene units), polyterpenes (many isoprene units). The isoprene units can be joined "head-to-tail" in a linear chain or arranged in rings. Terpenes can be hydrocarbons or can contain other atoms, such as oxygen (e.g., alcohols, aldehydes and ketones) which are typically referred to as terpenoids. Terpenes are commercially valuable compounds with a variety of uses in the healthcare, food, cosmetics and chemical industries, including, but not limited to, uses as pharmaceuticals (e.g., anticancer and antimalarial drugs), nutraceuticals, supplements, antioxidants, fragrances, flavoring agents, food colorants and agricultural pest control agents. Included in the tetraterpenes are organic pigments (e.g., β-carotene and astaxanthin) referred to as carotenoids. These terpenes have many uses such as, for example, additives in food and feed stocks, precursors to vitamin A, antioxidants and supplements (e.g., lutein and lycopene).

Cellular carbon flux

Cells can obtain carbon atoms from external carbon sources such as, for example, carbohydrates, hydrocarbons, acids and alcohols. Upon entering the cell, the source molecule is metabolized through various chemical reactions depending on the carbon source, as well as other factors (e.g., oxygen and nutrient availability). Carbon atoms flow or flux through these metabolic reactions and are utilized in generating energy and in the production of cellular materials. The multiple series of coordinated reactions involved in metabolizing different carbon sources for different purposes are referred to as metabolic pathways and can be catabolic or anabolic. In catabolic pathways, the carbon source is broken down through oxidative reactions in which electrons are removed from substrates or intermediates, and, in the process, energy is generated and stored as adenosine triphosphate (ATP). Glycolysis is an example of a catabolic pathway in which a carbohydrate carbon source (e.g., glucose) is converted to pyruvate which is oxidatively decarboxylated to acetyl-CoA by the pyruvate dehydrogenase multi-enzyme complex in the mitochondria of eukaryotic cells and in the cytosol of prokaryotic cells. In this series of

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1377 of 3077

reactions, a multi-carbon source molecule is degraded into a 2-carbon acetyl group and carbon dioxide. The 2 carbons of the acetyl group are then incorporated into a citrate molecule in the tricarboxylic acid cycle (also referred to as the TCA, citric acid and Krebs cycle) in which additional ATP molecules are generated. Intermediates in the TCA cycle (e.g., citrate, α-ketoglutarate, succinyl-CoA and oxaloacetate) provide precursors in the synthesis of essential cellular components such as amino acids, fatty acids, nucleotides and porphyrins. The TCA cycle is considered an amphibolic pathway which combines both catabolic and anabolic functions. In another catabolic pathway, β-oxidation, a fatty acid carbon source is broken down into acetyl-CoA and chain-shortened acyl-CoA which in turn can enter another cycle of β-oxidation for further degradation. The acetyl-CoA molecules generated in β-oxidation, which occurs in peroxisomes in eukaryotic microorganisms and the cytosol of prokaryotic cells, are then utilized in ATP generation in the TCA cycle.

Acetyl-CoA generated through β-oxidation in microorganisms (e.g., yeast and bacteria) can also be used in the glyoxylate cycle, which is an anabolic pathway wherein 2-carbon acetyl units are converted to 4-carbon molecules that can be used for the biosynthesis of macromolecules. The glyoxylate cycle thus allows these microorganisms to utilize non-fermentable carbon sources, such as fatty acids, acetate and ethanol, as a sole carbon source. In the glyoxylate cycle, which is similar to the TCA cycle, isocitrate is cleaved directly into the 4-carbon succinate molecule, and the 2-carbon glyoxylate molecule, through the enzyme isocitrate lyase without the two decarboxylation steps that occur in the same conversion in the TCA cycle. Glyoxylate then condenses with acetyl-CoA generated through β-oxidation to produce malate which in turn is converted to oxaloacetate and then isocitrate. Succinate generated in the glyoxylate cycle can also reenter the TCA cycle to produce oxaloacetate. Malate and oxaloacetate produced in the glyoxylate cycle can be converted into phosphoenolpyruvate, which is the product of the first enzyme-catalyzed reaction in gluconeogenesis. Gluconeogenesis is another anabolic pathway and provides for synthesis of carbohydrates when non-carbohydrate carbon sources are available to cells. Microorganisms growing on non-fermentable carbon sources utilize gluconeogenesis to synthesize glucose-6-phosphate which is used in the synthesis of ribonucleotides and deoxyribonucleotides. The carbon skeletons for generation of glucose-6-phosphate are contained within oxaloacetate from the glyoxylate and TCA cycles. In gluconeogenesis, oxaloacetate is converted into pyruvate through phosphoenolpyruvate carboxykinase, followed by several reactions that ultimately yield glucose-6-phosphate.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1378 of 3077

Additional anabolic pathways in cells include reactions in the synthesis of lipids, including, for example, triacylglycerols (referred to interchangeably as triglycerides and TAG) and phospholipids. Lipids are a diverse group of compounds that are soluble in non-polar organic solvents but not in water. Fatty acids serve as building blocks in the synthesis of storage lipids (e.g., triacylglycerols and steryl esters) and membrane lipids (e.g., phospholipids and sphingolipids). For example, triacylglycerol is an ester of glycerol and three fatty acids. In the synthesis of triacylglycerols from free fatty acids internalized into microbial cells from the environment, the fatty acids are first activated with coenzyme A to form an acyl-CoA. The acyl-CoA is involved in two pathways of triacylglycerol synthesis: the glycerol-3-phosphate (G3P) pathway and the dihydroxyacetone phosphate (DHAP) pathway. Both pathways proceed through formation of phosphatidic acid and subsequently diacylglycerol which is then acetylated to form triacylglycerol. Phosphatidic acid can also be converted to cytidine diphosphate-diacylglycerol which is the precursor of all major phospholipids in cells.

Modification of Cellular Carbon Flux

The multiple routes of carbon metabolism in cells provide opportunities for loss of carbon from a production pathway for desired organic molecules in a cell. Such losses can result in decreased product yields, increased production times and costs, and overall decreased production process efficiency and economy. Cells, organisms and microorganisms and methods described herein provide systems for enhanced production of target molecules. In one aspect, production is enhanced through alteration of carbon flux in cell-based and microbial production systems. Through alteration of cellular carbon flux, carbon atoms that may have flowed or been transported into other metabolic processes (e.g., energy and/or cellular composition generation) in a cell are redirected and/or recycled and made available for use in organic target molecule production processes. In so doing, starting material loss is reduced and carbon sources are utilized to a fuller extent in the production of the desired molecules.

Acetyl coenzyme A

Acetyl coenzyme A (acetyl-CoA; used interchangeably herein) is a major precursor in cell-based or microbial production of many industrially important chemicals. The fatty acid biosynthesis pathway begins with the conversion of acetyl-CoA to malonyl-CoA. Similarly, organic acids, such as, for example 3-hydroxypropionic acid, and polyketides, such as triacetic acid lactone,

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1379 of 3077

can be synthesized using acetyl-CoA as a starting material. Additional high-value products that can be synthesized in reactions beginning with acetyl-CoA include terpenes, which can be generated from isopentenyl diphosphate produced through microbial mevalonate pathways in cells. Acetyl-CoA is formed from an acetyl group and coenzyme A (a derivative of pantothenate and cysteine) which are linked through a thioester bond. Acetyl-CoA is a central metabolite in carbon metabolism. It is the final carbon form resulting from the catabolism of external carbon sources and is the initial precursor carbon form in many of the cellular anabolic pathways and energy generation processes. Acetyl-CoA is formed in multiple locations of a eukaryotic cell depending on the metabolic pathway and/or carbon source through which it is generated. For example, acetyl-CoA generated through glycolysis is localized in the mitochondria, whereas acetyl-CoA generated through peroxisomal β-oxidation is localized to peroxisomes. Acetyl-CoA generated through metabolism of acetate or ethanol is localized to the cytoplasm. Typically, acetyl-CoA formed in any of these cellular locations is transferred to the mitochondrial matrix for use in the TCA cycle for the generation of energy and precursors of cellular constituents, although some acetyl-CoA localized to the cytoplasm can be used in the synthesis of oxaloacetate via initial conversion to malate.

Acetyl-CoA, due to its chemical nature, cannot freely cross biological membranes. Therefore, acetyl-CoA formed in peroxisomes and the cytoplasm is modified for transport to the mitochondria. The primary mechanism for transfer of acetyl-CoA into the mitochondria in eukaryotic cells is the carnitine shuttle in which the acetyl group of acetyl-CoA is reversibly linked to a carrier molecule, carnitine, which is able to traverse biological membranes. Acetyl-carnitine is generated and degraded by the action of carnitine acetyltransferases (e.g., EC 2.3.1.7). Peroxisomal acetyl-CoA not destined for the glyoxylate cycle is converted to acetyl-carnitine by carnitine O-acetyltransferase. Due to its smaller size compared to acetyl-CoA, acetyl-carnitine is able to diffuse through pores in the peroxisomal membrane, across the cytoplasm to mitochondria where it is converted back to acetyl-CoA by mitochondrial carnitine O-acetyltransferase. Yeast also have carnitine acetyltransferases that localize to the cytosol and/or to the outer mitochondrial membrane which, in some species, are encoded by *YAT1* genes. The enzymes encoded by these genes may convert cytosolic acetyl-carnitine to acetyl-CoA and carnitine.

Acetyl-carnitine uptake into mitochondria involves an acetyl-carnitine translocase which, in some yeast species, is encoded by a *CRC1* gene. Mitochondria possess two membranes with

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1380 of 3077

the outer membrane allowing free diffusion of metabolites and the inner membrane controlling metabolite transport with multiple membrane transport proteins. A mitochondrial inner-membrane transport protein (e.g., Crc1p) may function as an acetyl-carnitine transporter providing for transport of acetyl-carnitine into the mitochondrial matrix.

Certain aspects of the cells, microorganisms, compositions and methods provided herein involve cellular carbon flux modifications to capture the carbon atoms in the acetyl group of acetyl-CoA formed in cellular metabolism. In some embodiments, carbon flux is modified to capture acetyl group carbon atoms generated in organelles (e.g., peroxisomes) or membranes as they are transported through the cytosol in the form of acetyl-carnitine. In other embodiments, acetyl group carbons of organelle-generated acetyl-CoA are re-directed from the carnitine-assisted transport system and toward conversion to acetate. Acetate, unlike acetyl-CoA, is able to traverse membranes and enter the cytosol from organelles. In further embodiments, carbon atoms of acetyl groups in mitochondrial acetyl-CoA can also be captured from intermediates of the TCA cycle that move into the cytosol.

Acetyl-carnitine capture/conversion

Included in embodiments of the cells, microorganisms, compositions and methods provided herein are cell-based or microbial production platform systems and components thereof in which the amount of (a) acetyl-carnitine in the cell cytosol is modified and/or (b) carnitine acetyltransferase and/or carnitine acetyltransferase activity in the cell cytosol is/are modified. In some instances, the amount of (a) acetyl-carnitine in the cell cytosol is increased and/or decreased and/or (b) carnitine acetyltransferase and/or carnitine acetyltransferase activity in the cell cytosol is/are increased and/or decreased. For example, in some aspects, a cell or microorganism may be modified to increase cytosolic acetyl-carnitine, may be modified to decrease cytosolic acetyl-carnitine or may be modified to alternately increase and decrease cytosolic acetyl-carnitine depending on the conditions in which the modified cell or microorganism is cultured.

In one embodiment, the amount of acetyl group carbons in the cytosol in the form of acetyl-carnitine in transit from the peroxisome and other areas to the mitochondria is increased in a cell or microorganism through a reduction in, and/or slowing of, the entry of acetyl-carnitine into mitochondria from the cytosol. This provides an increased availability of substrate for cytosolic

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1381 of 3077

carnitine acetyltransferase to convert to acetyl-CoA, and effectively results in an increase in the generation of cytosolic acetyl-CoA. Some of the acetyl group carbons are thereby diverted from the mitochondria, and from utilization in metabolic processes therein, and are instead retained in the cytosol. In some embodiments, the amount and/or activity of carnitine acetyltransferase in the cytosol of a cell or microorganism is/are increased. This provides an increased conversion of acetyl-carnitine, such as that which is in transit from the peroxisome to the mitochondria, into acetyl-CoA in the cytosol. In some embodiments, the entry of acetyl-carnitine into mitochondria from the cytosol is reduced in a cell or microorganism, and the amount and/or activity of carnitine acetyltransferase in the cytosol of the cell or microorganism is/are increased.

Acetate capture/conversion

Included in embodiments of the microorganisms, compositions and methods provided herein are microbial production platform systems and components thereof in which acetyl group carbons of organelle-generated acetyl-CoA are directed toward conversion to acetate. Modification of carbon flux in this manner provides for a tight and precise control of the movement of the acetyl carbons because acetate may pass through some membranes, e.g., peroxisomal membranes, more readily than other membranes, e.g., mitochondrial inner membranes. Therefore, carbon atoms captured in the form of cytosolic acetate will be less readily transported into mitochondria for further metabolism, unlike carbon atoms captured in the form of cytosolic acetyl-carnitine. Provided herein are cells, microorganisms, compositions and methods in which cellular carbon flux has been modified through the altered de novo generation of cellular acetate. In particular embodiments, cellular carbon flux has been modified to increase the production of acetate in a cell and/or a particular cellular location. For example, in certain aspects, cells are modified to increase the production of acetate in peroxisomes. In some embodiments, acetyl group carbons are directed toward conversion to acetate and away from the carnitine-carrier transport system.

In embodiments in which modification of cellular acetate generation yields increased amounts of cytosolic acetate, the amount and/or activity of cytosolic acetyl-CoA synthetase (also referred to as ACS or acetate-CoA ligase and used interchangeably herein) can also be increased to provide for increased conversion of acetate to acetyl-CoA. For example, the genomic copy number of nucleic acids encoding acetyl-CoA synthetase can be increased and/or the promoter

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1382 of 3077

for the acetyl-CoA synthetase-encoding nucleic acid can be replaced with a stronger promoter or one that provides for a different pattern of expression in the cell or microorganism.

Citrate capture/conversion

Carbon atoms of acetyl groups in mitochondrial acetyl-CoA can also be captured from intermediates of the TCA cycle such as, for example, citrate molecules generated in the first step of the cycle through the citrate synthase-catalyzed condensation of acetyl-CoA and oxaloacetate. In another embodiment of the cell and microbial production systems and methods provided herein, carbon atoms incorporated into citrate that has been transferred to the cytosol are captured through the cleavage of citrate to oxaloacetate and acetyl-CoA by the enzyme ATP citrate lyase (i.e., ACL, used interchangeably herein). The capture of metabolite carbon in this manner diverts it from use in other metabolic processes and also results in an increase in the level cytoplasmic acetyl-CoA.

Acyl coenzyme A

Acyl-CoA synthetases (e.g., EC 6.2.1.3) are enzymes that catalyze the activation of free fatty acids in the cytoplasm into CoA esters (fatty acyl-CoA) which are involved in several metabolic pathways. For example, free fatty acids internalized into cells that become activated with coenzyme A to form an acyl-CoA are used in the synthesis of triacylglycerols via two pathways: the glycerol-3-phosphate (G3P) pathway and the dihydroxyacetone phosphate (DHAP) pathway. When free fatty acids are activated and used in cellular processes, such as lipid biosynthesis, the carbon atoms in the free fatty acids are not available for use in cell or microbial production of commercially important chemicals. Certain aspects of the cells, microorganisms, compositions and methods provided herein include one or more modifications to reduce or eliminate cytosolic activation of free fatty acids into acyl-CoA. An example of a modification to reduce or eliminate activation of cytosolic free fatty acids is reduction or elimination of the amount and/or activity of acyl-CoA synthetase in the cytoplasm.

Malonyl-CoA

Malonyl-CoA is a coenzyme A derivative of the dicarboxylic acid malonic acid that can serve as a precursor in the synthesis of numerous valuable organic molecules, including fatty acids and

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1383 of 3077

polyketides. Cytoplasmic acetyl-CoA can be converted to malonyl-CoA by the enzyme acetyl-CoA carboxylase (e.g., EC 6.4.1.2). A modification of cellular carbon flux that increases cytosolic acetyl-CoA alone may not be optimal for enhancing fatty acid or other organic acid production in an engineered cell or microbial system if there is not a concurrent increase in conversion of acetyl-CoA to malonyl-CoA. To maximize production efficiencies, included in the cells, microorganisms, compositions and methods provided herein are cellular carbon flux modifications that increase the amount of cytosolic malonyl-CoA.

Cells and organisms

Provided herein are modified cells and organisms. In particular embodiments, the modified cells and organisms have been manipulated in ways designed to alter the cellular flux of carbon to direct carbon atoms toward one or more biochemical events or cellular locations and/or away from other metabolic pathways or locations. Also provided herein are methods of producing modified cells and organisms.

Host cells and organisms

Modified cells and organisms provided herein can be generated by manipulation of an existing cell or organism. The terms “host,” “starting” or “parental” as used herein in reference to a cell or organism refers to such an existing cell or organism. Host cells and organisms include, for example, wild-type or native cells or organisms as they occur in nature in their genetically unmodified, predominant form, and mutant cells or organisms that have one or more genetic differences compared to a wild-type cell or organism. A host cell or organism can also be a cell or organism that has been genetically modified. A host cell or organism thus serves as a reference cell or organism with respect to a modified or engineered cell or organism obtained by manipulation of a host. Organisms or cells that can be used as host organisms or cells, or as a source for a nucleic acid, are publicly available, from, for example, American Type Culture Collection (Manassas, Virginia), Thermo Fisher Scientific (Waltham, MA) and Agricultural Research Culture Collection (NRRL; Peoria, Illinois).

Host or modified organisms include multicellular and single cell, or unicellular, organisms. Microscopic organisms, referred to interchangeably herein as a “microorganism,” “microbial cell” or “microbe,” are an example of a host or modified organism and are included in the term

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1384 of 3077

"organism." Many microorganisms are unicellular and often are capable of dividing and proliferating. Cells from non-microbial organisms can also be utilized as a host or modified organism or source for a heterologous polynucleotide.

Organisms can be prokaryotic (e.g., bacteria) and non-prokaryotic (e.g., eukaryotic). Examples of eukaryotic organisms include yeast, filamentous fungi, protists, plants, algae and amoeba. An organism or microorganism can include one or more of the following features: aerobe, anaerobe, filamentous, non-filamentous, monoploid, haploid, diploid, oleaginous, non-oleaginous, auxotrophic and/or non-auxotrophic.

Host cells or organisms or modified cells or organisms can be selected based on a variety of criteria depending, for example, on the methods of generating modified cells or organisms therefrom and the uses of the modified cells or organisms from which they are derived. Selection criteria can include inherent metabolic mechanisms, suitability for genetic manipulation, adaptability to a variety of or particular growth or culture conditions, and ease of large-scale maintenance for use in industrial production processes. For example, microorganisms often can be cultured at cell densities useful for industrial production of a target product, including in a fermentation device. Included among microorganisms that may be selected as a host or modified organism or source for a heterologous polynucleotide are fungi. Examples of fungi include, but are not limited to, yeast, *Aspergillus* fungi (e.g., *A. parasiticus*, *A. nidulans*), *Thraustochytrium* fungi, *Schizochytrium* fungi and *Rhizopus* fungi (e.g., *R. arrhizus*, *R. oryzae*, *R. nigricans*). In some embodiments, a host organism can be a fungus such as a yeast strain, an *A. parasiticus* strain that includes, but is not limited to, strain ATCC 24690, and in certain embodiments, a fungus is an *A. nidulans* strain that includes, but is not limited to, strain ATCC 38163.

In some embodiments, a modified cell or microorganism provided herein can be derived from any one of the following cell lines: ATCC 20362, ATCC 8862, ATCC 18944, ATCC 20228, ATCC 76982, LGAM S(7)1, ATCC 20336, ATCC 20913, SU-2 (ura3-/ura3-), ATCC 20962, ATCC 24690, ATCC 38164, ATCC 38163, H5343, ATCC 8661, ATCC 8662, ATCC 9773, ATCC 15586, ATCC 16617, ATCC 16618, ATCC 18942, ATCC 18943, ATCC 18944, ATCC 18945, ATCC 20114, ATCC 20177, ATCC 20182, ATCC 20225, ATCC 20226, ATCC 20228, ATCC 20237, ATCC 20255, ATCC 20287, ATCC 20297, ATCC 20306, ATCC 20315, ATCC 20320, ATCC 20324, ATCC 20341, ATCC 20346, ATCC 20348, ATCC 20362, ATCC 20363,

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1385 of 3077

ATCC 20364, ATCC 20372, ATCC 20373, ATCC 20383, ATCC 20390, ATCC 20400, ATCC 20460, ATCC 20461, ATCC 20462, ATCC 20496, ATCC 20510, ATCC 20628, ATCC 20688, ATCC 20774, ATCC 20775, ATCC 20776, ATCC 20777, ATCC 20778, ATCC 20779, ATCC 20780, ATCC 20781, ATCC 20794, ATCC 20795, ATCC 20875, ATCC 22421, ATCC 22422, ATCC 22423, ATCC 22969, ATCC 32338, ATCC 32339, ATCC 32340, ATCC 32341, ATCC 32342, ATCC 32343, ATCC 32935, ATCC 34017, ATCC 34018, ATCC 34088, ATCC 34922, ATCC 38295, ATCC 42281, ATCC 44601, ATCC 46025, ATCC 46026, ATCC 46027, ATCC 46028, ATCC 46067, ATCC 46068, ATCC 46069, ATCC 46070, ATCC 46330, ATCC 46482, ATCC 46483, ATCC 46484, ATCC 48436, ATCC 60594, ATCC 62385, ATCC 64042, ATCC 74234, ATCC 76598, ATCC 76861, ATCC 76862, ATCC 90716, ATCC 90806, ATCC 90811, ATCC 90812, ATCC 90813, ATCC 90814, ATCC 90903, ATCC 90904, ATCC 90905, ATCC 96028, ATCC 201089, ATCC 201241, ATCC 201242, ATCC 201243, ATCC 201244, ATCC 201245, ATCC 201246, ATCC 201247, ATCC 201248, ATCC 201249, ATCC 201847, ATCC MYA-165, ATCC MYA-166, ATCC MYA-2613, and ATCC MYA-4467. That is, in certain embodiments, an engineered cell or microorganism described herein can be generated from one or more of the aforementioned ancestral cell lines.

Yeast that can serve as a host organism, and that can be modified organisms, include, but are not limited to, ascomycetes, non-Saccharomyces ascomycetes, and basidiomycetes. Non-limiting examples of yeast include *Yarrowia* yeast (e.g., *Y. lipolytica* (formerly classified as *Candida lipolytica*)), *Candida* yeast (e.g., *C. revkaufi*, *C. viswanathii*, *C. pulcherrima*, *C. tropicalis*, *C. utilis*), *Blastobotrys* (formerly classified as *Arxula*) (e.g., *Blastobotrys adeninivorans* (formerly classified as *Arxula adeninivorans*), *Blastobotrys mokoenaii*), *Rhodotorula* yeast (e.g., *R. glutinus*, *R. graminis*), *Rhodosporidium* yeast (e.g., *R. toruloides*), *Saccharomyces* yeast (e.g., *S. cerevisiae*, *S. bayanus*, *S. pastorianus*, *S. carlsbergensis*), *Cryptococcus* yeast, *Trichosporon* yeast (e.g., *T. pullans*, *T. cutaneum*), *Pichia* yeast (e.g., *P. pastoris*) and *Lipomyces* yeast (e.g., *L. starkeyii*, *L. lipoferus*). In some embodiments, a suitable yeast is of the genus *Arachniotus*, *Aspergillus*, *Aureobasidium*, *Auxarthron*, *Blastobotrys*, *Blastomyces*, *Candida*, *Chrysosporuim*, *Debaryomyces*, *Coccidiodes*, *Cryptococcus*, *Gymnoascus*, *Hansenula*, *Histoplasma*, *Issatchenkia, Kluyveromyces*, *Lipomyces*, *Lssatchenkia*, *Microsporum*, *Myxotrichum*, *Myxozyma*, *Oidiodendron*, *Pachysolen*, *Penicillium*, *Pichia*, *Rhodosporidium*, *Rhodotorula*, *Saccharomyces*, *Schizosaccharomyces*, *Scopulariopsis*, *Sepedonium*, *Trichosporon*, or *Yarrowia*. In some embodiments, a suitable yeast is of the species *Arachniotus flavoluteus*, *Aspergillus flavus*, *Aspergillus fumigatus*, *Aspergillus niger*,

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1386 of 3077

Aureobasidium pullulans, *Auxarthron thaxteri*, *Blastobotrys adeninivorans*, *Blastomyces dermatitidis*, *Candida albicans*, *Candida dubliniensis*, *Candida famata*, *Candida glabrata*, *Candida guilliermondii*, *Candida kefyr*, *Candida krusei*, *Candida lambica*, *Candida lipolytica*, *Candida lustitaniae*, *Candida parapsilosis*, *Candida pulcherrima*, *Candida revkaufi*, *Candida rugosa*, *Candida tropicalis*, *Candida utilis*, *Candida viswanathii*, *Candida xestobii*, *Chrysosporuim keratinophilum*, *Coccidiodes immitis*, *Cryptococcus albidus* var. *diffluens*, *Cryptococcus laurentii*, *Cryptococcus neofomans*, *Debaryomyces hansenii*, *Gymnoascus dugwayensis*, *Hansenula anomala*, *Histoplasma capsulatum*, *Issatchenkia occidentalis*, *Isstachenkia orientalis*, *Kluyveromyces lactis*, *Kluyveromyces marxianus*, *Kluyveromyces thermotolerans*, *Kluyveromyces waltii*, *Lipomyces lipoferus*, *Lipomyces starkeyii*, *Microsporum gypseum*, *Myxotrichum deflexum*, *Oidiodendron echinulatum*, *Pachysolen tannophilis*, *Penicillium notatum*, *Pichia anomala*, *Pichia pastoris*, *Pichia stipitis*, *Rhodosporidium toruloides*, *Rhodotorula glutinus*, *Rhodotorula graminis*, *Saccharomyces cerevisiae*, *Saccharomyces kluyveri*, *Schizosaccharomyces pombe*, *Scopulariopsis acremonium*, *Sepedonium chrysospermum*, *Trichosporon cutaneum*, *Trichosporon pullans*, *Yarrowia lipolytica*, or *Yarrowia lipolytica* (formerly classified as *Candida lipolytica*). In some embodiments, a yeast is a *Y. lipolytica* strain that includes, but is not limited to, ATCC 20362, ATCC 8862, ATCC 18944, ATCC 20228, ATCC 76982 and LGAM S(7)1 strains (Papanikolaou S., and Aggelis G., *Bioresour. Technol*. 82(1):43-9 (2002)).

In certain embodiments, a yeast is a *C. tropicalis* strain, a *C. viswanathii* strain, a *Y. lipolytica* strain or a yeast strain that includes, but is not limited to, ATCC 20336, ATCC 20913, SU-2 (ura3-/ura3-), ATCC 20962, H5343 (beta oxidation blocked; US Patent No. 5648247) ATCC 20362, ATCC 8862, ATCC 18944, ATCC 20228, ATCC 76982, LGAM S(7)1, ATCC 8661, ATCC 8662, ATCC 9773, ATCC 15586, ATCC 16617, ATCC 16618, ATCC 18942, ATCC 18943, ATCC 18944, ATCC 18945, ATCC 20114, ATCC 20177, ATCC 20182, ATCC 20225, ATCC 20226, ATCC 20228, ATCC 20237, ATCC 20255, ATCC 20287, ATCC 20297, ATCC 20306, ATCC 20315, ATCC 20320, ATCC 20324, ATCC 20341, ATCC 20346, ATCC 20348, ATCC 20362, ATCC 20363, ATCC 20364, ATCC 20372, ATCC 20373, ATCC 20383, ATCC 20390, ATCC 20400, ATCC 20460, ATCC 20461, ATCC 20462, ATCC 20496, ATCC 20510, ATCC 20628, ATCC 20688, ATCC 20774, ATCC 20775, ATCC 20776, ATCC 20777, ATCC 20778, ATCC 20779, ATCC 20780, ATCC 20781, ATCC 20794, ATCC 20795, ATCC 20875, ATCC 22421, ATCC 22422, ATCC 22423, ATCC 22969, ATCC 32338, ATCC 32339, ATCC 32340, ATCC 32341, ATCC 32342, ATCC 32343, ATCC 32935, ATCC 34017, ATCC 34018,

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1387 of 3077

ATCC 34088, ATCC 34922, ATCC 38295, ATCC 42281, ATCC 44601, ATCC 46025, ATCC 46026, ATCC 46027, ATCC 46028, ATCC 46067, ATCC 46068, ATCC 46069, ATCC 46070, ATCC 46330, ATCC 46482, ATCC 46483, ATCC 46484, ATCC 48436, ATCC 60594, ATCC 62385, ATCC 64042, ATCC 74234, ATCC 76598, ATCC 76861, ATCC 76862, ATCC 90716, ATCC 90806, ATCC 90811, ATCC 90812, ATCC 90813, ATCC 90814, ATCC 90903, ATCC 90904, ATCC 90905, ATCC 96028, ATCC 201089, ATCC 201241, ATCC 201242, ATCC 201243, ATCC 201244, ATCC 201245, ATCC 201246, ATCC 201247, ATCC 201248, ATCC 201249, ATCC 201847, ATCC MYA-165, ATCC MYA-166, ATCC MYA-2613, and ATCC MYA-4467.

In certain embodiments, a yeast is a *Candida* species (i.e., *Candida* spp.) yeast. In some embodiments, suitable *Candida* species include, but are not limited to *Candida albicans*, *Candida dubliniensis*, *Candida famata*, *Candida glabrata*, *Candida guilliermondii*, *Candida kefyr*, *Candida krusei*, *Candida lambica*, *Candida lipolytica*, *Candida lustitaniae*, *Candida parapsilosis*, *Candida pulcherrima*, *Candida revkaufi*, *Candida rugosa*, *Candida tropicalis*, *Candida utilis*, *Candida viswanathii*, *Candida xestobii* and any other *Candida* spp. yeast described herein. Non-limiting examples of *Candida* spp. strains include, but are not limited to, sAA001 (ATCC 20336), sAA002 (ATCC 20913), sAA003 (ATCC 20962), sAA496 (US2012/0077252), sAA106 (US2012/0077252), SU-2 (ura3-/ura3-), H5343 (beta oxidation blocked; US Patent No. 5648247) strains. Any suitable strains from *Candida* spp. yeast may be utilized as parental strains for modification.

Examples of ascomycetes fungi include, but are not limited to, *Candida* spp., *Yarrowia* spp., *Blastobotrys* spp., *Aspergillus* spp., *Penicillium* spp., *Saccharomyces* spp., *Debaryomyces* spp., *Lipomyces* spp., *Fusarium* spp., *Paecilomyces* spp., *Trichoderma* spp., *Cladosporium* spp., *Pichia* spp., and *Neurospora* spp. Examples of basidiomycetes fungi include, but are not limited to, *Trichosporon* spp., *Rhodotorula* spp., *Rhodosporidium* spp., *Cryptococcus* spp., *Phaffia* spp., and *Xanthophyllomyces* spp.

Prokaryote organisms that can serve as host organisms, and that can be modified organisms, include, for example, Gram negative or Gram positive bacteria. Examples of bacteria include, but are not limited to, *Bacillus* (e.g., *B. subtilis*, *B. megaterium*), *Acinetobacter*, *Norcardia*, *Xanthobacter*, *Escherichia* (e.g., *E. coli* (e.g., strains DH10B, Stbl2, DH5-alpha, DB3, DB3.1), DB4, DB5, JDP682 and ccdA-over (e.g., U.S. Application No. 09/518,188))), *Streptomyces*,

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1388 of 3077

Erwinia, *Klebsiella*, *Serratia*
(e.g., *S. marcessans*), *Pseudomonas* (e.g., P. aeruginosa), *Salmonella* (e.g., *S. typhimurium*, *S. typhi*), *Megasphaera* (e.g., *Megasphaera elsdenii*). Bacteria also include, but are not limited to, photosynthetic bacteria (e.g., green non-sulfur bacteria (e.g., *Choroflexus* (e.g., *C. aurantiacus*), *Chloronema* (e.g., *C. gigateum*)), green sulfur bacteria (e.g., *Chlorobium* bacteria (e.g., *C. limicola*), *Pelodictyon* (e.g., *P. luteolum*), purple sulfur bacteria (e.g., *Chromatium* (e.g., *C. okenii*)), and purple non-sulfur bacteria (e.g., *Rhodospirillum* (e.g., *R. rubrum*), *Rhodobacter* (e.g., *R. sphaeroides*, *R. capsulatus*), and *Rhodomicrobium* (e.g., *R. vanellii*)).

Examples of cells from non-microbial organisms that can be utilized as a host cell or organism, engineered cell or organism or source for a heterologous polynucleotide include, but are not limited to, insect cells (e.g., *Drosophila* (e.g., *D. melanogaster*), *Spodoptera* (e.g., *S. frugiperda* Sf9 or Sf21 cells) and *Trichoplusa* (e.g., High-Five cells); nematode cells (e.g., *C. elegans* cells); avian cells; amphibian cells (e.g., *Xenopus laevis* cells); reptilian cells; mammalian cells (e.g., NIH3T3, 293, CHO, COS, VERO, C127, BHK, Per-C6, Bowes melanoma and HeLa cells); and plant cells (e.g., *Arabidopsis thaliana*, *Nicotania tabacum*, *Cuphea acinifolia*, *Cuphea aequipetala*, *Cuphea angustifolia*, *Cuphea appendiculata*, *Cuphea avigera*, *Cuphea avigera* var. pulcherrima, *Cuphea axilliflora*, *Cuphea bahiensis*, *Cuphea baillonis*, *Cuphea brachypoda*, *Cuphea bustamanta*, *Cuphea calcarata*, *Cuphea calophylla*, *Cuphea calophylla* subsp. *mesostemon*, *Cuphea carthagenensis*, *Cuphea circaeoides*, *Cuphea confertiflora*, *Cuphea cordata*, *Cuphea crassiflora*, *Cuphea cyanea*, *Cuphea decandra*, *Cuphea denticulata*, *Cuphea disperma*, *Cuphea epilobiifolia*, *Cuphea ericoides*, *Cuphea flava*, *Cuphea flavisetula*, *Cuphea fuchsiifolia*, *Cuphea gaumeri*, *Cuphea glutinosa*, *Cuphea heterophylla*, *Cuphea hookeriana*, *Cuphea hyssopifolia* (Mexican-heather), *Cuphea hyssopoides*, *Cuphea ignea*, *Cuphea ingrata*, *Cuphea jorullensis*, *Cuphea lanceolata*, *Cuphea linarioides*, *Cuphea llavea*, *Cuphea lophostoma*, *Cuphea lutea*, *Cuphea lutescens*, *Cuphea melanium*, *Cuphea melvilla*, *Cuphea micrantha*, *Cuphea micropetala*, *Cuphea mimuloides*, *Cuphea nitidula*, *Cuphea palustris*, *Cuphea parsonsia*, *Cuphea pascuorum*, *Cuphea paucipetala*, *Cuphea procumbens*, *Cuphea pseudosilene*, *Cuphea pseudovaccinium*, *Cuphea pulchra*, *Cuphea racemosa*, *Cuphea repens*, *Cuphea salicifolia*, *Cuphea salvadorensis*, *Cuphea schumannii*, *Cuphea sessiliflora*, *Cuphea sessilifolia*, *Cuphea setosa*, *Cuphea spectabilis*, *Cuphea spermacoce*, *Cuphea splendida*, *Cuphea splendida* var. viridiflava, *Cuphea strigulosa*, *Cuphea subuligera*, *Cuphea teleandra*, *Cuphea thymoides*, *Cuphea tolucana*, *Cuphea urens*, *Cuphea utriculosa*, *Cuphea viscosissima*, *Cuphea watsoniana*, *Cuphea wrightii*, *Cuphea lanceolata*).

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1389 of 3077

In some embodiments, host organisms, or modified organisms, can be hydrocarbon-utilizing (e.g. alkane-utilizing), fatty acid-utilizing and/or fatty alcohol-utilizing microorganisms. These organisms are able to assimilate hydrocarbons, fatty acids and/or fatty alcohols for energy and biomass generation. Many of these organisms are able to utilize hydrocarbons, fatty acids and/or fatty alcohols as a sole carbon source. Some examples of hydrocarbon-, fatty acid- and/or fatty alcohol-utilizing microorganisms include some species of fungi (including, e.g., yeast), bacteria and algae. Non-limiting examples of such organisms include *Yarrowia* (e.g., *Y. lipolytica* (formerly classified as *Candida lipolytica*)), *Candida* (e.g., *C. apicola*, *C. maltosa*, *C. tropicalis*, *C. utilis*, *C. viswanathii*, *C. catenulate*, *C. rugose*, *C. vini*, *C. entamophila*, *C. intermedia*), *Aspergillus* (e.g., *A. niger*, *A. versicolor*, *A. ustus*, *A. fumigatus*, *A. oryzae*, *A. flavus*, *A. ficuum*, *A. terricola*, *A. japonicas*, *A. wentii*, *A. clavatus*, *A. terreus*), *Penicillium* (*P. cyclopium*, *P. chrysogenum*, P. *italicum*), *Fusarium* (e.g., *F. oxysporum*, *F. moniliforme*, *F. solani*), *Paecilomyces* (e.g., *Paec. lilacinus*), *Trichoderma* (e.g., *T. koningii*, *T. viride*, *T. virens*), *Cladosporium* (e.g., *C. herbarum*), *Stachybotrys*, *Trichosporon* (e.g., *T. veenhuisii*, *T. asahii*, *T. jirovecii*, *T. monteviblankiideense*), *Rhodotorula* (e.g., *R. glutinous*, *R. mucilaginosa*), *Rhodosporidium* (e.g., *R. toruloides*), *Cryptococcus* (e.g., *C. neoformans*, *C. albidus*), *Pichia* (e.g., *P. farinosa*, *P. stipitis*), *Debaryomyces* (e.g., *D. hansenii*), *Blastobotrys* (e.g., *Blastobotrys adeninivorans*), *Saccharomyces* (e.g., *S. cerevisiae*, *S. bayanus*, *S. pastorianus*, *S. carlsbergensis*), *Lipomyces* (e.g., *L. starkeyii*, *L. lipoferus*) and *Chlorella* algae (e.g., *Chlorella protothecoides*).

In some embodiments, a host organism or modified organism can be an oleaginous organism (e.g., an oleaginous microorganism). As used herein, an “oleaginous” organism is an organism capable of accumulating at least about 20% or more of its cell mass (by dry weight) as intracellular lipids (e.g., oil). In oleaginous organisms, a significant carbon flux towards lipid synthesis occurs and is enhanced under certain conditions (e.g., limited supply of nitrogen). These lipid-accumulating organisms can be characterized by the endogenous expression of cytosolic ATP citrate lyase, which catalyzes the degradation of citrate generated in the TCA cycle into acetyl-CoA and oxaloacetate, and/or a dependence on AMP concentration for the activity of isocitrate dehydrogenase in the TCA cycle. Generally, under certain conditions (e.g., limited nitrogen), AMP deaminase is activated in oleaginous yeast which can lead to a decrease in mitochondrial AMP concentration and isocitrate dehydrogenase activity. This, in turn, can cause an accumulation of mitochondrial citrate from the TCA cycle which is then exported to the

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1390 of 3077

cytosol and can serve as substrate for ATP citrate lyase. The acetyl-CoA that may be generated through the action of ATP citrate lyase can be used in synthesizing fatty acyl-CoA that can be converted into lipids which may be stored in lipid bodies in the cells. A "non-oleaginous" organism, as used herein, is an organism that is not capable of accumulating at least about 20% or more of its cell mass (by dry weight) as intracellular lipids. In some embodiments, a host organism or modified organism can be a non-oleaginous organism.

Oleaginous microorganisms include species of fungi, bacteria and algae. Examples of oleaginous fungi include, but are not limited to, *Blastobotrys* (e.g., *Blastobotrys adeninivorans*), *Yarrowia* (e.g., *Y. lipolytica*), *Trichosporon* (e.g., *T. fermentans*, *T. porosum*, *T. pullulan*), *Rhodotorula* (e.g., *R. graminis*, *R. glutinous*, *R. araucariae*, *R. minuta*, *R. bogoriensis*, *R. mucilaginosa*, *R. colostri*), *Rhodosporidium* (e.g., *R. toruloides*, *R. kractochvilovae*, *R. paludigenum*, *R. fluviale*, R. *babjevae*), *Lipomyces* (e.g., *L. starkeyii*, *L. lipofer*), *Debaryomyces* (e.g., *D. hansenii*), *Cryptococcus* (e.g., *C. podzolicus*, *C. phenolicus*, *C. curvatus*), *Pichia* (e.g., *P. segobiensis*), *Cystofilobasidium* (e.g., *C. informiminiatum*), *Leucosporidium* (e.g., *L. scottii*), *Sporobolomyces* (e.g., *S. singularis*, *S. poonsookiae*, *S. odoratus*, *S. metaroseus*, *S. bannaenis*), *Sporidiobolus* (e.g., *S. ruineniae*, *S. carnicolor*, S. pararoseus, *S. johnsonii*), *Schwanniomyces* (e.g., *S. occidentalis*), *Occultifur* (e.g., *O. externus*), *Blakeslea*, *Cunninghamella*, *Mortirella*, *Mucor*, *Phycomyces* and *Pythium*. Nonlimiting examples of oleaginous bacteria include *Morrococcus*, *Bacillus subtilis* and *Rhodococcus opacus*. Examples of oleaginous algae include, but are not limited to *Nannochloropsis* (e.g., *N. oceania*), *Chlorella* (e.g., *C. vulgaris*), *Thraustochyrtium* and *Schizochytrium*.

In particular embodiments, a host organism or modified organism can be a non-oleaginous yeast. In some embodiments, a host organism or modified organism can be a non-oleaginous, non-*Saccharomyces* yeast. Included among such yeast are non-oleaginous, non-*Saccharomyces* ascomycetes yeast as well as non-oleaginous, basidiomycetes yeast and non-oleaginous, ascomycetes yeast. In certain aspects, a host organism or modified organism can be an oleaginous, non-*Yarrowia* yeast, a non-*Yarrowia* ascomycetes yeast, or a non-*Yarrowia*, non-*Saccharomyces*, ascomycetes yeast. In another aspect, a host organism or modified organism can be an oleaginous yeast that accumulates 20% to 65% or 20% to 60% or 20% to 58%, or 20% to 55%, or 20% to 50% or 20% to 45%, or 20% to 40% or 20% to 35%, or 20% to 30% or 20% to 25% of its cell mass (by dry weight) as intracellular lipids (e.g., oil). In another embodiment, a host organism or modified organism can be an oleaginous yeast that

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1391 of 3077

accumulates at least 20% or at least about 25% of its cell mass (by dry weight) as intracellular lipids (e.g., oil). In a further embodiment, a host organism or modified organism can be an oleaginous yeast that accumulates at least 20% but less than 70%, or at least 20% but less than 60%, or at least 20% but less than 50%, or at least 20% but less than 40%, or at least 20% but less than 30%, of its cell mass (by dry weight) as intracellular lipids (e.g., oil). In another embodiment, a host organism or modified organism can be an oleaginous yeast in which linoleic acid is less than 50% or less than 45% or less than 40% or less than 35% or less than 30% or less than 25% or less than 20% of the intracellular accumulated lipid composition.

In some embodiments, a host cell or organism or modified cell or organism is one that is capable of ω-oxidation of alkanes and/or fatty acids. Such cells or organisms can endogenously produce enzymes of the ω-oxidation pathway. This pathway includes steps of ω-hydroxylation, oxidation and dehydrogenation of ω-carbon. The ω-hydroxylation step can be catalyzed by a hydroxylase complex including a cytochrome P450 monooxygenase (such as, for example, an alkane-inducible cytochrome P450, e.g., CYP52) and a cytochrome P450:NADPH oxidoreductase which yields an alcohol. In a subsequent oxidation step, the alcohol is further oxidized to an aldehyde in a reaction catalyzed by a fatty alcohol oxidase. A dicarboxylic acid is generated through dehydrogenation of the aldehyde by a fatty aldehyde dehydrogenase. In some aspects, a host cell or organism endogenously expresses proteins having cytochrome P450 monooxygenase and cytochrome P450:NADPH oxidoreductase activity. In some embodiments, a host cell or organism or modified cell or organism is one that is capable of synthesizing dicarboxylic acids, such as, for example, α,ω-dicarboxylic acids.

In some embodiments, the host cell or organism or modified cell or organism is a diploid cell or organism. In certain aspects, the host cell or organism or modified cell or organism is an anamorphic ascomycetes yeast.

In some embodiments, a host organism or modified organism can be a thermotolerant and/or osmotolerant organism. As used herein, “thermotolerant,” in reference to an organism, e.g., a microorganism, refers to the ability of the organism to survive at elevated temperatures. For example, a thermotolerant organism, e.g., a microorganism, such as yeast, is one that is able to survive and/or grow and/or assimilate fatty acids and/or aliphatic carbon sources at temperatures greater than 30°C, greater than 31°C, greater than 32°C, greater than 33°C, greater than 34°C, greater than 35°C, greater than 36°C, greater than 37°C, greater than 38°C,

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1392 of 3077

greater than 39°C, greater than 40°C, greater than 41°C, greater than 42°C, greater than 43°C, greater than 44°C, greater than 45°C, greater than 46°C, greater than 47°C, or greater than 48°C. A thermotolerant organism can be one that is able to survive and/or grow and/or assimilate fatty acids and/or aliphatic carbon sources at temperatures up to about 30°C, 32°C, 34°C, 35°C, 40°C, 41°C, 42°C, 43°C, 44°C, 45°C, 46°C, 47°C, 48°C, 49°C or 50°C or more. As used herein, "osmotolerant," in reference to an organism, e.g., a microorganism, refers to the ability of the organism to survive in elevated external osmotic pressure environments, e.g., high solute (such as salt or sugar) concentrations). For example, an osmotolerant organism, e.g., a microorganism, such as yeast, is one that is able to survive and/or grow and/or assimilate fatty acids and/or aliphatic carbon sources in media containing up to about 15%, 16%, 17%, 18%, 19%, 20%, 21%, 22%, 23%, 24%, 25% or greater NaCl. Thermotolerant and/or osmotolerant microorganisms include, for example, species of *Blastobotrys* yeast (e.g., *Blastobotrys adeninivorans*), *Candida* yeast (e.g., *C. Mexicana, C. glycerinogenes*, *C. zemplinina*), *Pichia* yeast (e.g., *P. mississippiensis*, *P. mexicana*, *P. farinosa*, *P. sorbitophila*), *Clavispora* yeast (e.g., *C. opuntiae*, *C. lusitaniae*), *Kluyveromyces* yeast (e.g., *K. thermotolerans*), *Debaryomyces* (e.g., *D. hansenii*), *Rhodotorula* (e.g., *R. mucilaginosa*), *Zygosaccharomyces* (e.g., *Z. rouxii*) and *Issatchenkia* (e.g., *I. orientalis*). Thermotolerant and/or osmotolerant organisms can be well suited for use in industrial production systems operating at elevated temperatures and/or osmotic pressures that would impair growth and/or metabolism and/or completely inactivate organisms that are not thermotolerant and/or osmotolerant. Furthermore, in many instances, production efficiency can be improved and production costs reduced in using such organisms due to decreases in losses and avoidance of implementation of cooling processes.

Host cells and microorganisms and engineered cells and microorganisms may be provided in any suitable form. For example, such cells and microorganisms may be provided in liquid culture or solid culture (e.g., agar-based medium), which may be a primary culture or may have been passaged (e.g., diluted and cultured) one or more times. Microorganisms and cells also may be provided in frozen form or dry form (e.g., lyophilized). Microorganisms and cells may be provided at any suitable concentration.

Modified cells and organisms

Provided herein are cells and organisms (including microorgansims) that have been modified in one or more aspects relative to the unmodified cell or organism (i.e., the cell or organism prior to

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1393 of 3077

the modification). For example, a cell or organism can be modified by altering one or more cellular activities and/or the sum total of a cell's or organism's activities. Thus, in this example, modifications can include alteration of cellular activities, addition of cellular activities and/or elimination of cellular activities. A "cellular activity," as used herein, refers to any process, functioning, or operation that can occur in a cell. In particular embodiments provided herein, a cell or organism has been modified to alter cellular carbon flux. Such modified cells and organisms have been manipulated in ways designed to direct carbon atoms toward one or more biochemical events, cellular activities or cellular locations and/or away from other metabolic pathways, cellular activities or locations. The alteration(s) can involve a single modification or multiple modifications of the original, or host, cell or organism in which carbon flux is altered. Also provided herein are methods of producing such modified cells and organisms. As described herein, there are multiple methods of altering cellular carbon flux by modifying one or more aspects of carbon processing in cells. Aspects of cellular carbon processing include, for example, but are not limited to, fatty acid metabolism, including fatty acid catabolism and synthesis, ω-oxidation, β-oxidation, fatty acid transport, acetyl group transfer/transport and processing, the TCA cycle, metabolite processing and triacylglyceride and lipid biosynthesis. For example, in altering carbon flux, certain cellular activities may be reduced, slowed or eliminated and/or other activities may be increased, accelerated, added or relocated. In particular embodiments, the amount and/or activity of one or more enzymes and/or transport proteins is/are modified in cells or microorganisms.

As such, the cells and organisms provided herein are "modified" or "engineered." The terms "engineered" or "modified," as used interchangeably herein, in reference to a cell, organism or microorganism refer to a cell or organism (including a microorganism) that has been manipulated or altered such that it is distinct (e.g., detectably changed or physically different) from a naturally occurring cell or organism. For example, the sum total of the cellular activities of a modified or engineered cell or microorganism can be distinct from those of a naturally occurring cell or microorganism, e.g., a modified cell or microorganism may include or lack one or more activities relative to the activities present in an unmodified cell or microorganism utilized as a starting point (e.g., host cell, host organism or host microorganism) for modification. In another example, one or more cellular activities of a modified or engineered cell or microorganism may be altered relative to the cellular activity or activities of the host cell or microorganism. A modified or engineered cell or organism can be genetically modified through any alteration in its genetic composition. For example, a genetically modified cell or organism

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1394 of 3077

can include one or more heterologous polynucleotides, can have one or more endogenous nucleic acid deletions and/or can have one or more genetic mutations. Mutations include point mutations, insertions and deletions of a single or multiple residues in a nucleic acid. In some embodiments, an engineered cell, organism or microorganism includes a heterologous polynucleotide, and in certain embodiments, an engineered cell, organism or microorganism has been subjected to selective conditions that alter an activity, or introduce an activity, relative to the host cell or microorganism. Thus, a modified or engineered cell, organism or microorganism has been altered directly or indirectly by a human being. It is understood that the terms "modified cell," "modified organism," "modified microorganism," "engineered cell," " engineered organism," "engineered microorganism," refer not only to the particular cell or organism but to the progeny or potential progeny of such a cell or organism. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

For example, a "genetically modified" or "genetically engineered" cell, organism or microorganism is one in which the genetic make-up of the cell, organism or microorganism has been modified. Genetic modification encompasses a variety of alterations and can be accomplished in numerous ways. A genetic modification includes, but is not limited to, any of the following alterations: modification of the expression of an endogenous gene (e.g., the amount, pattern, timing and/or regulation (e.g., inducibility) of expression of a gene), disruption or deletion of an endogenous gene, increasing the copy number of an endogenous gene, mutation of an endogenous gene (including the regulatory components, exons, introns and/or peptide- or protein-encoding portions of a gene), and introduction of heterologous nucleic acid in to a cell or cells. These genetic modifications, and others, are described herein.

A genetic modification of a cell or organism can be one that modifies the expression of one or more nucleic acids or polypeptides in the cell or organism. A genetic modification of a cell or organism can be one that modifies the amount and/or activity of a polypeptide in the cell or organism. For example, modified expression of a nucleic acid or protein (e.g., modified rate, amount and/or level of expression) or modified amount or activity of a polypeptide may be a reduction, slowing, decrease or elimination, or increase, acceleration, addition or elevation in expression of a nucleic acid or protein or in the amount and/or activity of a polypeptide.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1395 of 3077

Modified expression of a nucleic acid or modification of the amount and/or activity of a polypeptide may be a relocation of expression or activity within a cell.

In one example, a genetic modification of a cell or organism can be one that alters the expression of, or the amount and/or activity of, a polypeptide involved in a reaction that generates a product (e.g., cytosolic acetyl-CoA, cytosolic malonyl-CoA, peroxisomal acetate) in a cell or organism. A "polypeptide involved in a reaction that generates," as used herein with respect to generation of a product, refers to a polypeptide that participates in the direct generation of the product from reactants. A reaction that directly generates a particular product can be a single-step reaction or a multi-step reaction involving transient reaction intermediates. For example, a polypeptide involved in a reaction that generates cytosolic acetyl-CoA is one that participates in a reaction that directly yields acetyl-CoA in the cytosol. Exemplary polypeptides (e.g., enzymes) involved in a reaction that generates cytosolic acetyl-CoA include, but are not limited to, cytosolic carnitine acetyltransferase, cytosolic acetyl-CoA synthetase and cytosolic ATP citrate lyase. A polypeptide involved in a reaction that generates cytosolic malonyl-CoA is one that participates in a reaction that directly yields malonyl-CoA in the cytosol. A non-limiting example of a polypeptide involved in a reaction that generates cytosolic malonyl-CoA is acetyl-CoA carboxylase. A polypeptide involved in a reaction that generates peroxisomal acetate is one that participates in a reaction that directly yields acetate in the peroxisome. A non-limiting example of a polypeptide involved in a reaction that generates peroxisomal acetate is acetyl-CoA hydrolase.

The term "endogenous," as used herein in reference to an aspect (e.g., a gene, nucleic acid, peptide, polypeptide, activity, genetic composition, gene expression, and the like) of a cell or organism or microorganism refers to the inherent aspect, or condition thereof, in the cell, organism or microorganism that has not been modified or engineered (i.e., the reference cell, organism or microorganism). The term "heterologous," "exogenous" or "foreign" as used herein with respect to a composition or quality (e.g., a gene, nucleic acid, peptide, polypeptide, cellular activity, genetic composition, gene expression, and the like) refers to the composition or quality not being a physically existing part or attribute of a reference cell, organism or microorganism. For example, a heterologous, exogenous or foreign nucleic acid can be any nucleic acid that is introduced into a cell or microorganism as part of a genetic modification of the cell or microorganism. A heterologous, exogenous or foreign composition (such as, for example, a nucleic acid) includes compositions that may be identical to an endogenous composition (e.g., a

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1396 of 3077

nucleic acid gene sequence that is introduced into a cell or microorganism to increase the copy number and/or alter the positioning or expression of the same nucleic acid sequence existing therein) or may be different from an endogenous composition.

Coordination of carbon source, host organism and regulatory mechanisms to optimize carbon flux modification

In developing cell- and organism-based systems for enhanced production of target molecules, there are multiple factors, in addition to the design of cellular modifications for altering carbon flux, that can affect the overall efficiency and economics of the production process. These additional considerations include the sources of carbon available to the microorganism, the organism's ability to utilize various forms of carbon in the sources and the cellular regulatory systems that can be used in controlling carbon flux. Coordination of these factors can play a significant role in optimization of carbon flux alteration and, in turn, the efficiency of target molecule production.

Carbon sources used for culturing cells and microorganisms and/or fermentation processes sometimes are referred to as feedstocks. The term "feedstock" as used herein refers to a composition containing a carbon source that is provided to a cell or organism, which is used by the cell or organism to produce energy and metabolic products useful for growth. In order for cells and organisms to utilize carbon in vital processes, the carbon source is processed intracellularly in catabolic pathways to a form(s) that can be accommodated by energy generation and biosynthetic pathways. For example, glucose is processed in glycolytic pathways in cells whereas fatty acids are processed through β-oxidation. Thus, the carbon source used in microbial-based methods of target molecule production can influence which metabolic pathways will be involved in assimilating the carbon. A target molecule production system that incorporates elements of endogenous cellular metabolic pathways may not perform optimally if those pathways are not utilized in processing the carbon source.

Some cells and microorganisms are able to utilize a variety of carbon sources. However, many cells and microorganisms preferentially utilize particular carbon sources over others, and some cells and microorganisms are unable to utilize certain carbon sources. For example, *Saccharomyces cerevisiae* can utilize xylulose but not xylose. *Blastobotrys adeninivorans* and

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1397 of 3077

Arxula terestre are able to utilize carbon- and nitrogen-containing compounds, e.g., adenine, uric acid, butylamine and pentylamine, as a sole source of carbon and nitrogen.

One consideration in the design of an economically feasible cell-based system for the production of target molecules is production-associated costs. The carbon source used in cultivating cells and organisms can be a significant factor contributing to production costs. Many microorganisms, including yeast, preferentially use glucose over other carbon sources. However, glucose is a relatively high-cost carbon source. Therefore, from an economic perspective, it can be beneficial to utilize lower-cost sources of carbon in bioproduction systems. Non-fermentable carbon sources, including, for example, glycerol and fatty acids, may be lower-cost alternatives to glucose and other carbohydrates in feedstocks. For example, waste materials, such as waste cooking oil, can be used as feedstocks containing non-fermentable carbon sources.

Therefore, in developing a cost-effective, efficient cell- or microbial-based target molecule production system, the modifications made to an organism to alter carbon flux should be coordinated and compatible with, and complementary to, the carbon source and cell or organism that will be employed in the production methods. Additionally, for optimal target molecule production, the regulatory mechanisms that are used in the cell or organism for controlling the individual elements (e.g., enzyme expression) being modified should provide for timing and extent of each element's activity that correlates with the desired carbon flux alterations at each stage of the production process.

For example, some embodiments of the cell- and microbial-based methods for producing target molecules provided herein include cells or organisms in which carbon processing activities have been engineered to enhance carbon flow through cellular oxidative metabolism pathways, e.g, ω-oxidation and/or β-oxidation. One advantage of such modified bioproduction systems is that they are well suited for use with lower cost, alternative carbon sources, including, for example, non-carbohydrate and non-fermentable carbon sources such as aliphatic compounds and hydrocarbons (e.g, alkanes, fatty acids and fatty alcohols). Use of such carbon sources is not only more cost-effective but can also have the added advantage of reducing the environmental impact of harmful wastes (e.g., agro-industrial by-products, waste cooking oil and waste motor oil) that can be used as feedstocks in target molecule production instead of being discarded. Cells or organisms particularly compatible with such methods are those that are able to utilize

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1398 of 3077

non-fermentable, as well as fermentable, carbon sources. Generally, such cells and organisms contain endogenous metabolic pathways that form part of the basis for the desired carbon flux modifications. As also described herein, embodiments of the cell- and microbial-based systems in which carbon processing activities have been engineered to direct carbon flow through oxidative metabolism and away from mitochondrial metabolism can be controlled to provide for maximal, coordinated and highly efficient target molecule production based on, for example, use of carbon source-dependent transcription regulation of modified activities in the cells.

Transcription regulatory elements, including promoters, for some genes are responsive to the carbon source available to the cells. For example, transcription of some genes is subject to glucose repression in which the gene may not be expressed, or is less expressed, in the presence of glucose. Thus, in contrast to unregulated constitutive promoters, transcription regulatory elements for genes such as these are repressed, derepressible and/or inducible by varying carbon sources. When glucose is depleted, genes that were subject to glucose repression are then transcribed in a process referred to as glucose derepression. For some of these genes, this increase in transcription due to derepression represents the extent to which the genes will be expressed because they are not subject to induction and further increased transcription. For others of these genes, transcription may be increased several-fold over the derepressed level upon induction by, for example, certain carbon sources. Examples of such carbon sources include, but are not limited to, vegetable oils, triglycerides, fatty acids, e.g, oleic acid, esters of fatty acids and n-alkanes. Some genes encoding peroxisomal proteins (including enzymes involved in fatty acid catabolism) are subject to glucose repression/derepression. As described herein, the transcription regulatory elements for genes subject to glucose repression can advantageously be used in cell- and microbial-based methods for target molecule production involving alternative carbon sources.

Different carbon sources or feedstocks may be used in culturing cells or microorganisms at different phases of a target molecule production process. For example, one carbon source, e.g., glucose, may be used in preparing an initial starter culture of modified cells or microorganisms to establish a foundation of growing cells and a different carbon source (e.g., a lower-cost alternative such as fatty acids) may be used in a target molecule production phase subsequent to establishment of the starter culture. Accordingly, carbon source utilization can vary depending on the goal of a particular time or phase of a culture process.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1399 of 3077

In some embodiments of cell- and microbial-based target molecule production systems provided herein, modifications made to the cells or microorganisms include use of carbon source-dependent regulatory elements in altering carbon flux to enhance production efficiency. As described herein, in some embodiments, cytosolic activities for generating acetyl-CoA and/or malonyl-CoA (e.g., carnitine acetyltransferase, acetyl-CoA carboxylase, acetyl-CoA synthetase and/or ATP citrate lyase) can be increased for target molecule production during fatty acid or alkane assimilation, while mitochondrial and/or cytosolic activities for uptake and utilization of acetyl group carbons (e.g., cytosolic acyl-CoA synthetase, mitochondrial acetyl-carnitine transporters and/or mitochondrial carnitine acetyltransferase) are decreased. For optimal coordination of these activities with the overall production process, in some embodiments, the expression of proteins involved in some of the target molecule production activities can be engineered to be controlled by particular glucose-repressible and/or fatty acid-inducible transcription regulatory elements. For example, during initial cell-growth stages of a production method, a preferred carbon source may be, for example, glucose. Activities participating in target molecule production can be suppressed during this stage by using glucose-repressible elements, e.g., promoters, to regulate transcription of nucleic acids encoding proteins involved in those activities. At the same time, activities (e.g., mitochondrial metabolism) involved in cellular energy generation for growth can be permitted to function and/or elevated by using constitutive or glucose-inducible elements, e.g., promoters, to regulate transcription of nucleic acids encoding proteins involved in those activities. On the other hand, activities participating in target molecule production will be permited and/or increased following growth stages and during target molecule production stages when glucose is depleted and fatty acid carbon sources are provided by using glucose-repressible and/or fatty acid-inducible elements, e.g., promoters, to regulate transcription of nucleic acids encoding proteins involved in those activities. Also during those stages, the activities (e.g., mitochondrial metabolism) involved in cellular energy generation and growth can be unchanged or reduced or minimized by using weak, constitutive and/or glucose-inducible/fatty acid-inhibited elements, e.g., promoters, to regulate transcription of nucleic acids encoding proteins involved in those activities. This type of coordination of gene expression regulation with cellular modifications to alter carbon flux and use of alternative carbon sources can greatly enhance target molecule production efficiency and economy.

In some embodiments of cell- and microbial-based target molecule production systems provided herein, such as those in which target molecule production involves cellular oxidative metabolism pathways, e.g, ω-oxidation and/or β-oxidation, for processing of fatty acids and/or alkanes,

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1400 of 3077

modifications made to the cells or microorganisms may include optimization of carbon source-dependent regulatory elements within the pathways. For example, although expression of unmodified genes encoding some of the polypeptides (e.g., enzymes) in these pathways is regulated by glucose-repressible and/or fatty acid (and/or alkane)-inducible transcriptional control elements, it may be beneficial to utilize heterologous stronger, more active, fatty acid (and/or alkane)-inducible transcriptional control elements to increase expression and/or activity of pathway polypeptides (e.g., enzymes) in modified cells and organisms. Thus, for example, the promoter of an endogenous glucose-repressible and/or fatty acid-inducible gene (e.g., an acyl-CoA oxidase (such as Pox5p), a peroxisomal protein (such as Pex11p), a cytochrome P450 monooxygenase or reductase (such as CYP52A17 or CPRB proteins)) can be replaced with a glucose-repressible and/or fatty acid-inducible promoter from another gene (e.g., an *HDE* gene) to enhance carbon processing through oxidative metabolism in modified cells or organisms.

In some embodiments of cell- and microbial-based target molecule production systems provided herein, carbon flux can be altered to reduce acetyl carbon processing in mitochondria during target molecule production occurring with fatty acid or alkane assimilation. In these embodiments, transcriptional control elements of some endogenous glucose-repressible and/or fatty acid-inducible genes encoding mitochondria-assoicated polypepties (e.g., mitochondrial transporter proteins (such as Crc1p) and carnitine acetyltransferases (such as Cat2p)) can be replaced with a promoter that is not fatty acid-inducible (e.g., a weaker and/or constitutive promoter) from another gene (e.g., glucose-6-phosphate isomerase gene) to reduce acetyl carbon uptake by and/or metabolism in mitochondria in modified cells or organisms. As also described herein, acetyl carbon uptake by and/or metabolism in mitochondria can be reduced in modified cells and organisms by replacing genes encoding one or more endogenous mitochondria-associated polypeptides (e.g., enzymes, such as carnitine acetyltransferase) with nucleic acid encoding a less active polypeptide. The nucleic acid encoding a less active polypeptide may also be linked to a transcriptional control element that provides for weak and/or not inducible expression of the polypeptide. If, however, the activity of the less-active polypeptide is insufficient for optimal cell functioning, a more active, inducible (e.g., fatty acid-inducible) promoter can be used to regulate expression of the less-active polypeptide.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1401 of 3077

Aliphatic and hydrophobic carbon sources

The ability of cells and microorganisms to utilize alternative carbon sources for energy generation and growth is based in the multiple enzyme-mediated metabolic pathways and gene regulation systems in microbial cells. In general, glucose often is a preferred carbon and energy source for many cells and microorganisms, e.g., yeast. A number of genes encoding products, such as certain enzymes, involved in cellular pathways and processes that are not used in carbohydrate metabolism may be repressed when glucose is present in culture media. If glucose is depleted in the media, some of these genes may then be derepressed. If alternative carbon sources, e.g., non-fermentable carbon sources, are available, some of the genes may be induced, and may be induced by several-fold. For example, when aliphatic or hydrophobic carbon sources, (e.g., alkanes, alkenes, fatty acids) are the external carbon source, the expression of genes encoding enzymes involved in β-oxidation and proteins involved in peroxisome proliferation can be induced. One example of a gene that is repressed in glucose media, derepressed in non-fermentable carbon source-containing media and induced in fatty acid-containing media is the gene encoding the peroxisomal trifunctional enzyme hydratase-dehydogenase-epimerase (HDE) in yeast, such as, for example, *Candida* (see, e.g., Sloots et al. (1991) *Gene* 105:129-134). The upstream regulatory regions of the HDE gene include a glucose-responsive region controlling glucose repression, a non-fermentable carbon-responsive region controlling derepression and an oleic acid-responsive region controlling fatty acid induction of transcription of the gene.

In some embodiments of the methods for producing a target molecule provided herein, a carbon source used during the production phase of a culturing step in the method includes an aliphatic or hydrophobic carbon source. In particular embodiments, an aliphatic or hydrophobic carbon source is the primary carbon source or may be the sole, or only, carbon source used during the production phase of a culturing step in the method. In some embodiments, the carbon source is a fatty acid and/or alkane. In certain aspects, the carbon source is a fatty acid. In some embodiments, the carbon source is an 18-carbon fatty acid such as, for example, oleic acid (C18:1), linoleic acid (C18:2) or linolenic acid (C18:3). Embodiments of the methods in which an aliphatic or hydrophobic carbon source is used are particularly advantageous when target molecule production involves oxidation pathways such as ω-oxidation and/or β-oxidation pathways and/or involves peroxisomes.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1402 of 3077

Additional advantages of using aliphatic or hydrophobic carbon sources in some embodiments of the methods of producing target molecules as provided herein are reduced costs and positive environmental effects realized through their use.

Methods of modifying cellular carbon flux

Included in the cells, organisms and microorganisms and methods described herein are those that provide for enhanced production of desired target molecules. In one aspect, production is enhanced through modification of carbon flux in cell or microbial production systems. Through modification of cellular carbon flux, carbon atoms that may have flowed or been transported into other metabolic processes (e.g., energy and/or cellular composition generation) in the cell are redirected for use in a target molecule production process. Described herein are multiple cellular modifications that can be employed to beneficially alter carbon flux. A modification can be used alone or in combination with one or more other modifications depending on the target molecule produced and the carbon flux alteration that is best suited to maximize its production.

Modification of acetyl-carnitine entry into mitochondria

Included in embodiments of the cells, microorganisms, compositions and methods provided herein are cell and microbial production platform systems and components thereof in which the amount of (a) acetyl-carnitine in the cell cytosol and/or (b) carnitine acetyltransferase and/or carnitine acetyltransferase activity in the cell cytosol is/are modified. Carnitine and carnitine acetyltransferase are the primary elements of the carnitine shuttle system in which acetyl carbons from acetyl-CoA are transferred across intracellular membranes and transported throughout eukaryotic cells. For example, the carnitine shuttle is a mechanism through which acetyl carbons from acetyl-CoA generated in peroxisomes flow to mitochondria. Acetyl groups from peroxisomal acetyl-CoA can be transferred to carnitine in a reaction catalyzed by carnitine acetyltransferase and then move across the peroxisomal membrane and into the cytosol in the form of acetyl-carnitine. Cytosolic acetyl-carnitine can then be transported into mitochondria where mitochondrial carnitine acetyltransferase catalyzes the reverse reaction to transfer the acetyl moiety from carnitine to free coenzyme A to generate acetyl-CoA. The carnitine shuttle provides a main “highway” for the flow of carbon atoms into mitochondria, particularly in the assimilation of fatty acids and other non-carbohydrate and non-fermentable carbon source

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1403 of 3077

molecules by cells. It is thus one target for modification of carbon flux in cells and microorganisms provided herein.

In some embodiments of the cells, microorganisms, compositions and methods provided herein, the amount of (a) acetyl-carnitine in the cell cytosol is increased and/or decreased, and/or (b) carnitine acetyltransferase and/or carnitine acetyltransferase activity in the cell cytosol is/are increased and/or decreased. For example, in some aspects, a cell or microorganism may be modified to increase cytosolic acetyl-carnitine, may be modified to decrease cytosolic acetyl-carnitine or may be modified to alternately increase and decrease cytosolic acetyl-carnitine depending on the conditions in which the modified cell or microorganism is cultured.

In certain embodiments, a host cell or microorganism is modified to increase the amount of (a) acetyl-carnitine in the cell cytosol and/or (b) carnitine acetyltransferase and/or carnitine acetyltransferase activity in the cell cytosol. Increasing the amount and/or activity of cytosolic carnitine acetyltransferase provides for an increased conversion of acetyl-carnitine into acetyl-CoA in the cytosol. Increasing the amount of acetyl-carnitine in the cytosol provides an increased availability of substrate for cytosolic carnitine acetyltransferase to convert to acetyl-CoA. These modifications effectively result in an increase in the generation and amount of cytosolic acetyl-CoA which can then be used in the production of desired carbon-containing molecules.

Modification of a mitochondrial acetyl-carnitine transporter

In one embodiment, cells or organisms provided herein are modified to increase and/or decrease the amount of acetyl group carbons in the cytosol in the form of acetyl-carnitine in transit from the peroxisome and other areas to the mitochondria by increasing or reducing the entry of acetyl-carnitine into mitochondria from the cytosol. In one aspect, the amount and/or rate of acetyl-carnitine transfer into the mitochondria in a cell can be increased or reduced by increasing or decreasing the expression of an acetyl-carnitine translocase protein localized in mitochondrial inner membranes. This protein carrier transports acetyl-carnitine across the mitochondrial inner membrane and into the mitochondrial matrix in exchange for carnitine. For example, a mitochondrial carnitine translocase is encoded by a *CRC1* gene in *Saccharomyces* (see, e.g, Palmieri *et al*. (1999) *FEBS Lett* 462:472-276), which contains an oleate-responsive element in the promoter region, and by an *AcuH* gene in *Asperillgus* (see, e.g., De Lucas *et al*.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1404 of 3077

(2001) *FEMS Microbiol Lett* 201:193-198). Carnitine carrier proteins belong to a family of mitochondrial carrier proteins which generally contain three tandemly repeated ~100-amino acid domains. Each of the three domains typically contains two hydrophobic regions, spanning the membrane as α-helices, linked by a hydrophobic loop that extends into the mitochondrial matrix. Each domain also typically contains a version of a motif (PX[DE]XX[RK]XRK) involved in forming a salt bridge that closes off the matrix side of a channel generated by the α-helices (see, e.g., Indiveri et al. (2011) *Molecular Aspects of Medicine* 32:223-233).

There are a number of ways to increase or reduce expression of an acetyl-carnitine translocase in a cell. For example, a host acetyl-carnitine translocase activity can be decreased by disruption (e.g., knockout, insertion mutagenesis, the like and combinations thereof) of a host gene encoding the protein, or by decreasing the activity of the promoter (e.g., through addition of repressor sequences to the promoter or 5'UTR or replacing the promoter) that controls transcription of an acetyl-carnitine translocase gene using recombinant molecular biology techniques known in the art and/or described herein. One method for disrupting an endogenous acetyl-carnitine translocase gene is by recombinantly inserting a heterologous nucleic acid (e.g., a nucleotide sequence encoding a selectable marker such as an enzyme that restores an auxotrophic host cell or organism to prototrophy) into the endogenous gene, thereby generating an engineered cell or organism deficient in acetyl-carnitine translocase activity. This can be done, for example, through homologous recombination in which a heterologous nucleic acid containing sequences of the endogenous acetyl-carnitine translocase gene and a disrupting sequence (e.g., a knock-out gene cassette such as described herein) is introduced into a host cell or microorganism. Nucleic acids encoding an acetyl-carnitine translocase can be obtained from a number of sources, including, for example, yeast cells. For example, genomic DNA from cell sources can be amplified using oligonucleotide primers based on the nucleotide sequence of an acetyl-carnitine translocase-encoding gene. Provided herein, for example, are a nucleotide sequence (SEQ ID NO: 71) that encodes a *Candida viswanathii* acetyl-carnitine translocase and the corresponding amino acid sequence (SEQ ID NO: 14). Nucleotide sequences encoding additional acetyl-carnitine translocase proteins include, but are not limited to: *Saccharomyces cerevisiae CRC1* (Genbank accession number AJ250124) and *Aspergillus nidulans AcuH* (Genbank accession number AJ011563).

Presence, absence or amount of an acetyl-carnitine translocase activity can be detected by any suitable method known in the art and/or described herein. For example, detection can be

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1405 of 3077

performed using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cells or organisms exhibit increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism. Methods of evaluating the activity of an acetyl-carnitine translocase include, for example, measuring carnitine uptake into and/or efflux from liposomes reconstituted with acetyl-carnitine translocase protein purified from microbial cells expressing the protein (see, e.g, , Palmieri *et al*. (1999) *FEBS Lett* 462:472-276).

In another example, a host acetyl-carnitine translocase activity can be increased, for example, by increasing the number of copies of an acetyl-carnitine translocase gene (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the gene), by increasing the activity of a promoter that regulates transcription of an acetyl-carnitine translocase gene, or by increasing the number of copies of an acetyl-carnitine translocase gene and increasing the activity of a promoter that regulates transcription of an acetyl-carnitine translocase gene. In some embodiments, an acetyl-carnitine translocase is endogenous to the host microorganism. Acetyl-carnitine translocase activities can also be increased, for example, by using an inducible promoter, e.g, a glucose- or fatty acid-inducible promoter for regulating transcription of an acetyl-carnitine translocase-encoding nucleic acid, and culturing the recombinant cell or microorganism in media containing a transcription-inducing carbon source.

Thus, in another example of modifying the expression of an acetyl-carnitine translocase in a cell, the promoter used for expression of nucleic acid encoding a mitochondrial acetyl-carnitine transport protein can be modified relative to an endogenous promoter encoding a transport protein. A promoter that is weaker, stronger and/or differently regulated than any endogenous mitochondrial acetyl-carnitine translocase gene promoter will provide for modified expression levels of the translocase protein. To achieve such modified expression, an endogenous promoter of a gene encoding a mitochondrial inner membrane acetyl-carnitine translocase can, in effect, be replaced with another promoter. This can be accomplished, for example, by introducing into a cell or microorganism a heterologous nucleic acid construct that includes a translocase-encoding sequence of nucleotides operably linked to a promoter that provides modified transcription or expression in the cell or microorganism relative to the endogenous promoter. The cell or microorganism can be one in which the endogenous gene(s) encoding a mitochondrial acetyl-carnitine translocase has been disrupted or deleted. For example, a host

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1406 of 3077

organism could be a yeast, e.g., a *Candida* yeast, in which the endogenous promoter of the mitochondrial acetyl-carnitine translocase includes an oleate-responsive element allowing for fatty acid induction. An example of a weaker promoter that would not be fatty acid inducible and provide for decreased acetyl-carnitine translocase expression, particularly when exposed to fatty acids as a carbon source, could be a yeast glucose-6-phosphate isomerase gene promoter. Modifying a promoter in this way provides another method for decreasing the amount and/or activity of acetyl-carnitine translocase protein in a host. This method is particularly advantageous when decreasing or eliminating acetyl-carnitine translocase activity through gene disruption is detrimental to cell growth and/or viability.

Modification of mitochondrial carnitine acetyltransferase activity

Acetyl-carnitine can be generated and degraded by the action of carnitine acetyltransferases (e.g., EC 2.3.1.7). In another embodiment provided herein, the amount of acetyl group carbons in the cytosol in the form of acetyl-carnitine in transit from the peroxisome to the mitochondria can be modified through altering the amount and/or activity of mitochondrial carnitine acetyltransferase. For example, by decreasing the activity level of mitochondrial carnitine acetyltransferase, there can be a corresponding decrease in conversion of acetyl-carnitine to acetyl-CoA in the mitochondria. This can introduce a bottleneck in acetyl-carnitine processing in the mitochondria which can have the effect of diverting acetyl-carnitine from entering the mitochondria from the cytoplasm. Alternatively, by increasing the activity level of mitochondrial carnitine acetyltransferase, there can be a corresponding increase in conversion of acetyl-carnitine to acetyl-CoA in the mitochondria which can augment acetyl-carnitine processing and avoid slowing of mitochondrial entry of acetyl-carnitine due to bottlenecks that might occur in the presence of increased amounts of acetyl-carnitine in the cytoplasm.

In some organisms, such as, for example, certain yeast species, carnitine acetyltransferase is dually targeted to mitochondria and peroxisomes by N-terminal and C-terminal targeting signals, respectively (see, e.g., Elgersma *et al.* (1995) *EMBO J.* 14: 3472-3479 and Kawachi *et al.* (1996) *Eur. J. Biochem.* 238: 845-852). An N-terminal sequence is referred to as the mitochondrial targeting signal (mts) and a C-terminal sequence is referred to as the peroxisomal targeting sequence (pts). An example of such an enzyme is the carnitine O-acetyltransferase enzyme encoded by some yeast *CAT2* genes.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1407 of 3077

Modifying carnitine acetyltransferase activity in mitochondria can be accomplished by modifying the amount of mitochondrial carnitine acetyltransferase protein expression in a cell, for example, by replacing the wild-type promoter of an endogenous gene in a cell or organism with a weaker or stronger heterologous promoter, and/or replacing or modifying a gene encoding a wild-type carnitine acetyltransferase such that the encoded modified or substituted carnitine acetyltransferase protein has a reduced or increased enzyme activity. For example, a host carnitine acetyltransferase activity can be decreased by disruption (e.g., knockout, insertion mutagenesis, the like and combinations thereof) of a host gene encoding the protein, or by decreasing the activity of the promoter (e.g., through addition of repressor sequences to the promoter or 5'UTR or replacing the promoter) that controls transcription of a carnitine acetyltransferase gene using recombinant molecular biology techniques known in the art and/or described herein. In one embodiment, a diploid yeast, such as, for example, a *Candida* yeast, when used as a host microorganism can be subjected to genetic modification in which one of the two alleles of a mitochondrial carnitine acetyltransferase gene is disrupted or deleted. In so doing, a single allele of the gene remains for a reduced amount of carnitine acetyltransferase expression in the microorganism and a reduced amount of the protein in the mitochondria. This can effectively reduce and/or slow the amount of acetyl-carnitine that is processed into acetyl-CoA in the mitochondria without completely eliminating a minimal supply of acetyl carbons that may be required for cellular respiration that occurs in the mitochondria, yet provides for increased retention of acetyl-carnitine in the cytosol. In some instances, the amount of carnitine acetyltransferase activity remaining after disruption of a single allele of a mitochondrial carnitine acetyltransferase gene of a diploid cell or organism may be at a higher level than desired. In such cases, both alleles may be disrupted. A heterologous nucleic acid encoding a carnitine acetyltransferase that is less active than an endogenous carnitine acetyltransferase (or nucleic acid encoding a carnitine acetyltransferase that is linked to a weak promoter) can be introduced into host cells or organisms in which all alleles of the endogenous gene have been disrupted.

One method for disrupting an endogenous carnitine acetyltransferase gene is by recombinantly inserting a heterologous nucleic acid (e.g., a nucleotide sequence encoding a selectable marker such as an enzyme that restores an auxotrophic host cell or organism to prototrophy) into the endogenous gene, thereby generating an engineered cell or organism deficient in carnitine acetyltransferase activity. This can be done, for example, through homologous recombination in which a heterologous nucleic acid containing sequences of the endogenous carnitine acetyltransferase gene and a disrupting sequence (e.g., a knock-out gene cassette such as

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1408 of 3077

described herein) is introduced into a host cell or microorganism. Nucleic acids encoding a carnitine acetyltransferase can be obtained from a number of sources, including, for example, yeast cells. For example, genomic DNA from cell sources can be amplified using oligonucleotide primers based on the nucleotide sequence of a carnitine acetyltransferase-encoding gene. Provided herein, for example, are a nucleotide sequence (SEQ ID NO: 59) that encodes a *Candida viswanathii* carnitine acetyltransferase (*CAT2* gene) and the corresponding amino acid sequence (SEQ ID NO: 2). Nucleotide sequences encoding additional carnitine acetyltransferase proteins include, but are not limited to: *Saccharomyces cerevisiae CAT2* (Genbank accession numbers Z14021, NM_001182400), *Candida tropicalis CAT2* (Genbank accession number D84549), *Candida glabrata CAT2* (Genbank accession number AF2811), *Candida albicans CAT2* (Genbank accession numbers AF525684), *Aspergillus nidulans AcuJ* (Genbank accession number XM_658791) and *Cyberlindnera jadinii* (Genbank accession number AB641826).

In cells or organisms in which a mitochondrial carnitine acetyltransferase is encoded by a gene that generates a protein containing mitochondrial and peroxisomal targeting sequences, it may be desired to modify only the mitochondrial enzyme, and continue expression of the peroxisomal enzyme. In this instance, an endogenous mitochondrial/peroxisomal carnitine acetyltransferase gene can be disrupted or deleted and heterologous nucleic acids separately encoding a mitochondrial-targeted enzyme and a peroxisomal-targeted enzyme can be introduced into the cell or microorganism. For example, a peroxisomal-targeted enzyme that would not be expressed in the mitochondria can be produced in a cell or microorganism by introducing a heterologous nucleic acid that encodes a carnitine acetyltransferase that includes a peroxisomal targeting sequence of amino acids but lacks a mitochondrial targeting sequence of amino acids. An example of such a modified *Candida viswanathii* nucleic acid sequence (*CAT2*$^{\Delta mts}$; SEQ ID NO: 60), and the amino acid sequence encoded thereby (Cat2p$^{\Delta mts}$; SEQ ID NO: 3), are provided herein. A mitochondrial-targeted carnitine acetyltransferase that would not be expressed in peroxisomes can be produced in a cell or microorganism by introducing a heterologous nucleic acid that encodes a carnitine acetyltransferase that includes a mitochondrial targeting sequence of amino acids but lacks a peroxisomal targeting sequence of amino acids. An example of such a modified *Candida viswanathii* nucleic acid sequence (*CAT2*$^{\Delta pts}$; SEQ ID NO: 62), and the amino acid sequence encoded thereby (Cat2p$^{\Delta pts}$; SEQ ID NO: 5), are provided herein. A heterologous nucleic acid encoding a mitochondrial-targeted carnitine acetyltransferase that would not be expressed in peroxisomes can also include

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1409 of 3077

modifications that alter its expression and or activity in the mitochondria as described herein. For example, regulatory sequences of nucleic acids (e.g, promoter sequences, repressor sequences) can be included that provide for decreased or increased expression of the enzyme and/or an altered pattern of expression of the enzyme. A heterologous nucleic acid encoding a mitochondrial-targeted carnitine acetyltransferase can include modifications that alter its activity, e.g., providing for more active or less active enzymatic activity relative to an endogenous mitochondrial carnitine acetyltransferase. The carnitine acetyltransferase activities of host and modified cells and microorganisms can be evaluated and monitored using methods known in the art. For example, methods of isolating peroxisomal and mitochondrial components of yeast cells and of extracting carnitine acetyltransferase from subcellular fractions have been described by Ueda et al. [(1982) *Eur. J. Biochem.*124:205-210] and Kozulic et al. [(1987) *Eur. J. Biochem.*168:245-250]. Methods of measuring the enzymatic activity of carnitine acetyltransferase are also known in the art, see, e.g., Fritz and Schultz (1965) *J. Biol. Chem.* 240:2188-2192; Chase (1969) *Meth. Enzymol.*13:387-393.

In one embodiment provided herein, a heterologous nucleic acid encoding a yeast cytoplasmic carnitine acetyltransferase that has a reduced carnitine acetyltransferase activity relative to the activity of the enzyme encoded by a host microorganism's endogenous mitochondrial carnitine acetyltransferase gene can be introduced into a microbial host in which the endogenous mitochondrial carnitine acetyltransferase gene(s) has been disrupted or deleted. The heterologous nucleic acid encoding the less active carnitine acetyltransferase can be modified to include nucleotides encoding a mitochondrial targeting sequence for expression of the enzyme in the mitochondria. For example, in one aspect, a heterologous nucleic acid encoding a *Candida viswanathii* cytosolic carnitine acetyltransferase (*YAT1*) with added nucleotides encoding a mitochondrial targeting sequence (*YAT1*$^{+mts}$) can be introduced into a host cell or microorganism (e.g., a *Candida viswanathii* cell). Any sequence encoding a mitochondrial targeting from a protein that is localized to mitochondria can be used in generating the heterologous nucleic acid. Examples include, but are not limited to, nucleotides encoding mitochondrial targeting sequences from mitochondrial cytochrome oxidase subunit IV (Cox4p), mitochondrial citrate synthase (Cit1p) and mitochondrial carnitine acetyltransferase (Cat2p) proteins. Nucleotide sequences encoding (and the amino acid sequences of) *Candida viswanathii* Yat1p (amino acid SEQ ID NO: 6 and nucleotide SEQ ID NO: 63), YAT1^{+mts}p (amino acid SEQ ID NOS: 10, 11 and 12 and nucleotide SEQ ID NOS: 67, 68 and 69), and the mitochondrial targeting sequences of Cox4p (amino acid SEQ ID NO: 7 and nucleotide SEQ ID

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1410 of 3077

NO: 64), Cit1p (amino acid SEQ ID NO: 8 and nucleotide SEQ ID NO: 65) and Cat2p (amino acid SEQ ID NO: 9 and nucleotide SEQ ID NO: 66) are provided herein. Additional non-limiting examples of nucleic acids encoding cytoplasmic carnitine acetyltransferase include: *Saccharomyces cerevisiae YAT1* (Genbank accession number X74553), *Aspergillus nidulans FacC* (Genbank accession number AF023156), *Cyberlindnera jadinii YAT1* (Genbank accession number AB641829), *Candida dubliniensis YAT1* (Genbank accession number XM_002416790) and *Candida albicans* (Genbank accession number AF525683). Additional non-limiting examples of nucleic acids encoding mitochondrial targeting sequences include *Saccharomyces cerevisiae Cit1* (nucleotides in Genbank accession number NM_001183178 encoding N-terminal 37 amino acids) and *Saccharomyces cerevisiae Cox4* (nucleotides in Genbank accession number NM_001181052 encoding N-terminal 25 amino acids).

In another example, the amount and/or activity of carnitine acetyltransferase in a cell or microorganism can be increased, for example, by increasing the number of copies of a carnitine acetyltransferase gene (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the gene), by increasing the activity of a promoter that regulates transcription of a carnitine acetyltransferase gene, or by increasing the number of copies of a carnitine acetyltransferase gene and increasing the activity of a promoter that regulates transcription of a carnitine acetyltransferase gene. In some embodiments, a carnitine acetyltransferase is endogenous to the host cell or microorganism. Carnitine acetyltransferase activities can also be increased, for example, by using an inducible promoter, e.g, a glucose- or fatty acid-inducible promoter for regulating transcription of a carnitine acetyltransferase-encoding nucleic acid, and culturing the recombinant cell or microorganism in media containing a transcription-inducing carbon source.

Thus, in another example of modifying the expression of a carnitine acetyltransferase in a cell, the promoter used for expression of nucleic acid encoding a carnitine acetyltransferase protein can be modified relative to an endogenous promoter encoding a carnitine acetyltransferase protein. A promoter that is weaker, stronger and/or differently regulated than any endogenous carnitine acetyltransferase gene promoter will provide for modified expression levels of the protein. To achieve such modified expression, an endogenous promoter of a gene encoding a carnitine acetyltransferase can, in effect, be replaced with another promoter. This can be accomplished, for example, by introducing into a cell or microorganism a heterologous nucleic acid construct that includes a carnitine acetyltransferase-encoding sequence of nucleotides operably linked to a promoter that provides modified expression in the cell or microorganism

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1411 of 3077

relative to the endogenous promoter. The cell or microorganism can be one in which the endogenous gene(s) encoding a carnitine acetyltransferase has been disrupted or deleted. For example, a host organism could be a yeast, e.g., a *Candida* yeast, in which the endogenous promoter includes an oleic acid-responsive element allowing for fatty acid induction. An example of a weaker promoter that would not be fatty acid inducible and provide for decreased carnitine acetyltransferase expression, particularly when exposed to fatty acids as a carbon source, could be a yeast glucose-6-phosphate isomerase gene promoter.

Different combinations of transcription regulatory elements (e.g., promoters) and enzymes can be utilized to achieve an optimal level of activity of carnitine acetyltransferase (or other enzyme being modified) in a cell or microorganism modified to alter carbon flux therein. For example, in embodiments in which a decreased level, but not an elimination, of an activity, such as mitochondrial carnitine acetyltransferase, in a cell or organism is desired, an optimal activity level may be achieved by using a strong and/or inducible promoter to express nucleic acid encoding a protein having a decreased activity. In one embodiment described herein, the mitochondrial carnitine acetyltransferase activity of a host organism (e.g., *Candida* yeast) is decreased by disrupting both alleles of the endogenous gene encoding mitochondrial carnitine acetyltransferase and introducing heterologous nucleic acid encoding a mitochondrial-targeted carnitine acetyltransferase having a lower enzyme activity than the endogenous mitochondrial carnitine acetyltransferase. To ensure that the level of enzyme activity provided by the less active mitochondrial-targeted carnitine acetyltransferase is sufficient and optimal in the modified cell, a strong, fatty acid-inducible promoter (e.g., an *HDE* gene promoter) can be linked to the nucleic acid encoding the less active enzyme to regulate transcription and production of a desired amount of the enzyme.

Modification of cytosolic carnitine acetyltransferase activity

Included in embodiments of the cells, microorganisms, compositions and methods provided herein are microbial production platform systems and components thereof in which the amount of carnitine acetyltransferase in the cell cytosol and/or carnitine acetyltransferase activity in the cell cytosol is modified. In some instances, the amount of carnitine acetyltransferase in the cell cytosol is increased and/or decreased, and/or carnitine acetyltransferase activity in the cell cytosol is increased and/or decreased. For example, in some aspects, a cell or microorganism may be modified to increase cytosolic carnitine acetyltransferase and/or carnitine

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1412 of 3077

acetyltransferase activity, may be modified to decrease cytosolic carnitine acetyltransferase and/or carnitine acetyltransferase activity, or may be modified to alternately increase and decrease cytosolic carnitine acetyltransferase and/or carnitine acetyltransferase activity depending on the conditions in which the modified cell or microorganism is cultured.

In some embodiments, the capture of carbon atoms in the acetyl group of acetyl-CoA generated from metabolic processes such as peroxisomal β-oxidation can be accomplished by increasing the amount of carnitine acetyltransferase protein and/or activity in the cell cytosol of a cell or microorganism. In so doing, there is an increased conversion of acetyl-carnitine, such as that which is in transit from the peroxisome to the mitochondria, into acetyl-CoA in the cytoplasm. In one aspect, the amount and/or activity of a host cytosolic carnitine acetyltransferase can be increased, for example, by increasing the number of copies of a gene encoding a cytoplasmic carnitine acetyltransferase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the gene), by increasing the activity of a promoter that regulates transcription of a gene encoding a cytoplasmic carnitine acetyltransferase, or by increasing the number of copies of a gene encoding a cytoplasmic carnitine acetyltransferase and increasing the activity of a promoter that regulates transcription of a gene encoding a cytoplasmic carnitine acetyltransferase. In some embodiments, a cytoplasmic carnitine acetyltransferase is endogenous to the host cell or microorganism.

In one embodiment of the cell and microbial systems and methods provided herein, the amount of carnitine acetyltransferase protein expressed in the cytosol can be increased by introducing heterologous nucleic acid encoding a cytoplasmic carnitine acetyltransferase into a cell or microorganism. In some cells and microorganisms, e.g., some yeast strains, cytoplasmic carnitine acetyltransferase is encoded by a gene that is distinct from the gene(s) encoding mitochondrial and/or peroxisomal carnitine acetyltransferase. For example, in some yeast strains, a cytoplasmic carnitine acetyltransferase is encoded by a *YAT* gene, whereas a mitochondrial and/or peroxisomal carnitine acetyltransferase is encoded by a *CAT* gene. Nucleotide sequences encoding (and the amino acid sequences of) *Candida viswanathii* Yat1p (amino acid SEQ ID NO: 6 and nucleotide SEQ ID NO: 63) are provided herein. Additional non-limiting examples of nucleic acids encoding cytoplasmic carnitine acetyltransferase include *Saccharomyces cerevisiae YAT1* (Genbank accession number X74553), *Aspergillus nidulans FacC* (Genbank accession number AF023156), *Cyberlindnera jadinii YAT1* (Genbank accession

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1413 of 3077

number AB641829), *Candida dubliniensis YAT1* (Genbank accession number XM_002416790) and *Candida albicans* (Genbank accession number AF525683).

In another embodiment, the amount of cytoplasmic carnitine acetyltransferase can be modified by modifying the activity of a promoter that regulates transcription of a gene encoding a cytoplasmic carnitine acetyltransferase. Thus, in another example of modifying the expression of a cytosolic carnitine acetyltransferase in a cell, the promoter used for expression of nucleic acid encoding a cytosolic carnitine acetyltransferase protein can be modified relative to an endogenous promoter encoding a cytosolic carnitine acetyltransferase protein. A promoter that is weaker, stronger and/or differently regulated than any endogenous cytosolic carnitine acetyltransferase gene promoter will provide for modified expression levels of the protein. To achieve such modified expression, an endogenous promoter of a gene encoding a cytosolic carnitine acetyltransferase can, in effect, be replaced with another promoter. This can be accomplished, for example, by introducing into a cell or microorganism a heterologous nucleic acid construct that includes a cytosolic carnitine acetyltransferase-encoding sequence of nucleotides operably linked to a promoter that provides modified expression in the cell or microorganism relative to the endogenous promoter. The cell or microorganism can be one in which the endogenous gene(s) encoding a cytosolic carnitine acetyltransferase has been disrupted or deleted. For example, a host organism could be a yeast, e.g., a *Candida* yeast, in which the endogenous promoter does not include an oleic acid-responsive element allowing for fatty acid induction. An example of a stronger promoter that would be fatty acid inducible and provide for increased cytosolic carnitine acetyltransferase expression, particularly when exposed to fatty acids as a carbon source, is a peroxisomal protein gene and/or β-oxidation enzyme gene promoter, e.g., a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene promoter. The nucleotide sequence of a *Candida viswanathii* HDE gene promoter (SEQ ID NO: 113) is provided herein as are examples of additional fatty acid-inducible promoters.

In a further embodiment, cytosolic carnitine acetyltransferase activity can be modified by introducing into a cell or microorganism a heterologous nucleic acid encoding a carnitine acetyltransferase that is more active or less active than an endogenous cytosolic carnitine acetyltransferase. For example, a heterologous nucleic acid encoding an enzyme that has an increased carnitine acetyltransferase activity relative to the activity of a cytosolic carnitine acetyltransferase expressed in the host cell or microorganism can be introduced into a host to provide for increased generation of cytosolic acetyl-CoA from acetyl-carnitine. The host can be

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1414 of 3077

one in which the endogenous cytosolic carnitine acetyltransferase gene(s) has been disrupted or deleted. The heterologous nucleic acid encoding the more active carnitine acetyltransferase can, if required, be altered to exclude any nucleotides encoding a cell localization (e.g., mitochondria, peroxisomes) targeting sequence in order to provide for expression of the enzyme in the cytosol.

Thus, for example, engineered carnitine O-acetyltransferase proteins lacking amino acid sequence targeting signals that direct the enzyme to one or more cellular locations other than the cytoplasm can be expressed in host cells thereby increasing the amount of carnitine O-acetyltransferase in the cytoplasm. Such engineered proteins will remain in the cytoplasm after being produced by the cell or organism. One such modified carnitine O-acetyltransferase protein is a yeast Cat2p$^{\Delta mts\Delta pts}$ lacking a mitochondrial targeting signal (mts) and a perioxisomal targeting signal (pts). In some instances, a mitochondrial and/or peroxisomal carnitine acetyltransferase (e.g., a yeast Cat2p) may be more active than an endogenous cytosolic carnitine acetyltransferase (e.g., a yeast Yat1p). A more active Cat2p enzyme can be expressed in the cytosol of a host upon introduction of heterologous nucleic acid (e.g., *CAT2*$^{\Delta mts\Delta pts}$) encoding the more active enzyme lacking mitochondrial- and peroxisomal-targeting sequences. In a particular embodiment, the Cat2p enzyme can be a *Candida* yeast protein. An example of a *Candida viswanathii* nucleotide sequence (*CAT2*$^{\Delta mts\Delta pts}$; SEQ ID NO: 61) encoding a carnitine acetyltransferase lacking mitochondrial- and peroxisomal-targeting sequences (Cat2p$^{\Delta mts\Delta pts}$; SEQ ID NO: 4) is provided herein.

Additional examples of nucleotide sequences encoding carnitine acetyltransferase proteins include: *Saccharomyces cerevisiae CAT2* (Genbank accession numbers Z14021, NM_001182400), *Candida tropicalis CAT2* (Genbank accession number D84549), *Candida glabrata CAT2* (Genbank accession number AF2811), *Candida albicans CAT2* (Genbank accession numbers AF525684), *Aspergillus nidulans AcuJ* (Genbank accession number XM_658791), *Neurospora crassa* (Genbank accession number XM_957579) and *Cyberlindnera jadinii CAT2* (Genbank accession number AB641826). Any of these, and other such carnitine acetyltransferase-encoding nucleic acids, can be analyzed for the presence of 5' and 3' ORF nucleotides encoding possible mitochondrial- or peroxisomal-targeting sequences of amino acids and modified to eliminate such sequences. For example, the initial approximately 66 base pairs of the *Saccharomyces cerevisiae CAT2* or the *Candida tropicalis CAT2* coding sequence can be excluded to eliminate mitochondrial targeting of the enzyme, while deletion of the

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1415 of 3077

terminal 9 base pairs of the coding sequence that encode a 3-amino acid persoxisomal targeting sequence, i.e., PTS1, (AKL or PKL motif) eliminates peroxisomal targeting of the enzyme (see, e.g., Elgersma *et al.* (1995) *EMBO J.* 14: 3472-3479 and Kawachi et al. (1996) *J. Biochem.* 120:731-735). Similarly, the initial approximately 120 base pairs of the *Aspergillus nidulans AcuJ* coding sequence can be excluded to possibly eliminate mitochondrial targeting of the enzyme, while deletion of the terminal 9 base pairs of the coding sequence that encode a 3-amino acid PTS1 (AKL motif) may eliminate peroxisomal targeting of the enzyme (see, e.g., Hynes et al. (2011) *Eukarot. Cell* 10:547-555). In general, yeast mitochondrial targeting sequences occur within the initial 10-90 N-terminal amino acid residues of a mitochondrial protein, have a significant arginine composition and very few to no negatively charged residues. Prediction tools, e.g., MitoProt, TargetP, Predotar and TPpred2, can be used in evaluating an amino acid sequence for identification of possible mitochondrial targeting sequences (see, e.g., Claros (1995) *Comput. Appl. Sci.* 11:441-447; Emanuelsson et al. (2000) *J. Mol. Biol*. 300:1005-1016; Small et al. (2004) *Proteomics* 4:1581-1590; Savojardo et al. (2014) *Bioinformatics* 30:2973-2974). Yeast peroxisomal targeting sequences generally occur at the C-terminus of a peroxisomal protein. Generally, the 3-amino acid consensus sequence of a yeast PTS1 has an initial amino acid containing a small, uncharged side chain (e.g., serine, alanine and cysteine), followed by a positively charged residue (e.g., lysine, arginine and histidine) and ending with a leucine residue; however, variants (e.g., PKL and others) of the consensus sequence do occur. Another example of a peroxisomal targeting signal sequence is the tripeptide AKI of the *Candida tropicalis* trifunctional enzyme hydratase-dehydrogenase-epimerase (HDE) (see, e.g., Aitchison et al. (1991) *J. Biol. Chem.* 266(34):23197-23203. Prediction tools, e.g., PTSI Predictor (mendel.imp.ac.at/mendeljsp/sat/pts1/PTS1predictor.jsp), can be used in evaluating an amino acid sequence for identification of possible peroxisomal targeting signal sequences (see, e.g., Brocard and Hartig (2006) *Biochim. Biophys. ACTA* 1763:1565-1573).

The promoter used for regulating transcription of a heterologous nucleic acid encoding a carnitine acetyltransferase that is more active or less active than an endogenous cytosolic carnitine acetyltransferase can also be modified. For example, the amount of a more active carnitine acetyltransferase protein expressed in the cytosol may be increased by including in the heterologous nucleic acid a stronger heterologous promoter and/or a promoter that provides for a different pattern of expression in the cell or microorganism.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1416 of 3077

Alternatively, decreasing carnitine acetyltransferase activity in the cytosol can be accomplished by modifying the amount of cytosolic carnitine acetyltransferase protein expression in a cell, for example, by replacing the wild-type promoter of an endogenous cytosolic carnitine acetyltransferase gene in a cell or organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type cytosolic carnitine acetyltransferase such that the encoded modified or substituted carnitine acetyltransferase protein has a reduced enzyme activity. For example, expression of a host cytosolic carnitine acetyltransferase activity can be decreased by disruption (e.g., knockout, insertion mutagenesis, the like and combinations thereof) of a host gene encoding the protein, or by decreasing the activity of the promoter (e.g., through addition of repressor sequences to the promoter or 5'UTR or replacing the promoter) that controls transcription of a cytosolic carnitine acetyltransferase gene using recombinant molecular biology techniques known in the art and/or described herein. In one embodiment, a diploid yeast, such as, for example, a *Candida* yeast, when used as a host microorganism can be subjected to genetic modification in which one of the two alleles of a cytosolic carnitine acetyltransferase gene is disrupted or deleted. In so doing, a single allele of the gene remains for a reduced amount of carnitine acetyltransferase expression in the microorganism and a reduced amount of the protein in the cytosol. This effectively reduces and/or slows the amount of acetyl-carnitine that is processed into acetyl-CoA in the cytosol.

One method for disrupting an endogenous carnitine acetyltransferase gene is by recombinantly inserting a heterologous nucleic acid (e.g., a nucleotide sequence encoding a selectable marker such as an enzyme that restores an auxotrophic host organism to prototrophy) into the endogenous gene, thereby generating an engineered organism deficient in cytosolic carnitine acetyltransferase activity. This can be done, for example, through homologous recombination in which a heterologous nucleic acid containing sequences of the endogenous cytosolic carnitine acetyltransferase gene and a disrupting sequence (e.g., a knock out gene cassette as described herein) is introduced into a host cell or microorganism. Nucleic acids encoding a cytosolic carnitine acetyltransferase can be obtained from a number of sources, including, for example, yeast cells. Genomic DNA from cell sources can be amplified using oligonucleotide primers based on the nucleotide sequence of a cytosolic carnitine acetyltransferase-encoding gene. For example, in some yeast strains, a cytosolic carnitine acetyltransferase is encoded by a *YAT* gene. Nucleotide sequences encoding (and the amino acid sequences of) *Candida viswanathii* Yat1p (amino acid SEQ ID NO: 6 and nucleotide SEQ ID NO: 63) are provided herein.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1417 of 3077

Additional non-limiting examples of nucleic acids encoding cytoplasmic carnitine acetyltransferase include *Saccharomyces cerevisiae YAT1* (Genbank accession number X74553), *Aspergillus nidulans FacC* (Genbank accession number AF023156), *Cyberlindnera jadinii YAT1* (Genbank accession number AB641829), *Candida dubliniensis YAT1* (Genbank accession number XM_002416790) and *Candida albicans* (Genbank accession number AF525683).

Modification of acetyl-CoA generation through oxidative metabolism

Included in the biological production platform systems and components thereof provided herein are embodiments in which the generation of acetyl-CoA in a cell or organism is modified. In some embodiments, the cellular processing of fatty acids, such as those obtained from external carbon sources (e.g., non-fermentable carbon sources) and internal, cell-generated sources (including, for example, but not limited to, fatty acids generated by catabolism of alkanes, fatty alcohols and fatty aldehydes), can be directed toward acetyl-CoA-generating oxidative metabolism pathways in cells. In some embodiments, the processing of fatty acids can be directed toward oxidative metabolism (e.g., ω- and/or β-oxidation) and away from cellular pathways, such as lipid synthesis pathways, that may not be involved in target molecule production. Accordingly, provided herein are cells, organisms, compositions and methods in which cellular carbon flux has been modified through alterations in cellular oxidative metabolism and/or fatty acid activation. In particular embodiments, cellular carbon flux has been modified to increase the production of acetyl-CoA in a cell through altering oxidative metabolism and/or fatty acid activation. Carbon flux modifications involving oxidative metabolism are particularly useful in embodiments in which alternative, non-carbohydrate carbon sources (e.g., some non-fermentable carbon sources) are used as a feedstock for modified cells and organisms in target molecule production.

For example, some organisms (e.g., some species of *Candida*, *Yarrowia*, *Pichia*, *Debaryomyces*, *Acinetobacter, Bacillus, Mycobaterium, Pseudomonas, Sphingomonas, Alcanivorax and Rhodococcus*) are able to endogenously assimilate alkanes as a carbon source. A primary pathway for alkane assimilation (also referred to as monoterminal alkane oxidation), which occurs in association with the endoplasmic reticulum and peroxisomes in eukaryotes, is through the initial conversion of alkanes to fatty acids which can then be metabolized in the cells. The conversion can occur through a three-step process as follows: (1)

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1418 of 3077

terminal hydroxylation of alkane by a cytochrome P450-dependent monooxygenase system (e.g, *ALK* gene products of the CYP52 family as a terminal oxidase and an NADPH-dependent cytochrome P450 reductase (e.g., *CPR1*-encoded) for electron transfer) which yields a fatty alcohol; (2) conversion of the terminal hydroxy group of the alcohol to a fatty aldehyde in reactions involving fatty alcohol dehydrogenase (e.g., *ADH*) or fatty alcohol oxidase (e.g., *FAO*); and (3) conversion of the fatty aldehyde to a fatty acid by a fatty aldehyde dehydrogenase. The resulting fatty acid can then be subject to the same metabolic processing as is a fatty acid taken up directly by the organism.

Fatty acids can be metabolized in several ways depending on the type of cell or organism. Many fatty acid metabolic pathways, including β-oxidation, lipid biosynthesis, and protein acylation, require that a fatty acid be activated by thioesterification to coenzyme A (i.e., acyl-CoA), or to an acyl carrier protein (i.e., acyl-ACP), prior to being metabolized. As used herein, "activation" with reference to fatty acids refers to the thioesterification of a fatty acid with a carrier molecule such as coenzyme A (Co-A) or acyl carrier protein (ACP). A fatty acid that has undergone activation into an acyl-CoA or fatty acid-ACP molecule is referred to as an activated fatty acid. The thioesterification reaction can be catalyzed by acyl-CoA synthetase enzymes. There are multiple enzymes having acyl-CoA synthetase activity in cells which differ based on cellular localization (e.g., plasma membrane, cytosol, endoplasmic reticulum membrane, peroxisomes) and substrate (e.g., fatty acid carbon chain length) specificity. In general, plasma membrane-associated acyl-CoA synthetases often are more specific for very long chain fatty acids and are involved in transport of these hydrophobic molecules across the membrane concurrent with activation of the fatty acids to acyl-CoA. Once activated, the acyl-CoA can then be used in a number of metabolic pathways, only one of which is β-oxidation. Thus, fatty acids activated at the plasma membrane and/or in the cytosol can represent possible "losses" of carbon atoms to cellular synthesis pathways (e.g., lipid synthesis) at the expense of other target molecule production pathways (e.g., oxidative metabolism, malonyl-CoA production and organic acid, polyketide and/or terpene synthesis). Therefore, in some embodiments of cells and organisms provided herein for use in target molecule production systems, it can be beneficial to capture the carbon atoms of free fatty acids for use in target molecule production and decrease activated fatty acid (e.g., acyl-CoA) flow into cellular pathways not associated with target molecule production. As described herein, methods of enhancing fatty acid carbon flow through oxidative metabolism include, but are not limited to, modification of activities of cellular ω- and/or β-oxidation systems and acyl-CoA synthetase activities.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1419 of 3077

Modification of ω-oxidation

In the oxidative metabolism pathway referred to as ω-oxidation (or diterminal oxidation), fatty acids can be converted to dicarboxylic acids (diacids). Several enzyme activities (e.g., cytochrome P450 hydroxylase complex, fatty alcohol oxidase and fatty aldehyde dehydrogenase) can be involved in the process of ω-oxidation. The term "ω-oxidation pathway" as used herein, refers to a cellular metabolic pathway constituted by a series of enzymatic activities through which fatty acids and alkanes are converted to dicarboxylic acids. Some cells and microorganisms (e.g., species of yeast and bacteria) endogenously express the enzyme activities, and those that do not can be genetically modified to provide a heterologous ω-oxidation pathway by introducing nucleic acids encoding the required enzymes into cells and expressing the proteins therein. Free fatty acids internalized into cells, or generated within cells (e.g., by oxidation of internalized alkanes), can directly enter into and be processed in the ω-oxidation pathway without prior activation to acyl-CoA. In some embodiments of the cell-based production systems and methods provided herein, dicarboxylic acids can be a target molecule. In some embodiments, e.g., in which some shorter-chain dicarboxylic acids are a target molecule, or dicarboxylic acids are not a target molecule (or a co-target molecule along with one or more other desired products), dicarboxylic acids generated by ω-oxidation can be further oxidized through β-oxidation. Dicarboxylic acids can traverse peroxisomal membranes in eukaryotic cells and be metabolized to yield acetyl-CoA that can be used in target molecule generation. Because free fatty acids can be processed through ω-oxidation without being activated by thioesterification with Co-A, and diacids resulting from ω-oxidation of fatty acids readily move into peroxisomes, the ω-oxidation pathway can serve as a cellular gateway for funneling internalized fatty acids into oxidative metabolism and away from cytosolic activation that is required for use of fatty acids in other cellular pathways that may not be involved in target molecule production.

The term "ω-oxidation activity" refers to any of the activities in the ω-oxidation pathway utilized to metabolize alkanes and fatty acids. The activities that may be utilized in this metabolic pathway include, but are not limited to, monooxygenase activity (e.g., cytochrome P450 activity), monooxygenase reductase activity (e.g., cytochrome P450 reductase activity), alcohol dehydrogenase activity (e.g., fatty alcohol dehydrogenase activity or long-chain alcohol dehydrogenase activity), fatty alcohol oxidase activity and fatty aldehyde dehydrogenase

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1420 of 3077

activity. In some embodiments of the cells, organisms, compositions and methods provided herein, the ω-oxidation activity of a cell or organism is modified. In one embodiment, one or more of the activities in the ω-oxidation pathway can be modified. In particular embodiments, one or more of the activities in the ω-oxidation pathway can be increased.

Modification of a monooxygenase activity

The initial step in the ω-oxidation pathway is the conversion of a fatty acid to a corresponding fatty alcohol and involves NADPH and molecular oxygen. A cytochrome P450 enzyme (e.g., monooxygenase activity, EC 1.14.14.1) often catalyzes the insertion of one atom of oxygen bound to the heme group in cytochrome P450 into an organic substrate (RH) while the other oxygen atom is reduced to water. A cytochrome P450 reductase catalyzes the reductive splitting of the oxygen and transfer of electrons from NADPH to cytochrome P450. Insertion of the oxygen atom near the omega carbon of a substrate yields an alcohol derivative of the original starting substrate (e.g., yields a fatty alcohol). In some embodiments of the cells, organisms, compositions and methods provided herein, the amount and/or activity of a monooxygenase in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase the amount and/or activity of a monooxygenase, may be modified to decrease the amount and/or activity of a monooxygenase, or may be modified to alternately increase and decrease the amount and/or activity of a monooxygenase depending, for example, on the cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

In certain aspects, the amount and/or activity of a monooxygenase in a cell is increased. Increasing the amount and/or activity of a monooxygenase may be particularly beneficial in embodiments in which the flux of carbons from fatty acids is directed toward a particular target product pathway involving oxidative metabolism and away from other cellular metabolic pathways not involved in target molecule production.

In certain embodiments, the monooxygenase activity is unchanged in a host or engineered cell or organism. In one embodiment, the amount and/or activity of a host monooxygenase can be increased, for example, by increasing the number of copies of a nucleic acid encoding a monooxygenase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the nucleic acid), by increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1421 of 3077

monooxygenase, or by increasing the number of copies of a nucleic acid encoding a monooxygenase and increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a monooxygenase. In some embodiments, a monooxygenase is endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, the amount of monooxygenase protein expressed in a cell can be increased by introducing heterologous nucleic acid encoding a monooxygenase into a cell or microorganism. For example, introduction of heterologous nucleic acid encoding a monooxygenase can result in increased copy number of such nucleic acids and/or provide for modification of the cellular location in which the protein is expressed.

In some embodiments, a cytochrome P450 monooxygenase enzyme can be a fungal or bacterial protein. In a particular embodiment, the monooxygenase enzyme can be a *Candida* (e.g., *C. tropicalis*, *C. viswanathii*, *C. maltosa*, *C. cloacae*), *Yarrowia* (e.g., *Y. lipolytica*), *Fusarium* (e.g., *F. oxysporum*), *Bacillus* (e.g., *B. megaterium*, *B. subtilis*) protein. *Candida tropicalis* contains a family of cytochrome P450 genes referred to as *CYP* genes. Examples of *Candida viswanathii* nucleotide sequences encoding polypeptides having monooxygenase activities are provided herein (nucleotide SEQ ID NO: 99 and amino acid SEQ ID NO: 45) and in International patent application no. PCT/US2012/045615 (publication no. WO 2013/106730). Additional nonlimiting examples of nucleotide sequences encoding polypeptides having monooxygenase activity include: *Candida tropicalis CYP52A12* (Genbank accession no. AY230498), *Candida tropicalis CYP52A13* (Genbank accession no. AY230499), *Candida tropicalis CYP52A14* (Genbank accession no. AY230500), *Candida tropicalis CYP52A15* (Genbank accession no. AY230501), *Candida tropicalis CYP52A16* (Genbank accession no. AY230502), *Candida tropicalis CYP52A17* (Genbank accession no. AY230504), *Candida tropicalis CYP52A18* (Genbank accession no. AY230505), *Candida tropicalis CYP52A19* (Genbank accession no. AY230506), *Candida tropicalis CYP52A20* (Genbank accession no. AY230507), *Candida tropicalis CYP52D2* (Genbank accession no. AY230503), *Bacillus megaterium CYPBM3* (Genbank accession no. KC839476) and *Fusarium oxysporum CYP505* (Genbank accession no. AB030037).

Monooxygenase activity can be provided by any suitable polypeptide, such as a cytochrome P450 polypeptide (CYP450) in certain embodiments. Examples of a polypeptide having CYP450 activity include CYP52A12, a CYP52A13, a CYP52A14, a CYP52A15, a CYP52A16, a CYP52A17, a CYP52A18, a CYP52A19, a CYP52A20, a CYP52D2, and/or a BM3. In some

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1422 of 3077

embodiments, the activity can be a single polypeptide with both monooxygenase and monooxygenase reductase activities (e.g., *B. megaterium* cytochrome P450:NADPH P450 reductase, *Fusarium oxysporum* CYP505). Presence, absence or amount of cytochrome P450 activity can be detected by any suitable method known in the art. For example, detection can be performed by assaying a reaction containing cytochrome P450 (CYP52A family) and NADPH - cytochrome P450 reductase (see, e.g., Craft et al. (2003) *Appl. Environ. Microbiol*. 69: 5983 and 5992). Nucleic acid sequences encoding native and/or modified CYP450 sequences also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cell or organism exhibits increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism.

The promoter used for regulating transcription of a heterologous nucleic acid encoding a monooxygenase can also be modified. For example, the amount of a monooxygenase protein expressed in a particular cellular location may be increased by including in the heterologous nucleic acid a strong heterologous promoter and/or a promoter that provides for a different pattern of expression in the cell or microorganism. An example of one such heterologous promoter is a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene promoter. The nucleotide sequence of a *Candida viswanathii* HDE gene promoter is provided herein as are examples of additional fatty acid-inducible promoters. Promoter elements from different monooxygenase-encoding genes can have differing responsiveness to induction by various carbon sources. Thus, the amount of a monooxygenase protein expressed in a cell or organism can be modified by using heterologous promoters from different cytochrome P450 monoxygenase genes to regulate transcription of a monoxgenase-encoding nucleic acid that is introduced into a host cell and by the carbon source provided to the modified cell or organism. Non-limiting examples of assays suitable for assessing induction of cytochrome P450 (or other protein) expression by a carbon source or feedstock include RT-PCR or qRT-PCR after the host cell or microorganism has been exposed to the chosen carbon source or feedstock for varying amounts of time.

Decreasing monooxygenase activity in a cell can be accomplished by modifying the amount of monooxygenase protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous monooxygenase gene in a cell or organism with a weaker heterologous

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1423 of 3077

promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type monooxygenase such that the encoded modified or substituted monooxygenase protein has a reduced enzyme activity.

Modification of a cytochrome P450 reductase activity

A cytochrome P450 reductase (e.g., monooxygenase reductase activity or NADPH:cytochrome oxidoreductase (NCP); EC 1.6.2.4) can catalyze the reduction of the heme-thiolate moiety in cytochrome P450 by transferring electrons to the cytochrome P450. This activity recycles cytochrome P450 and makes it available for further use in catalyzing reactions that occur in ω-oxidation of fatty acids. In some embodiments of the cells, organisms, compositions and methods provided herein, the amount and/or activity of a cytochrome P450 reductase in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase the amount and/or activity of a cytochrome P450 reductase, may be modified to decrease the amount and/or activity of a cytochrome P450 reductase, or may be modified to alternately increase and decrease the amount and/or activity of a cytochrome P450 reductase depending, for example, on the cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

In certain aspects, the amount and/or activity of a cytochrome P450 reductase in a cell is increased. Increasing the amount and/or activity of a cytochrome P450 reductase may be particularly beneficial in embodiments in which the flux of carbons from fatty acids is directed toward a particular target product pathway involving oxidative metabolism and away from other cellular metabolic pathways not involved in target molecule production.

In certain embodiments, the cytochrome P450 reductase activity is unchanged in a host or engineered cell or organism. In one embodiment, the amount and/or activity of a host cytochrome P450 reductase can be increased, for example, by increasing the number of copies of a nucleic acid encoding a cytochrome P450 reductase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the nucleic acid), by increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a cytochrome P450 reductase, or by increasing the number of copies of a nucleic acid encoding a cytochrome P450 reductase and increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a cytochrome P450 reductase. In some embodiments, a cytochrome P450 reductase is endogenous to the host cell

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1424 of 3077

or microorganism. In one aspect of the cell or microbial systems and methods provided herein, the amount of cytochrome P450 reductase protein expressed in a cell can be increased by introducing heterologous nucleic acid encoding a cytochrome P450 reductase into a cell or microorganism. For example, introduction of heterologous nucleic acid encoding a cytochrome P450 reductase can result in increased copy number of such nucleic acids and/or provide for modification of the cellular location in which the protein is expressed.

In some embodiments, a cytochrome P450 reductase enzyme can be a yeast or bacterial protein. In a particular embodiment, the reductase enzyme can be a *Candida* (e.g., *C. tropicalis*, *C. viswanathii*, *C. maltosa*, *C. cloacae*), *Yarrowia* (e.g., *Y. lipolytica*) or *Bacillus* (e.g., *B. megaterium*) protein. In a particular embodiment, the cytochrome P450 reductase enzyme can be a *Candida* yeast protein. *Candida tropicalis* contains two alleles of a cytochrome P450 reductase gene referred to as *CPRa* and *CPRb*. Examples of *Candida viswanathii* nucleotide sequences encoding cytochrome P450 reductase activities are provided herein (nucleotide SEQ ID NO: 90 and amino acid SEQ ID NO: 45) and in International patent application no. PCT/US2012/045615 (publication no. WO 2013/106730). Additional non-limiting examples of nucleotide sequences encoding polypeptides having cytochrome P450 reductase activity include: *Candida tropicalis* (Genbank accession nos. AY705446, AY823228), *Candida bombicola* (Genbank accession no. EF050789), *Bacillus megaterium CYPBM3* (Genbank accession no. KC839476), *Bacillus megaterium* (Genbank accession no. FJ859036).

Presence, absence or amount of cytochrome P450 reductase activity can be detected by any suitable method known in the art. For example, detection can be performed by assaying a reaction containing cytochrome c and NADPH and measuring the rate of cytochrome reduction by monitoring absorbance (see, e.g., He and Chen (2005) *Yeast* 22:481-491; Van Bogaert et al. (2007) *Yeast* 7:922-928). Nucleic acid sequences encoding native and/or modified cytochrome P450 reductase also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cell or organism exhibits increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism.

The promoter used for regulating transcription of a heterologous nucleic acid encoding a cytochrome P450 reductase can also be modified. For example, the amount of a cytochrome

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1425 of 3077

P450 reductase protein expressed in a particular cellular location may be increased by including in the heterologous nucleic acid a strong heterologous promoter and/or a promoter that provides for a different pattern of expression in the cell or microorganism. An example of one such heterologous promoter is a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene promoter. The nucleotide sequence of a *Candida viswanathii* HDE gene promoter is provided herein as are examples of additional fatty acid-inducible promoters.

Alternatively, decreasing cytochrome P450 reductase activity in a cell can be accomplished by modifying the amount of cytochrome P450 reductase protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous cytochrome P450 reductase gene in an organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type cytochrome P450 reductase such that the encoded modified or substituted cytochrome P450 reductase protein has a reduced enzyme activity.

Modification of an alcohol dehydrogenase activity

A second step in the ω-oxidation pathway generally is the conversion of a fatty alcohol to a corresponding fatty aldehyde and involves NAD^+- or $NADP^+$-dependent fatty alcohol dehydrogenases and/or hydrogen peroxide-producing fatty alcohol oxidases. Oxidation of the alcohol to an aldehyde may be performed by an enzyme in the fatty alcohol oxidase family (e.g., long-chain fatty alcohol oxidase EC 1.1.3.20), or an enzyme in the alcohol dehydrogenase family (e.g., fatty alcohol dehydrogenase; EC 1.1.1.1). An alcohol dehydrogenase (e.g., fatty alcohol dehydrogenase, long-chain alcohol dehydrogenase) can catalyze the removal of a hydrogen from an alcohol to yield an aldehyde or ketone and a hydrogen atom and NADH.

In some embodiments of the cells, organisms, compositions and methods provided herein, the amount and/or activity of an alcohol dehydrogenase in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase the amount and/or activity of an alcohol dehydrogenase, may be modified to decrease the amount and/or activity of an alcohol dehydrogenase, or may be modified to alternately increase and decrease the amount and/or activity of an alcohol dehydrogenase depending, for example, on the cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1426 of 3077

In certain aspects, the amount and/or activity of an alcohol dehydrogenase in a cell is increased. Increasing the amount and/or activity of an alcohol dehydrogenase may be particularly beneficial in embodiments in which the flux of carbons from fatty acids is directed toward a particular target product pathway involving oxidative metabolism and away from other cellular metabolic pathways not involved in target molecule production.

In certain embodiments, the alcohol dehydrogenase activity is unchanged in a host or engineered cell or organism. In one embodiment, the amount and/or activity of a host alcohol dehydrogenase can be increased, for example, by increasing the number of copies of a nucleic acid encoding an alcohol dehydrogenase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the nucleic acid), by increasing the activity of a promoter that regulates transcription of a nucleic acid encoding an alcohol dehydrogenase, or by increasing the number of copies of a nucleic acid encoding an alcohol dehydrogenase and increasing the activity of a promoter that regulates transcription of a nucleic acid encoding an alcohol dehydrogenase. In some embodiments, an alcohol dehydrogenase is endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, the amount of alcohol dehydrogenase protein expressed in a cell can be increased by introducing heterologous nucleic acid encoding an alcohol dehydrogenase into a cell or microorganism. For example, introduction of heterologous nucleic acid encoding an alcohol dehydrogenase can result in increased copy number of such nucleic acids and/or provide for modification of the cellular location in which the protein is expressed.

In some embodiments, an alcohol dehydrogenase enzyme can be a yeast or bacterial protein. In a particular embodiment, the alcohol dehydrogenase enzyme can be a *Candida* (e.g., *C. tropicalis*, *C. viswanathii*, *C. maltosa*, *C. cloacae*), *Yarrowia* (e.g., *Y. lipolytica*) or *Bacillus* (*B. stearothermophilus*) protein. In a particular embodiment, the alcohol dehydrogenase enzyme can be a *Candida* yeast protein. *Candida tropicalis* contains at least 6 genes encoding alcohol dehydrogenases. Examples of *Candida viswanathii* nucleotide sequences encoding polypeptides having alcohol dehydrogenase activities are provided herein (nucleotide SEQ ID NO: 100 and amino acid SEQ ID NO: 46) and in International patent application no. PCT/US2012/045615 (publication no. WO 2013/106730). Additional examples of nucleotide sequences encoding polypeptides having alcohol dehydrogenase activity include, but are not limited to: *Candida tropicalis ADH1* (Genbank accession no. XM_002546589), *Candida utilis ADH1* (Genbank accession no. DQ397054), *Candida albicans ADH1* (Genbank accession no.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1427 of 3077

X81694), *Aspergillus flavus ADH1* (Genbank accession no. L27434), *Yarrowia lipolytica ADH1* (Genbank accession no. AF175271), *Yarrowia lipolytica ADH2* (Genbank accession no. AF175272), *Yarrowia lipolytica ADH3* (Genbank accession no. AF175273), *Bacillus stearothermophilus ADH-HT* (Genbank accession no. Z27089), *Pseudomonas putida ADHA* (Genbank accession no. AF052750).

Presence, absence or amount of alcohol dehydrogenase activity can be detected by any suitable method known in the art. For example, detection can be performed using spectrophotometric assays (see, e.g., Gatter et al. (2014) *FEMS Yeast Res.* 14:858-872). Nucleic acid sequences encoding native and/or modified alcohol dehydrogenase also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cell or organism exhibits decreased RNA and/or polypeptide levels as compared to the host cell or organism.

The promoter used for regulating transcription of a heterologous nucleic acid encoding an alcohol dehydrogenase can also be modified. For example, the amount of an alcohol dehydrogenase protein expressed in a particular cellular location may be increased by including in the heterologous nucleic acid a strong heterologous promoter and/or a promoter that provides for a different pattern of expression in the cell or microorganism. An example of one such heterologous promoter is a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene promoter. The nucleotide sequence of a *Candida viswanathii* HDE gene promoter is provided herein as are examples of additional fatty acid-inducible promoters.

Alternatively, decreasing alcohol dehydrogenase activity in a cell can be accomplished by modifying the amount of alcohol dehydrogenase protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous alcohol dehydrogenase gene in an organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type alcohol dehydrogenase such that the encoded modified or substituted alcohol dehydrogenase protein has a reduced enzyme activity.

Modification of fatty alcohol oxidase activity

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1428 of 3077

A fatty alcohol oxidase (e.g., long-chain alcohol oxidase, EC 1.1.3.20) enzyme can catalyze the oxidation of a fatty alcohol to yield a fatty aldehyde in the peroxisome of a cell. In some embodiments of the cells, organisms, compositions and methods provided herein, the amount and/or activity of a fatty alcohol oxidase in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase the amount and/or activity of a fatty alcohol oxidase, may be modified to decrease the amount and/or activity of a fatty alcohol oxidase, or may be modified to alternately increase and decrease the amount and/or activity of a fatty alcohol oxidase depending, for example, on the cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

In certain aspects, the amount and/or activity of a fatty alcohol oxidase in a cell is increased. Increasing the amount and/or activity of a fatty alcohol oxidase may be particularly beneficial in embodiments in which the flux of carbons from fatty acids is directed toward a particular target product pathway involving oxidative metabolism and away from other cellular metabolic pathways not involved in target molecule production.

In certain embodiments, the fatty alcohol oxidase activity is unchanged in a host or engineered cell or organism. In one embodiment, the amount and/or activity of a host fatty alcohol oxidase can be increased, for example, by increasing the number of copies of a nucleic acid encoding a fatty alcohol oxidase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the nucleic acid), by increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a fatty alcohol oxidase, or by increasing the number of copies of a nucleic acid encoding a fatty alcohol oxidase and increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a fatty alcohol oxidase. In some embodiments, a fatty alcohol oxidase is endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, the amount of fatty alcohol oxidase protein expressed in a cell can be increased by introducing heterologous nucleic acid encoding a fatty alcohol oxidase into a cell or microorganism. For example, introduction of heterologous nucleic acid encoding a fatty alcohol oxidase can result in increased copy number of such nucleic acids and/or provide for modification of the cellular location in which the protein is expressed.

In some embodiments, a fatty alcohol oxidase enzyme can be a yeast or bacterial protein. In a particular embodiment, the fatty alcohol oxidase enzyme can be a *Candida* (e.g., *C. tropicalis*, *C. viswanathii*, *C. maltosa*, *C. cloacae*), *Yarrowia* (e.g., *Y. lipolytica*) or *Bacillus* (e.g, *B.*

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1429 of 3077

stearothermophilus) protein. In a particular embodiment, the fatty alcohol oxidase enzyme can be a *Candida* yeast protein. *Candida tropicalis* contains two genes encoding fatty alcohol oxidase. Examples of *Candida viswanathii* nucleotide sequences encoding polypeptides having fatty alcohol oxidase activities are provided herein (nucleotide SEQ ID NO: 101 and amino acid SEQ ID NO: 47) and in International patent application no. PCT/US2012/045615 (publication no. WO 2013/106730). Additional examples of nucleotide sequences encoding polypeptides having fatty alcohol oxidase activity include, but are not limited to: *Candida tropicalis FAO1* (Genbank accession no. AY538780), *Candida tropicalis FAO2a* (Genbank accession no. AY538781), *Candida tropicalis FAO2b* (Genbank accession no. AY538782).

Presence, absence or amount of fatty alcohol oxidase activity can be detected by any suitable method known in the art. For example, detection can be performed using a two-enzyme coupled reaction assay (see, e.g., Eirich et al. (2004) *Appl. Environ. Microbiol.* 70(8):4872-4879). Nucleic acid sequences encoding native and/or modified fatty alcohol oxidase also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cell or organism exhibits decreased RNA and/or polypeptide levels as compared to the host cell or organism.

The promoter used for regulating transcription of a heterologous nucleic acid encoding a fatty alcohol oxidase protein expressed in a particular cellular location may be increased by including in the heterologous nucleic acid a strong heterologous promoter and/or a promoter that provides for a different pattern of expression in the cell or microorganism. An example of one such heterologous promoter is a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene promoter. The nucleotide sequence of a *Candida viswanathii* HDE gene promoter is provided herein as are examples of additional fatty acid-inducible promoters.

Alternatively, decreasing fatty alcohol oxidase activity in a cell can be accomplished by modifying the amount of fatty alcohol oxidase protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous fatty alcohol oxidase gene in an organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type fatty alcohol oxidase such that the encoded modified or substituted fatty alcohol oxidase protein has a reduced enzyme activity.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1430 of 3077

Modification of aldehyde dehydrogenase activity

A third step in the ω-oxidation pathway is generally the conversion of a fatty aldehyde to a corresponding fatty acid and involves NAD^+- or $NADP^+$-dependent fatty aldehyde dehydrogenases (e.g., long-chain-aldehyde dehydrogenase or fatty aldehyde dehydrogenase; EC 1.2.1.48). In some embodiments of the cells, organisms, compositions and methods provided herein, the amount and/or activity of a fatty aldehyde dehydrogenase in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase the amount and/or activity of a fatty aldehyde dehydrogenase, may be modified to decrease the amount and/or activity of a fatty aldehyde dehydrogenase, or may be modified to alternately increase and decrease the amount and/or activity of a fatty aldehyde dehydrogenase depending, for example, on the cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

In certain aspects, the amount and/or activity of a fatty aldehyde dehydrogenase in a cell is increased. Increasing the amount and/or activity of a fatty aldehyde dehydrogenase may be particularly beneficial in embodiments in which the flux of carbons from fatty acids is directed toward a particular target product pathway involving oxidative metabolism and away from other cellular metabolic pathways not involved in target molecule production.

In certain embodiments, the fatty aldehyde dehydrogenase activity is unchanged in a host or engineered cell or organism. In one embodiment, the amount and/or activity of a host fatty aldehyde dehydrogenase can be increased, for example, by increasing the number of copies of a nucleic acid encoding a fatty aldehyde dehydrogenase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the nucleic acid), by increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a fatty aldehyde dehydrogenase, or by increasing the number of copies of a nucleic acid encoding a fatty aldehyde dehydrogenase and increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a fatty aldehyde dehydrogenase. In some embodiments, a fatty aldehyde dehydrogenase is endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, the amount of fatty aldehyde dehydrogenase protein expressed in a cell can be increased by introducing heterologous nucleic acid encoding a fatty aldehyde dehydrogenase into a cell or microorganism. For example, introduction of heterologous nucleic acid encoding a

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1431 of 3077

fatty aldehyde dehydrogenase can result in increased copy number of such nucleic acids and/or provide for modification of the cellular location in which the protein is expressed.

In some embodiments, a fatty aldehyde dehydrogenase enzyme can be a yeast protein. In a particular embodiment, the fatty aldehyde dehydrogenase enzyme can be a *Candida* (e.g., *C. tropicalis*, *C. viswanathii*, *C. maltosa*, *C. cloacae*) or a *Yarrowia* (e.g., *Y. lipolytica*) yeast protein. In a particular embodiment, the fatty aldehyde dehydrogenase enzyme can be a *Candida* yeast protein. Examples of *Candida viswanathii* nucleotide sequences encoding polypeptides having fatty aldehyde dehydrogenase activities are provided, for example, in International patent application no. PCT/US2012/045615 (publication no. WO 2013/106730). Additional examples of nucleotide sequences encoding polypeptides having fatty aldehyde dehydrogenase activity include, but are not limited to: *Yarrowia lipolytica HFD1* (Genbank accession no. AB935099), *Yarrowia lipolytica HFD2A* (Genbank accession no. AB935101), *Yarrowia lipolytica HFD2B* (Genbank accession no. AB935103), *Yarrowia lipolytica HFD3* (Genbank accession no. AB935104), *Yarrowia lipolytica HFD4* (Genbank accession no. AB935106).

Presence, absence or amount of fatty aldehyde dehydrogenase activity can be detected by any suitable method known in the art. For example, detection can be performed using enzyme activity assays (see, e.g., Iwama et al. (2014) *J. Biol. Chem.* 289(48):33275-33286). Nucleic acid sequences encoding native and/or modified fatty aldehyde dehydrogenase also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cell or organism exhibits decreased RNA and/or polypeptide levels as compared to the host cell or organism.

The promoter used for regulating transcription of a heterologous nucleic acid encoding an fatty aldehyde dehydrogenase can also be modified. For example, the amount of a fatty aldehyde dehydrogenase protein expressed in a particular cellular location may be increased by including in the heterologous nucleic acid a strong heterologous promoter and/or a promoter that provides for a different pattern of expression in the cell or microorganism. An example of one such heterologous promoter is a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene promoter. The nucleotide sequence of a *Candida viswanathii* HDE gene promoter is provided herein as are examples of additional fatty acid-inducible promoters.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1432 of 3077

Alternatively, decreasing fatty aldehyde dehydrogenase activity in a cell can be accomplished by modifying the amount of fatty aldehyde dehydrogenase protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous fatty aldehyde dehydrogenase gene in an organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type fatty aldehyde dehydrogenase such that the encoded modified or substituted fatty aldehyde dehydrogenase protein has a reduced enzyme activity.

Modification of β-oxidation

Another oxidative metabolism pathway, referred to as β-oxidation, is generally a degradative pathway through which fatty acids, typically in the form of fatty acid-CoA esters, can be broken down to shorter chain acyl-CoA and acetyl-CoA. In fungi and plant cells, β-oxidation can occur in peroxisomes, whereas in animal cells it additionally can take place in mitochondria. The β-oxidation pathway generally includes four main reaction steps resulting in an acyl-CoA that is shortened by two carbon atoms which are released as acetyl-CoA. The shortened acyl-CoA molecule can re-enter the pathway after each cycle and be subjected to another removal of two carbons from the acyl carbon chain. As such, the β-oxidation pathway can generate significant amounts of acetyl-CoA and is a major source of acetyl-CoA in cells. Alteration of enzyme activities in the the β-oxidation pathway can also provide for the generation of fatty acid or diacid target molecules including, but not limited to, adipic acid, suberic acid, sebacic acid and dodecanedioic acid (DDDA). Provided herein are cells, organisms, compositions and methods in which cellular carbon flux has been modified through one or more alterations in the β-oxidation pathway. In some embodiments, the β-oxidation pathway is modified by modifying one or more activities in the pathway. In particular embodiments, the β-oxidation pathway is modified to increase the generation of acetyl-CoA in a cell for use in target molecule production. For example, the β-oxidation pathway can be modified to increase one or more activities in the pathway. In some embodiments, β-oxidation can be manipulated (e.g., decreasing one or more pathway activities and/or altering the specificity of one or more activities) to be used as a pathway for production of target fatty acids and diacids (e.g., adipic acid) of a particular carbon chain length.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1433 of 3077

The term "β-oxidation pathway" as used herein, refers to a series of cellular enzymatic activities utilized to metabolize fatty alcohols, fatty acids, or dicarboxylic acids. The activities utilized to metabolize fatty alcohols, fatty acids, or dicarboxylic acids can include, but are not limited to, acyl-CoA oxidase activity, acyl-CoA hydrolase activity, enoyl-CoA hydratase activity, 3-hydroxyacyl-CoA dehydrogenase activity, and acetyl-CoA C-acyltransferase activity. The term "β-oxidation activity" refers to any and/or all of the activities in the β-oxidation pathway utilized to metabolize fatty alcohols, fatty acids or dicarboxylic acids. Additional activities, referred to as β-oxidation peripheral or auxillary activities, can be involved in degradation of unsaturated fatty acids (i.e., fatty acid chains containing double bonds) and fatty acids containing modifications (e.g., hydroxyl, methyl, phenoxy groups) including, but not limited to enoyl CoA isomerase ((ECI) or enoyl-CoA delta isomerase 1, dodecenoyl-CoA isomerase, 3,2 trans-enoyl-CoA isomerase, acetylene-allene isomerase, delta3, delta2-enoyl-CoA isomerase, dodecenoyl-CoA delta isomerase, and EC 5.3.3.8), dienoyl CoA Isomerase (DCI, e.g., EC 5.3.3, Δ3,5,Δ2,4-dienoyl-CoA isomerase, Δ3,5,Δ2,4-dienoyl-coenzyme A isomerase) and 2,4-dienoyl-CoA reductase (DCR, e.g., EC 1.3.1.34).

There are also cellular compositions and activities that can be closely associated with β-oxidation. These include peroxisomal- and mitochondrial-related compositions and activities. For example, as described herein, such compositions and activities include, but are not limited to, acyl-CoA synthetases, thioesterases, peroxisomal transport proteins and peroxisomal biogenesis factors. Included in the cells, organisms, systems and methods provided herein are embodiments in which one or more of these β-oxidation-associated compositions and/or activities are modified. In some embodiments, a β-oxidation-associated composition or activity is modified to enhance β-oxidation activity.

Modification of acyl-CoA oxidase activity

Typically, the first step in the β-oxidation pathway is oxidation of acyl-CoA, which is carried out by the enzyme acyl-CoA oxidase (e.g., EC 1.3.3.6). This step can be a rate-limiting step in β-oxidation. The term "acyl-CoA oxidase activity" as used herein refers to the enzymatic activity (e.g., catalytic activity) of an acyl-CoA oxidase. An acyl-CoA oxidase can catalyze the following chemical reaction:

$$\text{acyl-CoA} + O_2 \leftrightarrow \text{trans-2-enoyl-CoA} + H_2O_2$$

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1434 of 3077

Acyl-CoA oxidase enzymes generally contain FAD from which two electrons are transferred to oxygen to generate H_2O_2.

Different cells contain different types, and numbers of types, of acyl-CoA oxidase activities. For example, *Saccharomyces cerevisiae* expresses only one acyl-CoA oxidase, Pox1p/Fox1p, which has activity on acyl-CoA substrates with a broad range of carbon chain lengths. In contrast, other organisms, e.g., species of *Candida*, *Yarrowia*, *Arabidopsis*, can have multiple genes encoding different proteins having acyl-CoA oxidase activities with varying substrate specificities. In some embodiments, acyl-CoA oxidase activity refers to its enzyme activity (or lack thereof) on a selective set of substrates. The activity of an acyl-CoA oxidase can be affected by its ability to bind a substrate, oxidize a substrate and/or release a product. In some embodiments, an acyl-CoA oxidase is active in one compartment of a cell and not in another compartment of the cell. In some embodiments, an acyl-CoA oxidase activity is from a peroxisome.

Different Acyl-CoA oxidases can display different carbon chain-length substrate specificities. Some acyl-CoA oxidases display broad chain-length specificity and can accept any fatty acyl-CoA (or diacyl-CoA) as a substrate. However, some acyl-CoA oxidases can display narrow chain-length specificity. For example, the acyl-CoA oxidase activity encoded by the *POX4* gene of *Candida* strain ATCC 20336 has a relatively broad carbon chain-length specificity and exhibits a higher specific activity for acyl-CoA molecules with shorter carbon chain lengths (e.g., less than 10 carbons). The Pox5 enzyme from *Candida* strain ATCC 20336 displays optimal activity on fatty acid substrates having 12 to 18 carbons (C12 - C18) in the carbon chain, a decreased activity on substrates having less than 10 carbons (C10) in the carbon chain and has low activity on C6 and C8 substrates. In a cell with such a *Candida* Pox5 as the only functional acyl-CoA oxidase, long chain fatty acyl-CoA or diacyl-CoA substrates can be shortened to about 8 carbons and do not typically enter another cycle of β-oxidation. The shorter substrates (e.g., a C8 fatty acid or dicarboxylic acid) are not typically recognized as a substrate by Pox5. In this instance, an acyl-CoA would not be completely broken down to acetyl-CoA units. Instead, the chain-length substrate specificity of the acyl-CoA oxidase in such a cell (which would limit further degradation of an acyl-CoA once it has been broken down into about an 8-carbon chain) effectively controls the chain length of an acid or diacid produced through break down of fatty acids through β-oxidation. The shorter substrates (e.g., a C8 acyl-CoA) would remain intact, the

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1435 of 3077

CoA would be removed by peroxisomal thioesterases and the fatty acid or dicarboxylic acid (e.g., an α,ω - dicarboxylic acid) product is secreted from the cell. In this way, β-oxidation can be manipulated to be used as a pathway for production of target fatty acids and diacids (e.g., adipic acid) of a particular carbon chain length.

In some embodiments of the cells, organisms, compositions and methods provided herein, the amount and/or activity of one or more acyl-CoA oxidases in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase the amount and/or activity of an acyl-CoA oxidase, may be modified to decrease the amount and/or activity of an acyl-CoA oxidase, or may be modified to alternately increase and decrease the amount and/or activity of one or more acyl-CoA oxidases depending, for example, on the substrate specificity, target molecule(s) being produced, cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

In certain aspects, the amount and/or activity of an acyl-CoA oxidase in a cell is increased. Increasing the amount and/or activity of an acyl-CoA oxidase may be particularly beneficial in embodiments in which the flux of carbons from fatty acids is directed toward a particular target product pathway involving oxidative metabolism and away from other cellular metabolic pathways not involved in target molecule production.

In certain embodiments, the acyl-CoA oxidase activity is unchanged in a host or engineered cell or organism. In one embodiment, the amount and/or activity of a host acyl-CoA oxidase can be increased, for example, by increasing the number of copies of a nucleic acid encoding an acyl-CoA oxidase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the nucleic acid), by increasing the activity of a promoter that regulates transcription of a nucleic acid encoding an acyl-CoA oxidase, or by increasing the number of copies of a nucleic acid encoding an acyl-CoA oxidase and increasing the activity of a promoter that regulates transcription of a nucleic acid encoding an acyl-CoA oxidase. In some embodiments, an acyl-CoA oxidase is endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, the amount of an acyl-CoA oxidase protein expressed in a cell can be increased by introducing heterologous nucleic acid encoding an acyl-CoA oxidase into a cell or microorganism. For example, introduction of heterologous nucleic acid encoding an acyl-CoA oxidase can result in increased copy number of such nucleic acids and/or provide for modification of the cellular location in which the protein is expressed.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1436 of 3077

Non-limiting examples of organisms that include, or can be used as donors for, an acyl-CoA oxidase enzyme include yeast (e.g., *Candida*, *Saccharomyces*, *Debaryomyces*, *Meyerozyma*, *Lodderomyces*, *Scheffersomyces*, *Clavispora*, *Yarrowia*, *Pichia*, *Kluyveromyces*, *Eremothecium*, *Zygosaccharomyces*, *Lachancea*, *Nakaseomyces*), animals (e.g., *Homo*, *Rattus*), bacteria (e.g., *Escherichia*, *Pseudomonas*, *Bacillus*), or plants (e.g., *Arabidopsis*, *Nictotania*, *Cuphea*). In a particular embodiment, an acyl-CoA oxidase enzyme can be a *Candida* yeast protein. Examples of *Candida viswanathii* nucleotide sequences encoding acyl-CoA oxidases are provided herein (nucleotide SEQ ID NOS: 92 and 93 and amino acid SEQ ID NOS: 36, 37, 38 and 39) and in International patent application no. PCT/US2012/045622 (publication no. WO 2013/006733) and International patent application no. PCT/US2013/076739 (publication no. WO 2014/100504). Additional examples of nucleotide sequences encoding polypeptides having acyl-CoA oxidase activity include: *Saccharomyces cerevisiae POX1* (Genbank accession no. M27515), *Candida albicans POX1-3* (Genbank accession no. XM_716636), *Candida tropicalis POX2* (Genbank accession no. XM_002548031), *Candida tropicalis POX5* (Genbank accession no. XM_002548378),*Candida tropicalis POX4* (Genbank accession nos. AB031271, AB031272), *Candida maltosa POX2* (Genbank accession no. D21228), *Yarrowia lipolytica ACO1* (Genbank accession no. AJ001299), *Yarrowia lipolytica ACO2* (Genbank accession no. A001300), *Yarrowia lipolytica ACO3* (Genbank accession no. AJ001301), *Yarrowia lipolytica ACO4* (Genbank accession no. AJ001302), *Yarrowia lipolytica ACO5* (Genbank accession no. AJ001303), *Debaryomyces fabyri* (Genbank accession no. XM_015613952).

Presence, absence or amount of acyl-CoA oxidase activity can be detected by any suitable method known in the art and/or described herein. For example, detection can be performed using enzyme activity assays (see, e.g., Shimizu et al. (1979) *Biochem. Biophys. Res. Commun*. 91:108-113; Yao et al. (2014) *J. Braz. Chem. Soc.* 25(4):777-782); Kvannes and Flatmark (1991) *J. Biochem. Biophys. Methods* 23(2):135-149). Native and/or disrupted nucleic acid sequences encoding acyl-CoA oxidase (or other polypeptide) also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), and the amount of the nucleic acids or encoded proteins can be asessed using quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cells or organisms exhibit increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1437 of 3077

The promoter used for regulating transcription of a heterologous nucleic acid encoding an acyl-CoA oxidase can also be modified. For example, the amount of an acyl-CoA oxidase protein expressed in a particular cellular location may be increased by including in the heterologous nucleic acid a strong heterologous promoter and/or a promoter that provides for a different pattern of expression in the cell or microorganism. An example of one such heterologous promoter is a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene promoter. The nucleotide sequence of a *Candida viswanathii* HDE gene promoter is provided herein as are examples of additional fatty acid-inducible promoters.

Alternatively, decreasing acyl-CoA oxidase activity in a cell can be accomplished by modifying the amount of acyl-CoA oxidase protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous acyl-CoA oxidase gene in an organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type acyl-CoA oxidase such that the encoded modified or substituted acyl-CoA oxidase protein has a reduced enzyme activity. Reducing or eliminating the amount and/or activity of an acyl-CoA oxidase may be particularly beneficial in embodiments in which a target molecule and/or precursor or intermediate in the production of a target molecule contains a carbon chain of a particular length. In this case, the processing of fatty acids of particular chain lengths may be decreased or eliminated by decreasing the amount and/or activity of a particular acyl-CoA oxidase having activity on fatty acids of the particular chain length in a cell. Certain aspects of the cells, microorganisms, compositions and methods provided herein include one or more modifications to reduce or eliminate an acyl-CoA oxidase. One approach to reducing or eliminating the amount and/or activity of an acyl-CoA oxidase is by disrupting or deleting nucleic acid encoding the acyl-CoA oxidase in a host cell or microorganism to reduce or eliminate the acyl-CoA oxidase activity in the host relative to a cell or microorganism in which the gene(s) have not been modified. For example, expression of a host acyl-CoA oxidase activity can be decreased or eliminated by disruption (e.g., knockout, insertion mutagenesis, the like and combinations thereof) of a host gene encoding the protein, or by decreasing the activity of the promoter (e.g., through addition of repressor sequences to the promoter or 5'UTR or replacing the promoter) that controls transcription of an acyl-CoA oxidase gene using recombinant molecular biology techniques known in the art and/or described herein.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1438 of 3077

One method for disrupting an endogenous acyl-CoA oxidase gene is by recombinantly inserting a heterologous nucleic acid (e.g., a nucleotide sequence encoding a selectable marker such as an enzyme that restores an auxotrophic host organism to prototrophy) into the endogenous gene, thereby generating an engineered organism deficient in acyl-CoA oxidase activity. This can be done, for example, through homologous recombination in which a heterologous nucleic acid containing sequences of an endogenous acyl-CoA oxidase gene and a disrupting sequence (e.g., a knock-out gene cassette such as described herein) is introduced into a host cell or microorganism. In some embodiments, the nucleotide sequence of one or more acyl CoA oxidases (e.g., a yeast POX4, POX 5, or POX4 and POX5) can be disrupted with a URA3 nucleotide sequence encoding a selectable marker, and introduced to a host cell or microorganism, thereby generating an engineered cell or organism deficient in an acyl-CoA oxidase activity. Nucleic acids encoding an acyl-CoA oxidase can be obtained from a number of sources, including, for example, yeast cells. Genomic DNA from cell sources can be amplified using oligonucleotide primers based on the nucleotide sequence of an acyl-CoA oxidase encoding gene, including examples provided herein.

In some embodiments, of the cells, organisms, compositions and methods provided herein, the amount and/or activity of a peroxisomal acyl-CoA oxidase in a cell is modified, e.g., a POX activity of a POX polypeptide. In particular embodiments, the acyl-CoA oxidase activity to be modified is encoded by the *POX4* and/or *POX5* genes of a species of *Candida* (e.g., ATCC 20336). In certain embodiments, the amount and/or activity of an endogenous acyl-CoA oxidase can be increased. In some embodiments, the amount and/or activity of acyl-CoA oxidases in a cell or organism containing one or more acyl-CoA oxidases can be independently modified (e.g., one or more acyl-CoA oxidases can be modified). In some embodiments, the amount and/or activity of POX4 acyl-CoA oxidase and a POX5 acyl-CoA oxidase can be altered independently of each other (e.g., increase amount and/or activity of POX4 alone, POX5 alone, increase amount and/or activity of one and decrease or eliminate the amount and/or activity of the other, and the like). Increasing the amount and/or activity of one acyl-CoA oxidase, while decreasing or eliminating the amount and/or activity of another acyl-CoA oxidase, may alter the specific activity of acyl-CoA oxidase in a cell or organism with respect to carbon chain length, while maintaining or increasing overall carbon flux through the β-oxidation pathway, in certain embodiments. Disruption of nucleotide sequences encoding one or more acyl-CoA oxidases (e.g., POX4, POX 5, or POX4 and POX5) sometimes can alter pathway efficiency, specificity and/or specific activity with respect to metabolism of carbon chains of different lengths (e.g.,

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1439 of 3077

carbon chains including fatty alcohols, fatty acids, paraffins, dicarboxylic acids, aliphatic molecules of between about 1 and about 26 carbons in length).

In some embodiments of the modified cells or organisms provided herein, a β-oxidation pathway in a yeast is active and includes a genetically modified acyl-CoA oxidase. In some embodiments, an acyl-CoA oxidase is genetically modified to prevent complete oxidation of fatty acyl-CoA or diacyl-CoA substrates. Genetic modification of an acyl-CoA oxidase can increase the production yield of a desired fatty acid or fatty dicarboxylic acid product. Therefore, in some embodiments, metabolic degradation of a fatty acid of a specified chain length (e.g., the chain length of a desired or target fatty acid or fatty dicarboxylic acid product) is reduced significantly, when an acyl-CoA oxidase is genetically modified. For example, metabolic degradation of a fatty dicarboxylic acid product (e.g., dodecanedioic acid or DDDA) by beta-oxidation can be reduced significantly when an acyl-CoA oxidase is genetically modified. This can be accomplished by modifying the substrate specificity of an acyl-CoA oxidase such that the enzyme has low activity (e.g., enzymatic activity) on chain lengths equal to or less than that of a desired product.

Nucleic acids encoding a genetically modified acyl-CoA oxidase can be engineered and expressed in a suitable organism (e.g., bacteria (e.g., *E. coli*) or a yeast) to test the substrate specificity of the modified enzyme in vitro. In some embodiments, nucleic acids encoding a genetically modified acyl-CoA oxidase are engineered and expressed in a suitable yeast and the substrate specificity is tested. Yeast that express a modified acyl-CoA oxidase can be tested for production of a desired molecule, e.g., a fatty acid or fatty dicarboxylic acid product. A modified acyl-CoA oxidase can be generated in any suitable manner (e.g., random or rational mutagenesis), non-limiting examples of which are provided herein and, for example, in International patent application no. PCT/US2012/045622 (publication no. WO 2013/006733) and International patent application no. PCT/US2013/076739 (publication no. WO 2014/100504).

In some instances, a modified endogenous acyl-coA oxidase polypeptide is a modified POX4 or POX5 polypeptide from a *Candida* spp. yeast (e.g., strain ATCC 20336 or ATCC 20962). In some cases a modified POX4 polypeptide contains a modified amino acid sequence of the wild-type *Candida* strain ATCC 20336 Pox4p sequence provided herein. Sometimes the POX4 polypeptide contains an amino acid modification at one or more amino acid positions chosen from 88, 90, 96, 98, 99, 100, 102, 103, 302, 309, 310, 473, 474, 475, 476, 477, 478, 479, 480,

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1440 of 3077

481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 501, 502, 503, 504 and 505. A modified endogenous acyl-coA oxidase polypeptide that is not a modified POX4 polypeptide can include an amino acid modification at one or more positions corresponding to one or more of the foregoing positions in the POX4 polypeptide. In some instances a modified POX5 polypeptide contains a modified amino acid sequence of the wild-type *Candida* strain ATCC 20336 Pox5p sequence provided herein. Sometimes the POX5 polypeptide contains an amino acid modification at one or more amino acid positions chosen from 81, 82, 83, 84, 85, 86, 88, 93, 94, 95, 96, 98, 102, 284, 287, 290, 291, 292, 294, 295, 436, 453, 454, 455, 456, 457, 458, 459, 460, 461, 462 and 463. A modified endogenous acyl-coA oxidase polypeptide that is not a modified POX5 polypeptide can include an amino acid modification at one or more positions corresponding to one or more of the foregoing positions in the POX5 polypeptide.

In some embodiments, the substrate specificity of an acyl-CoA oxidase is modified such that the enzyme has low activity for aliphatic molecules with chain lengths less than C24 (i.e., 24 carbons). In some embodiments, the substrate specificity of an acyl CoA oxidase is modified such that the enzyme has very low activity with chain lengths less than 24, 22, 20, 18, 16, 14, 12, 10, 8, 6 or 4 carbons. In some embodiments, the substrate specificity of an acyl-CoA oxidase is modified such that the enzyme has very low activity with chain lengths less than 18, 16, 14, 12, 10 or 8 carbons. In some embodiments, the substrate specificity of an acyl-CoA oxidase is modified such that the enzyme has very low activity with chain lengths less than C12. In some embodiments, the substrate specificity of an acyl-CoA oxidase is modified such that the enzyme has very low activity with chain lengths less than C10 or C8. For example, in one embodiment for producing a 6-carbon dicarboxylic acid (e.g., adipic acid), a host cell or organism can be modified to decrease or eliminate acyl-CoA oxidase activities that are active on a broad range of substrate chain lengths (e.g., Pox4p acyl-CoA oxidase of *Candida viswanathii* ATCC 20336), and, to further increase productivity, can additionally be modified to express a mutant acyl-CoA oxidase activity that is more active on substrates with chain lengths of C8 and greater with little or no activity on substrates with chain lengths less than C8 (e.g., *Candida viswanathii* ATCC 20336 Pox5p(F98G); SEQ ID NO: 37).

As described herein (and in International patent application no. PCT/US2012/045622 (publication no. WO 2013/006733) and International patent application no. PCT/US2013/076739 (publication no. WO 2014/100504)), catalytic specificity of acyl-CoA oxidases (e.g., POX4,

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1441 of 3077

POX5) can be altered by a variety of methods. Altering the binding and/or catalytic specificity of acyl-CoA oxidases may prove advantageous for generating novel acyl-CoA oxidases with altered chain length recognition, altered chain length catalytic activity, and/or generation of an acyl-CoA oxidase activity with a narrow or specific chain length specificity, thereby allowing further increases in pathway efficiency, specificity and/or specific activity with respect to metabolism of carbon chains of different lengths or metabolism of carbon chain distributions found in a particular chosen feedstock. In some embodiments the altered acyl-CoA oxidase sequences are identified and/or generated by; (i) screening naturally occurring variant populations; (ii) mutagenesis of endogenous sequences; (iii) introduction of heterologous sequences having a desired specificity; (iv) generation of chimeric sequences having a portion of the coding sequence from one polynucleotide source (e.g., gene, organism) and a portion of the coding sequence from another source and/or (v) intelligent design using nucleotide sequences and three dimensional structure analysis from an acyl-CoA oxidase having a desired specificity to remodel an endogenous acyl-CoA oxidase, thereby generating a novel specificity enzyme. In some embodiments, a chimeric acyl-CoA oxidase nucleic acid sequence can have polynucleotide sequence contributions from two or more sources. In some embodiments, a chimeric acyl-CoA oxidase nucleic acid sequence comprises a portion of the coding sequences from an endogenous polynucleotide and a portion of the coding sequence from a heterologous polynucleotide.

One method for generating modified acyl-CoA oxidase proteins having altered substrate specificity is through random mutagenesis. A library of genetically modified acyl-CoA oxidases can be generated using several methods known in the art (e.g., site-directed mutagenesis). Genetically modified acyl-CoA oxidase genes can then be transformed into a β-oxidation blocked strain of a suitable yeast strain (e.g., *Candida* spp. (e.g., *Candida viswanathii* or *Candida tropicalis*)). In some embodiments, a genetically modified acyl-CoA oxidase is expressed under the control of the *POX4* promoter or another strong constitutive or inducible promoter in a pox4Δ/ pox4Δ pox5Δ/pox5Δ (e.g., an organism that lacks some or all endogenous acyl-CoA oxidase activity) background. In some embodiments, the genetically modified acyl-CoA oxidase is expressed under the control of an endogenous promoter. In some embodiments, the genetically modified acyl-CoA oxidase is expressed under the control of a heterologous promoter. The transformants can be selected by growth in media containing a fatty acid or methyl-derivate fatty acid containing fatty acids with two more carbons than a fatty acid product of interest. For example, for an adipic acid product, the transformants can be

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1442 of 3077

grown in caprylic acid or methyl-caprylate. For example, for a dodecanedioic acid product, the transformants can be grown in tetradecanedioic acid. The group of transformants can then be moved to a medium with a carbon source of a fatty acid of interest (for example dodecanedioic acid) in the presence of an agent that kills growing cells (e.g., Nystatin) and cells that cannot metabolize the carbon source (e.g., dodecanedioic acid in this example) can be selected. The resulting modified strains can then be further characterized for acyl-CoA oxidase activity. This method can be used to select for any modified acyl CoA oxidase (e.g., those listed and/or described in International patent application no. PCT/US2012/045622 (publication no. WO 2013/006733) and International patent application no. PCT/US2013/076739 (publication no. WO 2014/100504)). In addition, this method can be used to select for any heterologous acyl-CoA oxidase expressed in a suitable organism.

Another method for generating modified acyl-CoA oxidase proteins having altered substrate specificity is through rational mutagenesis. Structural and sequence information and experimental data can be combined to determine specific mutations for testing in an acyl-CoA oxidase for altered specificity. For example, primary sequences of acyl-CoA oxidases tested can be compared and correlated with substrate specificity. Based on such an analysis, single amino-acids, small numbers of contiguous amino acids and/or domains can be proposed for providing a desired substrate specificity. Those amino acids positions can be targeted for specific or random mutations for improve specificity.

Acyl-CoA oxidase structure also can be modeled against a known tertiary structure using modeling methods known in the art. The models can be used to propose amino acids and regions pertaining to substrate selectivity. For example, biochemical, structure and sequence data suggest that the N-terminus of acyl-CoA oxidases often, in part, determines substrate specificity. Mutations or region replacements can be introduced based on such analyses and the specificity of the new acyl-CoA oxidase tested as described before. The resulting information can be used to go back to the models to postulate new potential mutations. As for random mutagenesis, any suitable acyl-CoA oxidase can be modified to alter substrate specificity (e.g., those listed in International patent application no. PCT/US2012/045622 (publication no. WO 2013/006733) and International patent application no. PCT/US2013/076739 (publication no. WO 2014/100504)).

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1443 of 3077

Examples of modified Pox5 enzymes encoded by mutated *POX5* genes from *Candida viswanathii* include Pox5p (F98G) and Pox5p(W249F) are provided herein. The design, generation and analysis of modified Pox5 enzymes encoded by mutated *POX5* genes from *Candida viswanathii* are described in the examples.

Modification of multifunctional enzyme activities

Next (e.g., second and third) steps of the β-oxidation pathway can be catalyzed by a multifunctional enzyme (referred to, for example, as Mfe2, Fox2 and HDE in fungi) having hydratase and dehydrogenase activities, or by separate hydratase and dehydrogenase enzymes. In these steps, a trans-2-enoyl-CoA can be converted to 3-ketoacyl-CoA via a (3*R*)-hydroxy intermediate. An enoyl-CoA hydratase enzyme (e.g., EC 4.2.1.17) can catalyze the addition of a hydroxyl group and a proton to the unsaturated β-carbon on a fatty-acyl CoA in a second step of the pathway to generate 3-hydroxyacyl-CoA. In a next (e.g., third) step, a 3-hydroxyacyl-CoA dehydrogenase enzyme (e.g., EC 1.1.1.35) can catalyze the formation of a 3-ketoacyl-CoA by removal of a hydrogen from the newly formed hydroxyl group created by the activity of an enoyl-CoA hydratase. Typically, fungi have one peroxisomal multifunctional enzyme (HDE, Mfe2 or Fox2), mammalian cells have two peroxisomal multifunctional enzymes (Mfe1 and Mfe2) and bacteria have a single multifunctional enzyme, Mfe1. In the yeast *Candida tropicalis*, the N-terminal portion of the MFE polypeptide typically contains two duplicate 3-hydroxyacyl-CoA dehydrogenase domains, referred to as the A and B domains, which have differing substrate specificities. The A domain can catalyze the reaction for substrates with medium-to-long carbon chains (e.g., C10 – C16). The catalytic activity of the B domain often is more active on substrates having shorter carbon chains (e.g., C4). The hydratase domain is generally located at the C-terminal region of the polypeptide. Thus, each Mfe2 monomer can contain a dehydrogenase heterodimer and a hydratase monomer.

Some multifunctional enzymes involved in the β-oxidation pathway have additional enzymatic activities, including, but not limited to, an isomerase (e.g., a Δ3,Δ2-enoyl-CoA isomerase) activity and/or an epimerase (e.g., 3-hydroxyacyl-CoA epimerase; EC 5.1.2.3) activity. These enzymes function as auxillary enzymes in the oxidation of polyunsaturated fatty acids. For example, 3-hydroxyacyl-CoA epimerase catalyzes the reversible conversion of *S*-3-hydroxyacyl-CoA to *R*-3-hydroxyacyl-CoA, which (unlike *S*-3-hydroxyacyl-CoA) is a substrate for 3-

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1444 of 3077

hydroxyacyl-CoA dehydrogenase contained within Mfe2-type enzymes. Therefore, β-oxidation can proceed through the third step once the *R* isomer has been generated.

In some embodiments of the cells, organisms, compositions and methods provided herein, the amount and/or activity of a multifunctional enzyme (or an enoyl-CoA hydratase or 3-hydroxyacyl-CoA dehydrogenase) in a cell is modified. In particular embodiments, the multifunctional enzyme is a peroxisomal protein. For example, in some aspects, a cell or microorganism may be modified to increase the amount and/or activity of a multifunctional enzyme (or an enoyl-CoA hydratase or 3-hydroxyacyl-CoA dehydrogenase), may be modified to decrease the amount and/or activity of a multifunctional enzyme (or an enoyl-CoA hydratase or 3-hydroxyacyl-CoA dehydrogenase), or may be modified to alternately increase and decrease the amount and/or activity of a multifunctional enzyme (or an enoyl-CoA hydratase or 3-hydroxyacyl-CoA dehydrogenase) depending, for example, on the substrate specificity, target molecule(s) being produced, cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured. In some embodiments, the amount and/or activity of one or more of the hydratase and dehydrogenase enzymes of a multifunctional enzyme may be independently modified.

In certain aspects, the amount and/or activity of one or more of the hydratase and dehydrogenase enzymes of a multifunctional enzyme in a cell is increased. Increasing the amount and/or activity of one or more of the hydratase and dehydrogenase enzymes of a multifunctional enzyme may be particularly beneficial in embodiments in which the flux of carbons from fatty acids is directed toward a particular target product pathway involving oxidative metabolism and away from other cellular metabolic pathways not involved in target molecule production.

In certain embodiments, the multifunctional enzyme (or an enoyl-CoA hydratase or 3-hydroxyacyl-CoA dehydrogenase) activity is unchanged in a host or engineered cell or organism. In one embodiment, the amount and/or activity of one or more of a hydratase and/or dehydrogenase enzyme, for example, of a multifunctional enzyme, can be increased, for example, by increasing the number of copies of a nucleic acid encoding one or more of a hydratase and/or dehydrogenase enzyme (for example, of a multifunctional enzyme) (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the nucleic acid), by increasing the activity of a promoter that regulates transcription of a nucleic acid encoding one or more of a hydratase

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1445 of 3077

and dehydrogenase enzyme (for example, of a multifunctional enzyme), or by increasing the number of copies of a nucleic acid encoding one or more of a hydratase and/or dehydrogenase enzyme (for example, of a multifunctional enzyme) and increasing the activity of a promoter that regulates transcription of a nucleic acid encoding one or more of a hydratase and/or dehydrogenase enzyme (for example, of a multifunctional enzyme). In some embodiments, a multifunctional enzyme (or an enoyl-CoA hydratase and/or 3-hydroxyacyl-CoA dehydrogenase) is endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, the amount of one or more of the hydratase and dehydrogenase enzymes of a multifunctional enzyme protein expressed in a cell can be increased by introducing heterologous nucleic acid encoding one or more of the hydratase and dehydrogenase enzymes of a multifunctional enzyme into a cell or microorganism. For example, introduction of heterologous nucleic acid encoding one or more of the hydratase and dehydrogenase enzymes of a multifunctional enzyme can result in increased copy number of such nucleic acids and/or provide for modification of the cellular location in which the protein is expressed.

Non-limiting examples of organisms that include, or can be used as donors for, one or more of the hydratase and dehydrogenase enzymes of a multifunctional enzyme include yeast (e.g., *Candida*, *Saccharomyces*, *Yarrowia*), animals (e.g., *Homo*, *Rattus*), bacteria. In a particular embodiment, one or more of the hydratase and dehydrogenase enzymes of a multifunctional enzyme can be a *Candida* yeast protein. Additional examples of nucleotide sequences encoding multifunctional enzyme polypeptides include: *Saccharomyces cerevisiae FOX2* (Genbank accession nos. NM_001179799, M86456), *Candida tropicalis* (strain PK 233) *HDE* (Genbank accession nos. X57854, M22765), *Yarrowia lipolytica MFE2* (Genbank accession no. AF198225).

Presence, absence or amount of one or more of the hydratase and dehydrogenase enzymes of a multifunctional enzyme can be detected by any suitable method known in the art and/or described herein. For example, detection can be performed using enzyme activity assays (see, e.g., Hiltunen et al. (1992) J. *Biol. Chem.* 267(10):6646-6653). Nucleic acid sequences representing native and/or modified multifunctional enzyme sequences also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or the amounts of the nucleic acids or encoded proteins can be assessed using quantitative expression based analysis (e.g., RT-PCR, western blot analysis,

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1446 of 3077

northern blot analysis, the like and combinations thereof), where the engineered cells or organisms exhibit increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism.

The promoter used for regulating transcription of a heterologous nucleic acid encoding one or more of the hydratase and dehydrogenase enzymes (for example, of a multifunctional enzyme) can also be modified. For example, the amount of one or more of the hydratase and dehydrogenase enzymes (for example, of a multifunctional enzyme protein) expressed in a particular cellular location may be increased or decreased by including in the heterologous nucleic acid a stronger or weaker heterologous promoter and/or a promoter that provides for a different pattern of expression in the cell or microorganism.

Alternatively, decreasing the activity of one or more of the hydratase and dehydrogenase enzymes (for example, of a multifunctional enzyme) in a cell can be accomplished by modifying the amount of expression of one or more of the hydratase and dehydrogenase enzymes (for example, of a multifunctional enzyme) in the cell, for example, by replacing the wild-type promoter of an endogenous multifunctional enzyme gene (or an enoyl-CoA hydratase or 3-hydroxyacyl-CoA dehydrogenase gene) in a cell or organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type multifunctional enzyme (or an enoyl-CoA hydratase or 3-hydroxyacyl-CoA dehydrogenase) such that the encoded modified or substituted protein has a reduced enzyme(s) activity.

Modification of 3-ketoacyl-CoA thiolase activity

In a final step of the β-oxidation pathway, 3-ketoacyl-CoA can undergo thiolytic cleavage to yield a fatty acyl-CoA shortened by 2 carbons and acetyl-CoA. The reaction can be catalyzed by 3-ketoacyl-CoA thiolase (e.g., EC 2.3.1.16; also referred to as β-ketothiolase, acetyl-CoA acyltransferase) and involves cleavage of the 3-ketoacyl-CoA by the thiol group of another molecule of CoA. The thiol is inserted between C-2 and C-3, which yields an acetyl CoA molecule and an acyl CoA molecule that is two carbons shorter.

In some embodiments of the cells, organisms, compositions and methods provided herein, the amount and/or activity of a 3-ketoacyl-CoA thiolase in a cell is modified. For example, in some

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1447 of 3077

aspects, a cell or microorganism may be modified to increase the amount and/or activity of a 3-ketoacyl-CoA thiolase, may be modified to decrease the amount and/or activity of a 3-ketoacyl-CoA thiolase, or may be modified to alternately increase and decrease the amount and/or activity of a 3-ketoacyl-CoA thiolase depending, for example, on the substrate specificity, cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

In certain aspects, the amount and/or activity of a 3-ketoacyl-CoA thiolase in a cell is increased. Increasing the amount and/or activity of a 3-ketoacyl-CoA thiolase may be particularly beneficial in embodiments in which the flux of carbons from fatty acids is directed toward a particular target product pathway involving oxidative metabolism and away from other cellular metabolic pathways not involved in target molecule production.

In certain embodiments, the 3-ketoacyl-CoA thiolase activity is unchanged in a host or engineered cell or organism. In one embodiment, the amount and/or activity of a host 3-ketoacyl-CoA thiolase can be increased, for example, by increasing the number of copies of a nucleic acid encoding a 3-ketoacyl-CoA thiolase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the nucleic acid), by increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a 3-ketoacyl-CoA thiolase, or by increasing the number of copies of a nucleic acid encoding a 3-ketoacyl-CoA thiolase and increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a 3-ketoacyl-CoA thiolase. In some embodiments, a 3-ketoacyl-CoA thiolase is endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, the amount of a 3-ketoacyl-CoA thiolase protein expressed in a cell can be increased by introducing heterologous nucleic acid encoding a 3-ketoacyl-CoA thiolase into a cell or microorganism. For example, introduction of heterologous nucleic acid encoding a 3-ketoacyl-CoA thiolase can result in increased copy number of such nucleic acids and/or provide for modification of the cellular location in which the protein is expressed.

Non-limiting examples of organisms that include, or can be used as donors for, a 3-ketoacyl-CoA thiolase enzyme include yeast (e.g., *Candida*, *Saccharomyces*, *Debaryomyces*, *Meyerozyma*, *Lodderomyces*, *Scheffersomyces*, *Clavispora*, *Yarrowia*, *Pichia*, *Kluyveromyces*, *Eremothecium*, *Zygosaccharomyces*, *Lachancea*, *Nakaseomyces*), animals (e.g., *Homo*, *Rattus*), bacteria (e.g., *Escherichia*, *Pseudomonas*, *Bacillus*), or plants (e.g., *Arabidopsis*,

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1448 of 3077

Nictotania, *Cuphea*). In a particular embodiment, a 3-ketoacyl-CoA thiolase enzyme can be a *Candida* yeast protein. Examples of nucleotide sequences encoding polypeptides having 3-ketoacyl-CoA thiolase activity include, but are not limited to: *Saccharomyces cerevisiae FOX1* (Genbank accession no. NM_001179508), *Candida tenuis* (Genbank accession no. XM_006688917), *Candida tropicalis CT-T3A* (Genbank accession no. AB025647), *Candida tropicalis CT-T3B* (Genbank accession no. AB025648), *Yarrowia lipolytica POT1* (Genbank accession no. XM_504109, X69988), *Scheffersomyces stipitis POT11* (Genbank accession no. XM_001386372), *Debaryomyces fabyri* (Genbank accession no. XM_015611011), *Arabidopsis thaliana KAT2* (Genbank accession no. NM_128874), *Lillium* cultivar Belladonna (Genbank accession no. KR998331) and *Populus davidianna KAT* (Genbank accession no. KU297273).

Presence, absence or amount of 3-ketoacyl-CoA thiolase activity can be detected by any suitable method known in the art and/or described herein. For example, detection can be performed using enzyme activity assays (see, e.g., Staack et al. (1978) *J. Biol. Chem* 253: 1827-1831; Kurihara et al. (1988) *FEBS Lett.* 229(1):215-218). Nucleic acid sequences representing native and/or modified 3-ketoacyl-CoA thiolase-encoding sequences also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or the amounts of the nucleic acids or encoded proteins can be assessed using quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cells or organisms exhibit increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism.

The promoter used for regulating transcription of a heterologous nucleic acid encoding a 3-ketoacyl-CoA thiolase can also be modified. For example, the amount of a 3-ketoacyl-CoA thiolase protein expressed in a particular cellular location may be increased by including in the heterologous nucleic acid a strong heterologous promoter and/or a promoter that provides for a different pattern of expression in the cell or microorganism. An example of one such heterologous promoter is a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene promoter. The nucleotide sequence of a *Candida viswanathii* HDE gene promoter is provided herein as are examples of additional fatty acid-inducible promoters.

Alternatively, decreasing 3-ketoacyl-CoA thiolase activity in a cell can be accomplished by modifying the amount of 3-ketoacyl-CoA thiolase protein expression in the cell, for example, by

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1449 of 3077

replacing the wild-type promoter of an endogenous 3-ketoacyl-CoA thiolase gene in an organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type 3-ketoacyl-CoA thiolase such that the encoded modified or substituted 3-ketoacyl-CoA thiolase protein has a reduced enzyme activity.

Modification of enoyl-CoA isomerase activity

Feedstocks, such as, for example, fatty acid distillates and soapstocks, can comprise unsaturated fatty acids, for example, such as oleic acid (C18:1), linoleic acid (C18:2), and linolenic acid (C18:3). In some embodiments, unsaturated fatty acids are converted to dicarboxylic acids that maintain the position and orientation of the double bonds. Unsaturated fatty acids generally are degraded through the same reactions that degrade saturated fatty acids until a Δ3-cis-acyl-CoA or Δ2-cis-acyl-CoA is formed in the process of β-oxidation. Cells can employ additional enzymes to allow the oxidation of these types of unsaturated fatty acids or diacids. In some instances, an enzyme enoyl-CoA isomerase (ECI) is required for the beta-oxidation of substrates with double bonds at odd numbered positions. In some instances, the enzyme dienoyl-CoA reductase (DCR) is required for the beta-oxidation of substrates with double bonds at even numbered positions.

Enoyl-CoA isomerase (ECI) can also be known as enoyl-CoA delta isomerase 1, dodecenoyl-CoA isomerase, 3,2 trans-enoyl-CoA isomerase, acetylene-allene isomerase, Δ3 Δ2-enoyl-CoA isomerase, dodecenoyl-CoA delta isomerase, and EC 5.3.3.8 (in human for example). Several alternatively spliced transcript variants are also known. ECI is a member of the hydratase/isomerase superfamily. ECI can be a key mitochondrial enzyme involved in beta-oxidation of unsaturated fatty acids. This enzyme can isomerize both 3-cis and 3-trans double bonds into the 2-trans form in a range of ECI enzymes from different species. ECI can catalyze the transformation of 3-cis and 3-trans-enoyl-CoA esters arising during stepwise degradation of cis-, mono-, and polyunsaturated fatty acids to the 2-trans-enoyl-CoA intermediates. ECI is present in many microorganisms and several species of yeast have at least two ECI enzymes. Nucleotide sequences (and corresponding amino acid sequences) encoding enoyl-CoA isomerase enzymes from *Candida* strain ATCC 20336 are provided herein (nucleotide SEQ ID NOS: 106 and 107 and animo acid SEQ ID NOS: 50 and 51). Examples of nucleotide sequences encoding polypeptides having enoyl-CoA isomerase activity include, but are not

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1450 of 3077

limited to: *Saccharomyces cerevisiae ECI1* (Genbank accession no. AF090442) and *Candida albicans* (Genbank accession no. XM_711189).

In some embodiments, ECI is utilized in generating a target fatty acid product through β-oxidation of an unsaturated fatty because of its activity and the normal position of double bonds in some feedstocks (e.g., soapstocks and fatty acid distillates). Many unsaturated fatty acids have a cis-Δ9 double bond. During the β -oxidation of an 18-carbon diacid with a cis-Δ9 double bond, the double bond is encountered when it has been chain shortened to 12 carbons. At this stage the 12-carbon molecule can have a cis-Δ3 double bond that is not a substrate for an acyl-CoA oxidase. ECI can convert the cis-Δ3 double bond to a trans-Δ2 double bond. In some instances, the product of the ECI reaction is a substrate for the second step in beta-oxidation (e.g., a substrate for enoyl-CoA hydratase), and ECI can effectively bypass acyl-CoA oxidase in a particular round of beta-oxidation. In some instances, even if a yeast strain lacks any acyl-CoA oxidase that is active on fatty acids of less than or equal to C12 (i.e., 12 carbons), an active ECI can effect the shortening of one more rounds of β-oxidation, which can produce a 10-carbon product for substrates with a cis-Δ9 double bond. Therefore, in some embodiments, the ECI gene is disrupted (e.g., knocked out or deleted) in a yeast (e.g., in a *Candida* strain) to prevent chain shortening past a desired chain-length (e.g., in this instance, 12 carbons). In some embodiments, disrupting the expression (e.g. knocking out the expression) of an ECI gene can result in an increase in the production of a fatty dicarboxylic acid containing 10 to 18 carbons. In some embodiments, disrupting the expression (e.g. knocking out the expression) of an ECI gene can result in an increase in the production of a fatty dicarboxylic acid containing 10, 12, 14, 16 or 18 carbons. In some embodiments, disrupting the expression of an enoyl-CoA isomerase can increase the production of fatty dicarboxylic acid containing 10, 12, 14, 16 or 18 carbons when using certain feedstocks (e.g., certain soapstocks or fatty acid distillates).

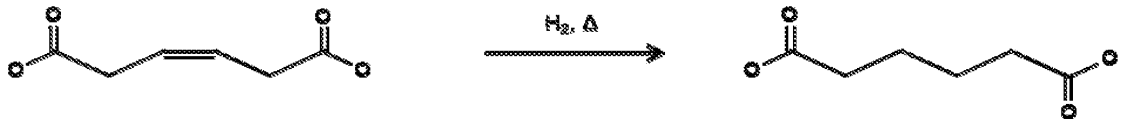
In some embodiments, an ECI knock out (i.e., eciΔ or Eci^{-}) strain is able to produce DDDA from from fatty acid feedstocks containing unsaturated fatty acids (e.g., oleic acid, linoleic acid, linolenic acid) even in the presence of acyl-CoA oxidase with activity on substrates of chain-length less than 12 carbons (but with little or nor activity on substrates having 12 carbons in the chain). This can be accomplished, for example, by discontinuation of β-oxidation after obtaining 3-dodecendioic acid (e.g., from oleic acid feedstock), 3,6-dodecenedioc acid (e.g., from linoleic acid feedstock) or 3,6,9-dodecenedioc acid (e.g., from linolenic acid feedstock) through an initial

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1451 of 3077

three rounds of β-oxidation (due to the lack of enoyl-CoA isomerase activity), and then hydrogenation of the dodecendioic acids to yield a fully saturated DDDA.

Thus, in some embodiments, a 12-carbon dicarboxylic acid produced from fatty acid feedstocks containing unsaturated fatty acids can be hydrogenated to generate a the fully saturated DDDA product. An unsaturated diacid sometimes is produced from a feedstock containing an unsaturated fatty acid, and production of a fully saturated diacid in such situations can involve hydrogenation of the unsaturated diacid. For example, an unsaturated C6:1 diacid generated from one or more long chain unsaturated fatty acids in an Eci$^-$ yeast strain which also lacks acyl-CoA oxidase enzymes having activity on substrates of 6-carbon chain lengths (e.g., a pox4Δ yeast strain) can be converted to a fully saturated C6:0 diacid by reducing the double bond by a suitable method. Non-limiting examples of hydrogenation methods include the use of a metallic chemical catalyst, non-metallic chemical catalyst, enzymatic catalyst, the like or combination thereof.

A non-limiting example of a hydrogenation reaction is shown below. Sometimes source hydrogen is provided from molecular hydrogen (e.g., in the case of chemical catalysis) and sometimes source hydrogen is provided from enzymatic cofactors, non-limiting examples of which include NADH, NADPH, FADH2, the like or combination thereof (e.g., in the case of enzymatic catalysis).



In some embodiments, catalytic hydrogenation is carried out with a suitable metallic catalyst, non-limiting examples of which include platinum, palladium, rhodium, ruthenium, nickel, the like or combination thereof. Sometimes a catalyst is a homogenous catalyst and sometimes a catalyst is a heterogeneous catalyst. An elevated temperature and/or pressure can be employed to increase reaction rate. For example, an unsaturated diacid (e.g., cis, cis-muconic acid) can be hydrogenated and converted to adipic acid using a 10% Pt on carbon catalyst at 3400 kPa for 2.5 hours at ambient temperature (Niu et al., (2002) *Biotechnol.Prog.* 18:201-211). In some embodiments, catalytic hydrogenation can occur with nonmetallic catalysts such as frustrated Lewis pair compounds (Welch et al., (2006) Science 314:1124-1126).

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1452 of 3077

In certain embodiments, enzymatic hydrogenation is conducted in vivo or in vitro with a suitable native or engineered enzyme that can catalyze a redox reaction with an unsaturated diacid or fatty acid as a substrate or a product. An enzyme can be utilized in vivo in some embodiments by increasing expression of a native enzyme or expressing a non-native enzyme capable of catalyzing a desired hydrogenation reaction in an organism that produces an unsaturated diacid precursor of a saturated diacid product. A lysate of an organism containing an enzyme capable of catalyzing a desired hydrogenation reaction, or a purified or isolated enzyme preparation, sometimes is utilized in an in vitro reaction. Non-limiting examples of a suitable native or engineered enzyme include acyl-CoA dehydrogenase (EC 1.3.1.8), trans-2-enoyl-CoA reductase (EC 1.3.1.44), stearoyl-CoA 9-desaturase (EC 1.14.19.1), the like or combination thereof. In some embodiments, a desired reaction product (e.g., a saturated diacid) is produced by an enzyme operating in a forward or a reverse direction (e.g., a forward or reverse reaction).

Modification of dienoyl-CoA reductase activity

Dienoyl CoA reductase (DCR, e.g., EC 1.3.1.34) is a peripheral enzyme that can convert trans-2, cis-4 dienoyl-CoA substrates to trans-3-enoyl-CoA products (Gurvitz A, et al., (1997) *J.Biol. Chem.* 272:22140-22147).

The trans-3-enoyl-CoA is then converted by the enzyme enoyl-CoA isomerase (ECI) to trans-2-enoyl-CoA which is then the substrate for the second enzyme (enoyl-CoA hydratase) in β-oxidation. Complete β-oxidation of fatty acids, including diacids, with double bonds at even numbered positions (e.g., linoleic acid (C18:2) and linolenic acid (C18:3)) can be achieved by including the DCR reaction in the β-oxidation pathway. Diacids are capable of being oxidized starting from either end (diterminal β-oxidation), and sometimes the enzymes used to rearrange and degrade the double bonds are the same from either direction. This is because even-numbered diacids with double bonds at even-numbered positions maintain the even-numbered position from either end (similarly with double bonds at odd-numbered positions).

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1453 of 3077

The DCR reaction can be used for complete β-oxidation of fatty acids with double bonds at even numbered positions, such as linoleic acid and linolenic acid. Depending on the carbon chain length of a desired final diacid product, it may be useful to either amplify or reduce the activity of one or more DCR enzymes in the host cell or engineered organism. For diacid products that have a carbon chain length of eight or greater, it may be useful or desirable to reduce or eliminate one or all DCR enzymes in the host cell. For diacid products that have a carbon chain length of less than eight carbons, it may be useful or desirable to amplify the activity of one or more DCR enzymes in the host cell or engineered organism.

Table 1 is a table of diacid products that may be produced from unsaturated fatty acids using a yeast strain in which the *ECI* and/or *DCR* genes have been disrupted or deleted. A Dcr$^-$ strain that does not include mutations of other genes encoding enzyme activities of the β-oxidation pathway typically can produce exclusively a C8:3 diacid. In some embodiments, a DCR polypeptide is not decreased, such as by disrupting a Dcr-encoding polynucleotide, in a strain utilized to produce a diacid product (e.g., adipic acid sebacic acid, DDDA). In certain embodiments, Dcr polypeptide production is increased (e.g., introducing additional copy numbers of an endogenous Dcr-encoding polynucleotide; introducing one or more copies of a heterologous Dcr-encoding polynucleotide) to produce adipic acid from polyunsaturated fatty acids, such as those prevalent in soybean or corn oil.

TABLE 1

	Diacids Produced in a Pox4$^+$, Pox5$^+$ Background		
Fatty Acid Carbon Source	**Eci$^-$ Product**	**Dcr$^-$ Product**	**Eci$^-$,Dcr$^-$ Product**
Oleic acid (C18:1)	C6:1 (3-hexenedioic acid)	None	C6:1 (3-hexenedioic acid)
Linoleic acid (C18:2)	C10:2 (3,7-decenedioc acid)	C8:3 (2,4,6-octenedioic acid)	C10:3 (2,4,7-decenedioic acid)
Linolenic acid (C18:3)	C12:3 (3,6,9-dodecenedioc acid)	C8:3 (2,4,6-octenedioic acid)	C12:3 (3,6,9-dodecenedioic acid)
Eicosenoic acid (C20:1)	C6:1 (3-hexenedioic acid)	None	C6:1 (3-hexenedioic acid)
Erucic acid (C22:1)	C6:1 (3-hexenedioic acid)	None	C6:1 (3-hexenedioic acid)

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1454 of 3077

In yeast such as *Candida tropicalis* and *Candida viswanathii*, there are two *DCR* homologs, often referred to as *DCR1* and *DCR2*. The yeast *Saccharomyces cerevisiae* includes one Dcr enzyme, while the yeast *Yarrowia lipolytica* includes at least three *DCR* homologs, referred to herein as "*DCR1*", "*DCR2*", and "*DCR3*". Nucleotide sequences (and corresponding amino acid sequences) encoding dienoyl-CoA reductase enzymes from *Candida* strain ATCC 20336 are provided herein (nucleotide SEQ ID NOS: 108 and 109 and amino acid SEQ ID NOS: 52 and 53). Examples of nucleotide sequences encoding polypeptides having dienoyl-CoA reductase activity include, but are not limited to: *Saccharomyces cerevisiae SPS19* (Genbank accession no. NM_001183040), *Candida tropicalis SPS19* (Genbank accession no. XM_002545237) and *Yarrowia lipolytica* (Genbank accession nos. XM_501382, XM_503937, XM_502296).

Accordingly, there are multiple possible genotypes of yeast strains (having varying combinations of wild-type and mutant acyl-CoA oxidase activity specificities and functional or non-functional ECI and/or DCR genes) for the production of fatty acids and diacids of differing carbon chain lengths and degrees of saturation/unsaturation. The fatty acid or diacid produced can depend on the carbon source in the feedstock. Table 2 provides non-limiting examples of some of the yeast strain (e.g., *Candida* spp.) genotype combinations and carbon sources for the production of adipic acid, suberic acid, sebacic acid and DDDA.

TABLE 2	**STRAIN GENOTYPE (with respect to *POX4/ECI/DCR*) AND CARBON SOURCE COMBINATIONS**		
DIACID PRODUCT	**OLEIC ACID**	**LINOLEIC ACID**	**LINOLENIC ACID**
Adipic acid (C6)	*pox4Δ/ECI/DCR* (yields saturated diacid)	*pox4Δ/ECI/DCR* (yields saturated diacid)	*pox4Δ/ECI/DCR* (yields saturated diacid*)
	POX4/Eci⁻/DCR (yields 3-hexenedioic acid*)		
Suberic acid (C8)	*pox4Δ/ECI/DCR* (yields saturated diacid)	*pox4Δ/ECI/DCR* (yields saturated diacid)	
	pox4Δ/ Eci⁻/DCR (yields 3-octenedioic acid*)	*POX4/ECI/Dcr⁻* (yields 2,4,6-octenedioic acid*)	*POX4/ECI/Dcr⁻* (yields 2,4,6-octenedioic acid*)

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1455 of 3077

TABLE 2	STRAIN GENOTYPE (with respect to *POX4/ECI/DCR*) AND CARBON SOURCE COMBINATIONS		
DIACID PRODUCT	**OLEIC ACID**	**LINOLEIC ACID**	**LINOLENIC ACID**
		POX4/Eci⁻/DCR (yields 3,6-octenedioic acid*)	
Sebacic acid (C10)		*POX4/Eci⁻/DCR* or *pox4Δ/ Eci⁻/DCR* (yields 3,7-decenedioic acid*)	
		POX4/Eci⁻/ Dcr⁻ (yields 2,4,7-decenedioic acid*)	
Dodecanedioic acid (C12)			*POX4/Eci⁻/DCR* or *pox4Δ/ Eci⁻/DCR* (yields 3,6,9-dodecenedioic acid*)

* can be hydrogenated to saturate

Modification of dienoyl-CoA isomerase activity

Dienoyl-CoA isomerase (DCI, e.g., EC 5.3.3, Δ3,5,Δ2,4-dienoyl-CoA isomerase, Δ3,5,Δ2,4-dienoyl-coenzyme A isomerase) is a peripheral β-oxidation enzyme that catalyzes the isomerization of a Δ3,5-dienoyl-CoA to a Δ2,4-dienoyl-CoA. This reaction is part of a minor β-oxidation pathway that occurs when the 3,2-enoyl-CoA isomerase (ECI) converts a Δ2,5-dienoyl-CoA to a Δ3,5-dienoyl-CoA. In order to fully oxidize this product DCI converts the Δ3,5-dienoyl-CoA to a Δ2,4-dienoyl-CoA, the latter of which is a substrate for the 2,4-dienoyl-CoA reductase (DCR). The product of the DCR reaction is a 3-enoyl-CoA, which is a substrate for ECI that converts it to a 2-enoyl-CoA that can be fully oxidized through β-oxidation.

In some embodiments, the amount and/or activity of a Dci enzyne in a cell or organism is decreased or increased, depending upon the chain-length of a desired target diacid product to be generated through β-oxidation of a fatty acid. For example, for adipic acid production, a DCI

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1456 of 3077

activity can be increased to improve productivity of unsaturated fatty acids in a host cell or organism (e.g., DCI activity can be increased by introducing one or more copies of a polynucleotide encoding a polypeptide having DCI activity into the cell or organism (e.g., introducing one or more copies of an endogenous or exogenous polynucleotide)). In some embodiments, for production of C8 and longer diacids, the amount and/or activity of a DCI enzyme in a cell or organism can be decreased (e.g., by introducing a disruption, deletion or knockout of a polynucleotide that encodes a polypeptide having DCI activity, or replacing a promoter of a DCI gene with a weaker promoter (for example, introducing a nucleic acid containing a weak promoter operably linked to a polynucleotide that encodes a polypeptide having DCI activity into a cell in which an endogenous DCI gene has been disrupted or deleted). An example of a nucleotide sequence encoding a *Saccharomyces cerevisiae* DCI enzyme is Genbank accession no. NM_001183599.

Modification of β-oxidation-associated activities

There are also cellular compositions and activities that are closely associated with β-oxidation and support the core degradative functioning of the pathway. These include peroxisomal- and mitochondrial-related compositions and activities. For example, as described herein, such compositions and activities include, but are not limited to, compositions and activities involved in: generating acyl-CoA through thioesterification of fatty acids, movement of fatty acids and/or acyl-CoA into cellular sites of β-oxidation (e.g., peroxisomes), regulation of β-oxidation activities, synthesis of compositions involved in β-oxidation, and maintenance/amount of sites of β-oxidation (e.g., peroxisomes). Included in the cells, organisms, systems and methods provided herein are embodiments in which one or more of these β-oxidation-associated compositions and/or activities are modified. In some embodiments, a β-oxidation-associated composition or activity is modified to enhance β-oxidation activity.

Modification of peroxisomal transport activity

In order for fatty acids to undergo peroxisomal β-oxidation, they must first move into the peroxisomes. Generally, medium-chain free fatty acids present in the cytosol can traverse the peroxisomal membrane and become activated once in the peroxisome by a persoxisomal acyl-CoA synthetase to then be processed as an acyl-CoA in β-oxidation. Long-chain fatty acids that have entered a cell from the extracellular medium tend to quickly be activated by acyl-CoA

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1457 of 3077

synthetases located at or near the cell membrane or in the cytosol. These acyl-CoA esters typically are not able to traverse the peroxisomal membrane and thus require a peroxisomal transporter in order to move into peroxisomes. Peroxisomal transporter proteins can be a target for modifying entry of fatty acids into peroxisomes. Free fatty acids internalized into cells, or generated within cells (e.g., by oxidation of internalized alkanes), can directly enter into and be processed in the ω-oxidation pathway without prior activation to acyl-CoA.

In some embodiments of the microorganisms, compositions and methods provided herein, the amount and/or activity of a peroxisomal transporter protein in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase the amount of a peroxisomal transporter protein and/or a peroxisomal transporter protein activity, may be modified to decrease the amount of a peroxisomal transporter protein and/or a peroxisomal transporter protein activity, or may be modified to alternately increase and decrease the amount of a peroxisomal transporter protein and/or a peroxisomal transporter protein activity depending, for example, on the cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

In certain aspects, the amount and/or activity of a peroxisomal transporter protein in a cell is decreased. Reducing or eliminating the amount and/or activity of a peroxisomal transporter protein may be particularly beneficial in embodiments in which the flux of carbons from fatty acids is directed toward a particular target product pathway (e.g., in peroxisomes) and away from other cellular metabolic pathways involving activated fatty acids (acyl-CoA). For example, in embodiments of the production systems in which a target molecule, or intermediate/precursor of a target molecule, is a dicarboxylic acid, it may be optimal to decrease or eliminate fatty acid entry into peroxisomes through modes other than as a dicarboxylic acid which moves freely into peroxisomes after formation though an initial ω-oxidation of a free fatty acid. Certain aspects of the cells, microorganisms, compositions and methods provided herein include one or more modifications to reduce or eliminate transport of acyl-CoA into peroxisomes. One approach to reducing or eliminating such transport is to decrease the amount and/or activity of a peroxisomal transporter protein. For example, one or more endogenous genes encoding a peroxisomal transporter protein (e.g., yeast *PXA1* and/or *PXA2*) can be disrupted or deleted in a host cell or microorganism to reduce or eliminate the amount of and/or activity of a peroxisomal transporter protein in the host relative to a cell or microorganism in which the gene(s) have not been modified.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1458 of 3077

Methods for decreasing the amount and/or activity of a peroxisomal transporter protein in a cell include, but are not limited to, modifying the amount of peroxisomal transporter protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous a peroxisomal transporter protein gene in an organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type peroxisomal transporter protein such that the encoded modified or substituted peroxisomal transporter protein has a reduced activity. For example, expression of a host peroxisomal transporter protein activity can be decreased or eliminated by disruption (e.g., knockout, insertion mutagenesis, the like and combinations thereof) of a host gene encoding the protein, or by decreasing the activity of the promoter (e.g., through addition of repressor sequences to the promoter or 5'UTR or replacing the promoter) that controls transcription of a peroxisomal transporter protein gene using recombinant molecular biology techniques known in the art and/or described herein. In one embodiment, a diploid yeast, such as, for example, a *Candida* yeast, when used as a host microorganism can be subjected to genetic modification in which one of the two alleles of a peroxisomal transporter protein gene is disrupted or deleted. In so doing, a single allele of the gene remains for a reduced amount of peroxisomal transporter protein expression in the microorganism and a reduced amount of the protein in the cell.

One method for disrupting an endogenous peroxisomal transporter protein gene is by recombinantly inserting a heterologous nucleic acid (e.g., a nucleotide sequence encoding a selectable marker such as an enzyme that restores an auxotrophic host organism to prototrophy) into the endogenous gene, thereby generating an engineered organism deficient in a peroxisomal transporter protein activity. This can be done, for example, through homologous recombination in which a heterologous nucleic acid containing sequences of an endogenous peroxisomal transporter protein gene and a disrupting sequence (e.g., a knock-out gene cassette such as described herein) is introduced into a host cell or microorganism. Nucleic acids encoding a peroxisomal transporter protein can be obtained from a number of sources, including, for example, yeast cells. Genomic DNA from cell sources can be amplified using oligonucleotide primers based on the nucleotide sequence of a peroxisomal transporter protein encoding gene, including examples provided herein. Nucleotide sequences encoding the subunits of (and the amino acid sequences of) a *Candida viswanathii* peroxisomal transporter protein, Pxa1 and Pxa2, are provided herein (nucleotide SEQ ID NOS: 94 and 95 and amino acid SEQ ID NOS: 40 and 41). Additional non-limiting examples of nucleic acids encoding a

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1459 of 3077

peroxisomal transporter protein include *Saccharomyces cerevisiae PXA1* (Genbank accession numbers NM_001183961 and U17065), *Saccharomyces cerevisiae PXA2* (Genbank accession numbers NM_001179754 and U93584), *Schizosaccharomyces pombe PXA1* (Genbank accession number NM_001018794), *Candida albicans PXA1* (Genbank accession number XM_713564), *Yarrowia lypolytica PXA1* (Genbank accession number XM_499814), *Yarrowia lypolytica PXA2* (Genbank accession number XM_502396), *Candida orthopsilosis PXA1* (Genbank accession number XM_003865834), *Aspergillus nomius PXA1* (Genbank accession number XM_015554863), *Clavispora lusitaniae PXA1* (Genbank accession number JQ710938), *Aspergillus niger PXA1* (Genbank accession number XM_001388761) and *Arabidopsis thalina ABCD1* (Genbank accession number NM_001204043).

In other embodiments, the amount and/or activity of a host peroxisomal transporter protein can be increased, for example, by increasing the number of copies of a gene encoding a peroxisomal transporter protein (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the gene), by increasing the activity of a promoter that regulates transcription of a gene encoding a peroxisomal transporter protein, or by increasing the number of copies of a gene encoding a peroxisomal transporter protein and increasing the activity of a promoter that regulates transcription of a gene encoding a peroxisomal transporter protein. In some embodiments, a peroxisomal transporter protein is endogenous to the host cell or microorganism. In particular embodiments, the amount and/or activity of a host peroxisomal transporter protein is increased.

The presence, absence or amount of peroxisomal transporter protein can be detected by any suitable method known in the art. Non-limiting examples of suitable detection methods include nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cells or organisms exhibit increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism.

Modification of peroxisome biogenesis activity

Peroxisomes can be found in eukaryotic cells and are a cellular location for β-oxidation (i.e., the site for β-oxidation in fungi and plant cells and one of two sites (the other being mitochondria) for β-oxidation of fatty acids in animal cells). Consistent with this function, peroxisome

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1460 of 3077

proliferation may occur in cells exposed to fatty acids as a sole source of carbon, and peroxisome degradation may occur in cells in the presence of glucose. Thus, the number of and volume of peroxisomes in cells can be regulated. Although most of the more than 30 peroxisomal membrane proteins, referred to as peroxins or Pex proteins, play a role in importing proteins into the peroxisomal matrix from the cytosol (e.g., Pex5, Pex7, Pex13, Pex14 Pex16, Pex17), some (e.g. Pex 11, Pex 25, Pex 27, Pex 34) are involved in peroxisome proliferation.

In some embodiments of the cells, organisms, compositions and methods provided herein, the amount and/or activity of a Pex protein in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase the amount and/or activity of a Pex protein, to decrease the amount and/or activity of a Pex protein, or to alter the pattern of expression of a Pex protein. In particular embodiments, the Pex protein is one that is involved in peroxisome proliferation, e.g., Pex11.

In certain aspects, the amount and/or activity of a Pex protein in a cell or organism is increased. Increasing the amount and/or activity of a Pex protein may be particularly beneficial in embodiments in which the flux of carbons from fatty acids is directed toward a particular target product pathway involving oxidative metabolism and away from other cellular metabolic pathways not involved in target molecule production. In some embodiments, the amount and/or activity of a Pex protein involved in peroxisome proliferation is increased in a cell or organism to provide for increased numbers of peroxisomes as sites for β-oxidation. In particular embodiments, the Pex protein is Pex11.

In certain embodiments, the Pex protein activity is unchanged in a host or engineered cell or organism. In one embodiment, the amount and/or activity of a host Pex protein can be increased, for example, by increasing the number of copies of a nucleic acid encoding a Pex protein (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the nucleic acid), by increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a Pex protein, or by increasing the number of copies of a nucleic acid encoding a Pex protein and increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a Pex protein. In some embodiments, a Pex protein is endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, the amount of a Pex protein expressed in a cell can be increased by introducing heterologous nucleic acid encoding a Pex protein into a cell or microorganism. For example, introduction of heterologous

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1461 of 3077

nucleic acid encoding a Pex protein can result in increased copy number of such nucleic acids and/or provide for modification of the cellular location in which the protein is expressed.

Non-limiting examples of organisms that include, or can be used as donors for, a Pex protein include yeast (e.g., *Candida*, *Saccharomyces*, *Debaryomyces*, *Meyerozyma*, *Lodderomyces*, *Scheffersomyces*, *Clavispora*, *Yarrowia*, *Pichia*, *Kluyveromyces*, *Eremothecium*, *Zygosaccharomyces*, *Lachancea*, *Nakaseomyces*), animals (e.g., *Homo*, *Rattus*), or plants (e.g., *Arabidopsis*, *Nictotania*, *Cuphea*). In a particular embodiment, a Pex protein can be a *Candida* yeast protein. An example of a *Candida viswanathii* nucleotide sequence (and corresponding amino acid sequence) encoding a Pex11 protein is provided herein (nucleotide SEQ ID NO: 89 and amino acid SEQ ID NO: 33). Additional examples of nucleotide sequences encoding polypeptides having Pex protein activity include, but are not limited to: *Saccharomyces cerevisiae PEX11* (Genbank accession no. NM_001183401), *Candida albicans* (Genbank accession no. XM_707009), *Candida orthopsilosiis PEX11* (Genbank accession no. XM_003870517), *Yarrowia lipolytica PEX11* (Genbank accession nos. XM_503276, XM_501447, XM_501425), *Arabidopsis thaliana PEX11A* (Genbank accession no. NM_103668), *Neurospora crassa PEX11* (Genbank accession no. XM_011396615), *Pichia angusta PEX11* (Genbank accession no. DQ645582).

Presence, absence or amount of Pex protein or nucleic acids encoding the protein can be detected by any suitable method known in the art and/or described herein. For example, detection can be performed using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cells or organisms exhibit increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism.

The promoter used for regulating transcription of a heterologous nucleic acid encoding a Pex protein can also be modified. For example, the amount of a Pex protein expressed in a particular cellular location may be increased by including in the heterologous nucleic acid a strong heterologous promoter and/or a promoter that provides for a different pattern of expression in the cell or microorganism. An example of one such heterologous promoter is a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene promoter. The nucleotide sequence

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1462 of 3077

of a *Candida viswanathii* HDE gene promoter is provided herein as are examples of additional fatty acid-inducible promoters.

Modification of acetyl-CoA processing in peroxisomes

Included in the cell-based and microbial production platform systems and components thereof provided herein are embodiments in which the processing of acetyl-CoA in organelles of a cell or microorganism is modified. In some embodiments, acetyl group carbons of organelle-generated acetyl-CoA are directed toward conversion to acetate. In particular embodiments, acetyl group carbons are directed toward conversion to acetate and away from the carnitine-carrier transport system. Accordingly, provided herein are cells, microorganisms, compositions and methods in which cellular carbon flux has been modified through the altered (e.g., increased or decreased) de novo generation of cellular acetate. In particular embodiments, cellular carbon flux has been modified to increase the production of acetate in a cell and/or a particular cellular location. In certain aspects, cells or microorganisms are modified to increase the production of acetate in peroxisomes.

Modification of acetyl-CoA hydrolase activity

In some embodiments of the cells, microorganisms, compositions and methods provided herein, the amount and/or activity of acetyl-CoA hydrolase in a cell is modified. Acetyl-CoA hydrolase (e.g. EC 3.1.2.1) is an enzyme that catalyzes the hydrolysis of acetyl-CoA to form acetate and CoA. For example, in some aspects, a cell or microorganism may be modified to increase acetyl-CoA hydrolase and/or acetyl-CoA hydrolase activity, may be modified to decrease acetyl-CoA hydrolase and/or acetyl-CoA hydrolase activity, or may be modified to alternately increase and decrease acetyl-CoA hydrolase and/or acetyl-CoA hydrolase activity depending, for example, on the cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

In some aspects, the amount and/or activity of acetyl-CoA hydrolase in a cell and/or a particular cellular location is increased. For example, the amount and/or activity of acetyl-CoA hydrolase in peroxisomes of a cell can be increased. In some embodiments, the pattern of expression of acetyl-CoA hydrolase can be modified such that the enzyme is produced in a cellular location

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1463 of 3077

where it is not produced in an unmodified cell and/or is no longer produced in a cellular location where it is produced in an unmodified cell.

In one aspect, the amount and/or activity of a host acetyl-CoA hydrolase can be increased, for example, by increasing the number of copies of a nucleic acid encoding an acetyl-CoA hydrolase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the nucleic acid), by increasing the activity of a promoter that regulates transcription of a nucleic acid encoding an acetyl-CoA hydrolase, or by increasing the number of copies of a nucleic acid encoding an acetyl-CoA hydrolase and increasing the activity of a promoter that regulates transcription of a nucleic acid encoding an acetyl-CoA hydrolase. In some embodiments, an acetyl-CoA hydrolase is endogenous to the host cell or microorganism. In one aspect of the cell-based and microbial systems and methods provided herein, the amount of acetyl-CoA hydrolase protein expressed in a cell can be increased by introducing heterologous nucleic acid encoding acetyl-CoA hydrolase into a cell or microorganism. For example, introduction of heterologous nucleic acid encoding acetyl-CoA hydrolase can result in increased copy number of such nucleic acids and/or provide for modification of the cellular location in which the protein is expressed.

Acetyl-CoA hydrolase is typically localized to the mitochondrial compartment in eukaryotes. In one embodiment of the cells, microorganisms and methods provided herein, host cells are modified to express engineered acetyl-CoA hydrolase proteins that include targeting signals that direct the enzyme to peroxisomes, thereby introducing, or increasing the amount of, peroxisomal acetyl-CoA hydrolase in the cells. In particular embodiments, the engineered acetyl-CoA hydrolase protein has also been modified to exclude amino acids of a mitochondrial targeting sequence. One such modified acetyl-CoA hydrolase protein is a yeast Ach1p $^{\Delta mts+pts}$ which includes a heterologous peroxisomal targeting signal (pts) and excludes a mitochondrial targeting sequence (mts). In order to express engineered acetyl-CoA hydrolase in a targeted location, such as the peroxisomes, heterologous nucleic acid encoding the modified enzyme can be introduced into host cells. Acetate generated through the action of peroxisomal acetyl-CoA hydrolase can freely diffuse out of the peroxisome into the cytosol where it can be converted back to acetyl-CoA by the enzyme acetyl-CoA synthetase (e.g., EC 6.2.1.1), thereby increasing the generation and amount of cytosolic acetyl-CoA. In a particular embodiment, the acetyl-CoA hydrolase enzyme can be a *Candida* yeast protein. An example of a *Candida viswanathii* nucleotide sequence (*ACH1*$^{\Delta mts+pts}$; SEQ ID NO: 73) encoding a modified acetyl-CoA hydrolase lacking a mitochondrial-targeting sequence and including a peroxisomal-targeting

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1464 of 3077

sequence (Ach1p$^{\Delta mts+pts}$; SEQ ID NO: 16) is provided herein. Additional examples of nucleotide sequences encoding acetyl-CoA hydrolase proteins include but are not limited to: *Saccharomyces cerevisiae ACH1* (Genbank accession numbers M31036, NM_001178255), *Candida tropicalis ACH1* (Genbank accession number XM_002550976), *Candida orthopsilosis ACH1* (Genbank accession number XM_003870486), *Candida albicans ACH1* (Genbank accession number XM_709496), *Aspergillus flavus ACH1* (Genbank accession number XM_002372714), *Neurospora crassa ACU8* (Genbank accession number XM_953261), *Cyberlindnera jadinii ACH1* (Genbank accession number AB641818), *Debaryomyces fabryi ACH1* (Genbank accession number XM_015614474), *Schizosaccharomyces octosporus ACH1* (Genbank accession numberXM_013163018), *Schizosaccharomyces japonicus ACH1* (Genbank accession numberXM_002173925), *Penicillium digitatum ACH1* (Genbank accession number XM_014683672), *Penicillium marneffei ACH1* (Genbank accession number XM_002152968) and *Talaromyces stipitatis ACH1* (Genbank accession number XM_002487448). Any of these, and other such acetyl-CoA hydrolase-encoding nucleic acids, can be analyzed for the presence of 5' ORF nucleotides encoding possible mitochondrial-targeting sequences of amino acids and modified to eliminate such sequences. Nucleotides encoding a peroxisomal-targeting sequence (e.g, a PTS1 sequence such as AKL or SKL) can also be added to the 3' terminus of the coding sequences of the nucleic acids.

The promoter used for regulating transcription of a heterologous nucleic acid encoding an acetyl-CoA hydrolase can also be modified. For example, the amount of an acetyl-CoA hydrolase protein expressed in a particular cellular location may be increased by including in the heterologous nucleic acid a strong heterologous promoter and/or a promoter that provides for a different pattern of expression in the cell or microorganism. An example of one such heterologous promoter is a fatty acid-inducible promoter that can provide for increased acetyl-CoA hydrolase expression, particularly when exposed to fatty acids as a carbon source. Such promoter elements include those that regulate expression of peroxisomal proteins and/or β-oxidation enzymes in microbes, e.g., a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene promoter. The nucleotide sequence of a *Candida viswanathii* HDE gene promoter is provided herein as are examples of additional fatty acid-inducible promoters.

The acetyl-CoA hydrolase activities of host and modified cells and microorganisms can be evaluated and monitored using methods known in the art. Examples of acetyl-CoA hydrolase activity assays include colorimetric assays (see, e.g., Connerton et al. (1992) *J. Gen. Microbiol.*

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1465 of 3077

138:1797-1800; Robinson et al (1976) *Biochem. Biophys. Res. Commun.* 21:959-965) and radioactivity-based and acetylation inhibition assays (see, e.g., U.S. Patent no. 5,487,990 to Smith et al.). Nucleic acid sequences representing native and/or modified acetyl-CoA hydrolase-encoding sequences also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or the amounts of the nucleic acids or encoded proteins can be assessed using quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cells or organisms exhibit increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism.

Modification of peroxisomal carnitine acetyltransferase activity

Also provided herein are cells, microorganisms, compositions and methods in which the amount and/or activity of peroxisomal carnitine acetyltransferase in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase the amount of peroxisomal carnitine acetyltransferase and/or peroxisomal carnitine acetyltransferase activity, may be modified to decrease the amount of peroxisomal carnitine acetyltransferase and/or peroxisomal carnitine acetyltransferase activity, or may be modified to alternately increase and decrease the amount of peroxisomal carnitine acetyltransferase and/or peroxisomal carnitine acetyltransferase activity depending, for example, on the conditions in which the modified cell or microorganism is cultured.

In some aspects, the amount and/or activity of peroxisomal carnitine acetyltransferase in a cell is decreased. Reducing or eliminating the amount and/or activity of peroxisomal carnitine acetyltransferase may be particularly beneficial in embodiments in which the flux of peroxisomal acetyl moiety carbons is directed toward generation of acetate within peroxisomes. In these embodiments, reducing or eliminating the amount and/or activity of peroxisomal carnitine acetyltransferase decreases the amount of peroxisomal acetyl group carbon atoms that are converted to acetyl-carnitine and provides increased peroxisomal acetyl-CoA availability for generation of peroxisomal acetate. Methods for decreasing peroxisomal carnitine acetyltransferase activity in a cell include, but are not limited to, modifying the amount of peroxisomal carnitine acetyltransferase protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous peroxisomal carnitine acetyltransferase gene in an organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene,

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1466 of 3077

and/or replacing or modifying a gene encoding a wild-type peroxisomal carnitine acetyltransferase such that the encoded modified or substituted peroxisomal carnitine acetyltransferase protein has a reduced enzyme activity. For example, expression of a host peroxisomal carnitine acetyltransferase activity can be decreased by disruption (e.g., knockout, insertion mutagenesis, the like and combinations thereof) of a host gene encoding the protein, or by decreasing the activity of the promoter (e.g., through addition of repressor sequences to the promoter or 5'UTR or replacing the promoter) that controls transcription of a peroxisomal carnitine acetyltransferase gene using recombinant molecular biology techniques known in the art and described herein. In one embodiment, a diploid yeast, such as, for example, a *Candida* yeast, when used as a host microorganism can be subjected to genetic modification in which one of the two alleles of a peroxisomal carnitine acetyltransferase gene is disrupted or deleted. In so doing, a single allele of the gene remains for a reduced amount of peroxisomal carnitine acetyltransferase expression in the microorganism and a reduced amount of the protein in the cell.

One method for disrupting an endogenous peroxisomal carnitine acetyltransferase gene is by recombinantly inserting a heterologous nucleic acid (e.g., a nucleotide sequence encoding a selectable marker such as an enzyme that restores an auxotrophic host organism to prototrophy) into the endogenous gene, thereby generating an engineered organism deficient in peroxisomal carnitine acetyltransferase activity. This can be done, for example, through homologous recombination in which a heterologous nucleic acid containing sequences of the endogenous peroxisomal carnitine acetyltransferase gene and a disrupting sequence (e.g., a knock-out gene cassette such as described herein) is introduced into a host cell or microorganism. Nucleic acids encoding a peroxisomal carnitine acetyltransferase can be obtained from a number of sources, including, for example, yeast cells. Genomic DNA from cell sources can be amplified using oligonucleotide primers based on the nucleotide sequence of a peroxisomal carnitine acetyltransferase encoding gene, including examples provided herein.

In some instances, a host gene, e.g., certain yeast genes, encoding a peroxisomal carnitine acetyltransferase also encodes a mitochondrial carnitine acetyltransferase. In these organisms, a peroxisomal carnitine acetyltransferase is encoded by a gene that generates a protein containing mitochondrial and peroxisomal targeting sequences. Therefore, in such an instance, disruption or deletion of a gene encoding a carnitine acetyltransferase that is localized to peroxisomes will result in reducing or eliminating mitochondrial, as well as peroxisomal,

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1467 of 3077

carnitine acetyltransferase protein expression. In order to reduce or eliminate peroxisomal carnitine acetyltransferase expression in such cells without eliminating mitochondrial carnitine acetyltransferase expression, a heterologous nucleic acid encoding a mitochondria-targeted carnitine acetyltransferase can be introduced into the cell after disruption of the endogenous gene. For example, a mitochondrial-targeted enzyme that would not be expressed in peroxisomes can be produced in a cell or microorganism by introducing a heterologous nucleic acid that encodes a carnitine acetyltransferase that includes a mitochondrial targeting sequence of amino acids but lacks a peroxisomal targeting sequence of amino acids. An example of such a modified *Candida viswanathii* nucleic acid sequence (*CAT2*$^{\Delta pts}$; SEQ ID NO: 62), and the amino acid sequence encoded thereby (Cat2p$^{\Delta pts}$; SEQ ID NO: 5), are provided herein.

In another embodiment provided herein, a heterologous nucleic acid encoding a peroxisomal carnitine acetyltransferase that has a reduced carnitine acetyltransferase activity relative to the activity of the enzyme encoded by a host cell's or microorganism's endogenous peroxisomal carnitine acetyltransferase gene can be introduced into a host cell in which the endogenous peroxisomal carnitine acetyltransferase gene(s) has been disrupted or deleted. The heterologous nucleic acid encoding the less active carnitine acetyltransferase can be modified to include nucleotides encoding a peroxisomal targeting sequence for expression of the enzyme specifically in peroxisomes and not in other areas, such as mitochondria. For example, in one aspect, a heterologous nucleic acid encoding a *Candida viswanathii* cytoplasmic carnitine acetyltransferase (*YAT1*) with added nucleotides encoding a peroxisomal targeting sequence (e.g, a PTS1 sequence such as AKL or SKL, or slight variant thereof (PKL, PKF)) can be introduced into a host cell or microorganism (e.g., a *Candida viswanathii* cell). A nucleotide sequence encoding (and the amino acid sequence of) a *Candida viswanathii* YAT1^{+pts}p are provided herein (nucleotide SEQ ID NO: 70 and amino acid SEQ ID NO: 13). Additional non-limiting examples of nucleic acids encoding cytoplasmic carnitine acetyltransferase include *Saccharomyces cerevisiae YAT1* (Genbank accession number X74553), *Aspergillus nidulans FacC* (Genbank accession number AF023156), *Cyberlindnera jadinii YAT1* (Genbank accession number AB641829), *Candida dubliniensis YAT1* (Genbank accession number XM_002416790) and *Candida albicans* (Genbank accession number AF525683). A sequence of nucleotides encoding a peroxisomal targeting sequence can be added to the 3' end of the coding sequence of any such nucleic acid using methods as described herein.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1468 of 3077

The peroxisomal carnitine acetyltransferase activities of host and modified cells and microorganisms can be evaluated and monitored using methods known in the art. For example, methods of isolating peroxisomal and mitochondrial components of yeast cells and of extracting carnitine acetyltransferase from subcellular fractions have been described by Ueda et al. [(1982) *Eur. J. Biochem.*124:205-210] and Kozulic et al. [(1987) *Eur. J. Biochem.*168:245-250]. Methods of measuring the enzymatic activity of carnitine acetyltransferase are also known in the art, see, e.g., Fritz and Schultz (1965) *J. Biol. Chem*. 240:2188-2192; Chase (1969) *Meth. Enzymol.*13:387-393. Nucleic acid sequences representing native and/or modified peroxisomal carnitine acetyltransferase-encoding sequences also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or the amounts of the nucleic acids or encoded proteins can be assessed using quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cells or organisms exhibit increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism.

In other embodiments, the amount and/or activity of a host peroxisomal carnitine acetyltransferase can be increased, for example, by increasing the number of copies of a gene encoding a peroxisomal carnitine acetyltransferase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the gene), by increasing the activity of a promoter that regulates transcription of a gene encoding a peroxisomal carnitine acetyltransferase, or by increasing the number of copies of a gene encoding a peroxisomal carnitine acetyltransferase and increasing the activity of a promoter that regulates transcription of a gene encoding a peroxisomal carnitine acetyltransferase. In some embodiments, a peroxisomal carnitine acetyltransferase is endogenous to the host cell or microorganism.

Modification of acetyl-CoA synthetase

Acetyl-CoA synthetase (EC 6.2.1.1) is an enzyme that can catalyze the ligation of acetate and coenzyme A to produce acetyl-CoA. In many cells and organisms, the enzyme is encoded by one or more *ACS* genes. For example, in some yeast, acetyl-CoA synthetase is encoded by two genes, *ACS1* and *ACS2*, which may be differentially expressed in response to growth on differing carbon sources. In some cells, the proteins encoded by the two genes may also be differentially distributed within the nucleus, mitochondria, peroxisomes and cytoplasm of cells.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1469 of 3077

Acs1p and Acs2p are expressed in the cytoplasm, but, in some cells, only Acs2p is present when cells are grown in glucose. *ACS1* expression may be repressed in some cells grown in glucose and derepressed when glucose is limited and/or in the presence of non-fermentable carbon sources, e.g., acetate and ethanol. Generally, *ACS2* is constitutively expressed in yeast cells. In some instances, the affinity of Acs1p for acetate may be higher, e.g, about 30-fold higher, than that of Acs2p.

Provided herein are cells, microorganisms, compositions and methods in which the amount and/or activity of acetyl-CoA synthetase (also referred to as ACS or acetate-CoA ligase and used interchangeably herein) in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase the amount of acetyl-CoA synthetase and/or acetyl-CoA synthetase activity, may be modified to decrease the amount of acetyl-CoA synthetase and/or acetyl-CoA synthetase activity, or may be modified to alternately increase and decrease the amount of acetyl-CoA synthetase and/or acetyl-CoA synthetase activity depending, for example, on the cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

For example, in embodiments in which modification of cellular acetate generation yields increased amounts of cytosolic acetate, the amount and/or activity of cytosolic acetyl-CoA synthetase can also be increased to provide for increased conversion of acetate to acetyl-CoA. Heterologous nucleic acid encoding Acs1p and/or Acs2p can be introduced into a host cell to increase the amount of cytosolic acetyl-CoA synthetase. For example, the amount and/or activity of a host cytosolic acetyl-CoA synthetase can be increased by increasing the number of copies of a gene encoding a cytosolic acetyl-CoA synthetase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the gene), by increasing the activity of a promoter that regulates transcription of a gene encoding a cytosolic acetyl-CoA synthetase, or by increasing the number of copies of a gene encoding a cytosolic acetyl-CoA synthetase and increasing the activity of a promoter that regulates transcription of a gene encoding a cytosolic acetyl-CoA synthetase. In some embodiments, a cytosolic acetyl-CoA synthetase is endogenous to the host cell or microorganism. Additionally, a heterologous promoter can be used to regulate expression of a recombinant acetyl-CoA synthetase-encoding nucleic acid. An example of one such heterologous promoter is a fatty acid-inducible promoter that can provide for increased acetyl-CoA synthetase expression, particularly when exposed to fatty acids as a carbon source. Such promoter elements include those that regulate expression of peroxisomal proteins and/or β-

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1470 of 3077

oxidation enzymes in microbes, e.g., a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene promoter.

In other aspects, the amount and/or activity of acetyl-CoA synthetase in a cell is decreased. Methods for decreasing acetyl-CoA synthetase activity in a cell include, but are not limited to, modifying the amount of acetyl-CoA synthetase protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous acetyl-CoA synthetase gene in an organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type acetyl-CoA synthetase such that the encoded modified or substituted acetyl-CoA synthetase protein has a reduced enzyme activity. For example, in some instances, it may be desirable to decrease the amount and/or activity of a peroxisomal protein having acetyl-CoA synthetase activity.

In a particular embodiment, the acetyl-CoA synthetase enzyme can be a *Candida* yeast protein. Examples of *Candida viswanathii* nucleotide sequences (and corresponding amino acid sequences) encoding acetyl-CoA synthetase are provided herein (nucleotide SEQ ID NOS: 76 and 77 and amino acid SEQ ID NOS: 20 and 21). Additional examples of nucleotide sequences encoding acetyl-CoA synthetase proteins include, but are not limited to: *Saccharomyces cerevisiae ACS1* (Genbank accession number NM_001178197), *Saccharomyces cerevisiae ACS2* (Genbank accession number NM_001182040), *Candida tropicalis ACS1* (Genbank accession number XM_002547679), *Candida albicans ACS2* (Genbank accession number AF535132), *Cyberlindnera jadinii ACS1* (Genbank accession number AB641819), *Cyberlindnera jadinii ACS2* (Genbank accession number AB641820), *Kluyveromyces lactis ACS2* (Genbank accession number AF134491).

The acetyl-CoA synthetase activities of host and modified cells and microorganisms can be evaluated and monitored using methods known in the art. Examples of acetyl-CoA synthetase activity assays include a continuous coupled enzymatic assay (see, e.g., Castano-Cerezo et al. (2012) *Bio-protocol* 2(17) and Frenkel and Kitchens (1977) *J. Biol. Chem*. 252(2): 504-507). Nucleic acid sequences representing native and/or modified acetyl-CoA synthetase-encoding sequences also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or the amounts of the nucleic acids or encoded proteins can be assessed using quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1471 of 3077

thereof), where the engineered cells or organisms exhibit increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism.

Modification of citrate processing

Carbon atoms of acetyl groups in mitochondrial acetyl-CoA can also be captured from intermediates of the TCA cycle such as, for example, citrate molecules generated in the first step of the cycle through the citrate synthase-catalyzed condensation of acetyl-CoA and oxaloacetate. Under certain conditions, citrate can be transported from mitochondria into the cytoplasm via a mitochondrial inner membrane citrate transport protein (CTP). This transport protein provides for the efflux of citrate from mitochondria generally in exchange for the influx of a carboxylate molecule (e.g., malate) from the cytosol. Cytosolic citrate can be converted to isocitrate which can serve as a substrate in the NADPH-generating oxidation reaction through which it is converted to α-ketoglutarate. Some yeast, typically oleaginous yeast, express an endogenous ATP citrate lyase which can catalyze the cleavage of citrate into oxaloacetate and acetyl-CoA. Thus, in such instances, cytosolic citrate can serve as a source of acetyl carbons that can be converted to acetyl-CoA. In general, yeast ATP citrate lysate is a dimer and can be heterodimeric (e.g., Acl1p/Acl2p) or homomeric.

In another embodiment of the cells and microorganisms, target molecule production systems and methods provided herein, carbon atoms incorporated into citrate that has been transferred to the cytosol can be captured through the cleavage of citrate to oxaloacetate and acetyl-CoA by the enzyme ATP citrate lyase (i.e., ACL, used interchangeably herein; e.g., EC 2.3.3.8). The capture of metabolite carbon in this manner can divert it from use in other metabolic processes and also can result in an increase in the level cytoplasmic acetyl-CoA. In one aspect of this embodiment, the amount and/or activity of ATP citrate lyase in the cytosol of modified cells is increased relative to the unmodified host cell. The amount and/or activity of a host ATP citrate lyase can be increased, for example, by introducing and/or increasing the number of copies of a gene encoding an ATP citrate lyase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the gene), by increasing the activity of a promoter that regulates transcription of a gene encoding an ATP citrate lyase, or by increasing the number of copies of a gene encoding an ATP citrate lyase and increasing the activity of a promoter that regulates transcription of a gene encoding an ATP citrate lyase. In some embodiments, an ATP citrate lyase is endogenous to

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1472 of 3077

the host cell or microorganism. In other embodiments, a host cell or microorganism does not express an endogenous cytosolic ATP citrate lyase.

Thus, for example, heterologous nucleic acids encoding an ATP citrate lyase can be introduced into a host cell or microorganism to provide for an increased amount and/or activity of cytosolic ATP citrate lyase. In a particular embodiment, the ATP citrate lyase enzyme encoded by the heterologous nucleic acid can be an oleaginous yeast protein. An example of an oleaginous yeast ATP citrate lyase is formed by the *Yarrowia lipolytica* Acl1 and Acl2 proteins. Examples of *Y. lipolytica* Acl1p and Acl2p amino acid sequences are provided herein (SEQ ID NOS: 42 and 43). If a host cell or microorganism is a different species than the heterologous ATP citrate lyase that will be expressed in the host, it can be desirable to introduce nucleic acids encoding the ATP citrate lyase proteins that have been optimized for codons used in the host species. As a non-limiting example, the nucleotide sequences encoding *Yarrowia lipolytica* Acl1p and Acl2p that have been optimized for expression in a different yeast species (*Candida viswanathii*) are provided herein (SEQ ID NOS: 96 and 97). Additional examples of nucleotide sequences encoding ATP citrate lyase proteins include, but are not limited to: *Phaffia rhodozyma ACL1 and ACL2* (Genbank accession numbers KM503045, KM510496) and *Sordaria macrospora ACL1* and *ACL2* (Genbank accession numbers AJ224922, XM_003344949).

The promoter used for regulating transcription of a heterologous nucleic acid encoding an ATP citrate lyase can also be modified. For example, the amount of an ATP citrate lyase protein expressed in a cell may be increased by including in the heterologous nucleic acid a strong heterologous promoter and/or a promoter that provides for a different pattern of expression in the cell or microorganism. An example of one such heterologous promoter is a fatty acid-inducible promoter that can provide for increased ATP citrate lyase expression, particularly when exposed to fatty acids as a carbon source. Such promoter elements include those that regulate expression of peroxisomal proteins and/or β-oxidation enzymes in microbes, e.g., a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene promoter.

ATP citrate lyase activity can be determined using assays known in the art and/or described herein. Such assays include, for example, methods described by Linn and Srere [(1979) *J. Biol. Chem.* 254:1691-1698] and Pentyala and Benjamin [(1995) *Biochemistry* 34:10961-10969]. Nucleic acid sequences representing native and/or modified ATP citrate lyase-encoding sequences also can be detected using nucleic acid detection methods (e.g., PCR, primer

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1473 of 3077

extension, nucleic acid hybridization, the like and combinations thereof), or the amounts of the nucleic acids or encoded proteins can be assessed using quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cells or organisms exhibit increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism.

In some embodiments of the cells, organisms and methods provided herein involving capture of acetyl carbons from cytosolic citrate, it may be beneficial to increase the amount of citrate in the cytosol. One approach to increasing cytosolic citrate levels is by increasing efflux of citrate from mitochondria into the cytosol. One method of increasing mitochondrial citrate efflux involves increasing the amount and/or activity of citrate transporter protein (CTP) in mitochondria of the modified cells. The amount and/or activity of a host citrate transporter protein can be increased, for example, by introducing and/or increasing the number of copies of a gene encoding a mitochondrial citrate transporter (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the gene), by increasing the activity of a promoter that regulates transcription of a gene encoding a mitochondrial citrate transporter, or by increasing the number of copies of a gene encoding a mitochondrial citrate transporter and increasing the activity of a promoter that regulates transcription of a gene encoding a mitochondrial citrate transporter. In some embodiments, a mitochondrial citrate transporter is endogenous to the host cell or microorganism.

Thus, for example, heterologous nucleic acids encoding a mitochondrial citrate transporter can be introduced into a host cell or microorganism to provide for an increased amount and/or activity of a mitochondrial citrate transporter. In a particular embodiment, the mitochondrial citrate transporter encoded by the heterologous nucleic acid can be a yeast protein. If a host cell or microorganism is a different species than the heterologous mitochondrial citrate transporter that will be expressed in the host, it can be desirable to introduce nucleic acids encoding the mitochondrial citrate transporter that have been optimized for codons used in the host species. Examples of nucleotide sequences encoding a protein that may have mitochondrial citrate transporter activity include, but are not limited to: *Candida albicans CTP1* (Genbank accession number XM_019475315), *Candida orthopsilosis CTP1* (Genbank accession number XM_003868562), *Saccharomyces cerevisiae CTP1* (Genbank accession number NM_001178639) and *Candida tropicalis* (Genbank accession number XM_002548023).

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1474 of 3077

Modification of acyl-CoA formation, hydrolysis and use

Acyl-CoA is a molecule containing a carboxylic acid and coenzyme A joined through a thioester bond. In cells, acyl-CoA can be generated from carboxylic acids entering the cytosol from the extracellular environment, fatty acids synthesized within cells, lipid sidechains resulting from membrane turnover, products of the hydrolysis of triglyceride and sterol esters and carboxylation of acetyl-CoA. Acyl-CoA participates in multiple cellular pathways including lipid synthesis, β-oxidation, fatty acid synthesis and protein acylation. One reaction through which acyl-CoA is formed is the condensation between a thiol group of coenzyme A and a carboxy group of a carboxylic acid. This reaction between a fatty acid and coenzyme A is referred to as activation of the free fatty acid and can be catalyzed by an acyl-CoA synthetase enzyme (e.g. EC 6.2.1.3). A short-chain acyl-CoA, for example, malonyl-CoA, can also be generated through carboxylation of acetyl-CoA in a reaction catalyzed by an acetyl-CoA carboxylase (e.g., EC 6.4.1.2). Conversely, free fatty acids can be liberated from acyl-CoA through the action of a thioesterase (e.g., EC 3.1.2.20). Because acyl-CoA is a major carrier molecule of cellular carbons, its formation and hydrolysis represent certain aspects of methods of modifying carbon flux in cells.

Modification of acyl-CoA synthetase activity

Acyl-CoA synthetases (also referred to as fatty acid or acyl Co-A ligases and used interchangeably herein), are a family of enzymes in the enzyme classification subgroup 6.2.1 with varying substrate affinities, expression patterns and cellular localizations. In many microorganisms, there are multiple, distinct genes encoding separate acyl-CoA synthetases. Many yeast species (e.g., *Candida* spp. and *Saccharomyces* spp.) have five or six or more acyl-CoA synthetase genes encoding distinct enzymes. For example, *Saccharomyces cerevisiae* has 4 *FAA* genes (*FAA1*, *FAA2*, *FAA3* and *FAA4*) and 2 *FAT* genes (*FAT1* and *FAT2*) encoding acyl-CoA synthetase enzymes. Generally, *FAA* gene-encoded enzymes catalyze activation of acyl chains containing about 8-20 carbon atoms whereas the enzyme encoded by *FAT1* typically catalyzes activation of acyl chains containing 20 or more carbon atoms. Faa1p and Faa4p, which tend to be located in the cytosol and associated with membranes, are involved in activation of fatty acids internalized into cells from the extracellular medium and intracellular fatty acids arising from degradation of lipids, triacylglycerides and steryl esters. The Faa1p isozyme can exhibit broad substrate chain-length specificity, represents 90% of the cellular acyl-

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1475 of 3077

CoA synthetase activity, and is localized in the cytosolic and microsomal fractions. Faa4p has broad chain-length specificity and has been shown to be important in protein myristoylation. Faa2p is localized to peroxisomes, has broad chain-length specificity, and participates in activation of fatty acids occurring during β-oxidation. Fat1p is typically a dual function protein localized to the cellular membrane that has activity for both fatty acid transport and fatty acid activation. Fat2p tends to be targeted to the peroxisomal membrane for medium chain fatty acid transport and activation.

Homologs for *FAA1* and *FAT1* have been identified in *Candida* strains. Acyl-CoA synthetase has six isoforms encoded by FAA1, FAT1, ACS2A, ACS2B, ACS2C and ACS2D, respectively, in *Candida* spp. (e.g., homologous to FAA1, FAT1, and FAA2 in *S. cerevisiae*). Two of the homologs display 95% identity to one another and are most likely alleles of the same gene. Four *FAA2* homologs have been identified in *Candida* strain ATCC 20336 (also referred to in the art as acyl-CoA synthetase-encoding genes *ACS2A* through *ACS2D* in *Candida*). Examples of *Candida viswanathii* nucleotide sequences (and corresponding amino acid sequences) encoding acyl-CoA synthetases are provided herein (nucleotide SEQ ID NOS: 91 and 98 and amino acid SEQ ID NOS: 35 and 44) and in International patent application no. PCT/US2012/045615 (publication no. WO 2013/106730). Acetyl-CoA synthetase-encoding genes are also referred to as *ACS* genes, as described herein. For clarity, acyl-CoA synthetase-encoding genes are referred to as *FAA* or *FAT* herein and not as *ACS* genes (which herein refer to acetyl-CoA synthetase-encoding genes).

In some embodiments of the microorganisms, compositions and methods provided herein, the amount and/or activity of acyl-CoA synthetase in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase acyl-CoA synthetase and/or acyl-CoA synthetase activity, may be modified to decrease acyl-CoA synthetase and/or acyl-CoA synthetase activity, or may be modified to alternately increase and decrease acyl-CoA synthetase and/or acyl-CoA synthetase activity depending, for example, on the cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

In certain aspects, the amount and/or activity of an acyl-CoA synthetase in a cell is decreased. Reducing or eliminating the amount and/or activity of an acyl-CoA synthetase may be particularly beneficial in embodiments in which the flux of carbons from fatty acids is directed

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1476 of 3077

toward a particular target product pathway and away from other cellular metabolic pathways involving activated fatty acids. When free internalized or cytosolic fatty acids are activated by acyl-CoA synthetase and used in cellular processes, such as lipid biosynthesis, the carbon atoms in the free fatty acids are not available for use in the cell or microorganism production of commercially important chemicals. Without being limited by theory, it is believed that reduction in the amount of fatty-acyl-CoA available for various cellular processes can increase the amount of fatty acids available for conversion into target molecules, for example, a fatty dicarboxylic acid (e.g., adipic acid, suberic acid, sebacic acid and dodecanedioic acid) by other engineered pathways in the same host cell or organism (e.g., omega oxidation pathway, beta oxidation pathway, omega oxidation pathway and beta oxidation pathway).

In some embodiments, one strategy is to control the subcellular location of acyl-CoA synthetase enzyme activity so that it is present only in the peroxisome. FAA1 and FAT1 mutants, faa1Δ and fat1Δ, of *Candida* were constructed and should have very little acyl-CoA synthetase activity targeted to the cytoplasm. In these mutant strains, exogenously supplied long-chain free fatty acids tend to accumulate in the cytoplasm since they cannot be transported into the peroxisome unless they are activated to the acyl-CoA thioester. High concentrations of free fatty acid can be toxic, so the cell acts to detoxify itself by oxidizing the free fatty acids to dicarboxylic acids that are much less toxic. Unlike long-chain fatty acids, long-chain dicarboxylic acids are able to diffuse into the peroxisomal compartment where they can then be activated to diacyl-CoA thioesters and enter into the beta-oxidation pathway. With multiple peroxisomal acyl-CoA synthetase isozymes it may be that each isozyme has different substrate specificity. In some embodiments, it is desired to retain those peroxisomal acyl-CoA synthetase enzymes with substrate specificity matching the chain-length of the fatty acid feedstock but without activity (or low activity) on diacids of chain-length ≤ 6, 8, 10, 12, 14, 16, 18 or 20 carbons. With this strategy, any long-chain dicarboxyl-CoA that is chain-shortened by beta-oxidation to 12 carbons, for example, that is subsequently hydrolyzed to a dicarboxylic acid and free CoA cannot be reactivated to a dicarboxyl-CoA for re-entry into beta-oxidation for further chain shortening. In some embodiments, in combination with controlling the substrate chain-length specificity of the peroxisomal acyl-CoA synthetase, a peroxisomal thioesterase activity is amplified with maximum activity at the desired chain-length of a target product. This strategy can control the chain-length of the dicarboxylic acid produced by beta-oxidation.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1477 of 3077

Certain aspects of the microorganisms, compositions and methods provided herein include one or more modifications to reduce or eliminate cytosolic activation of free fatty acids into acyl-CoA. One approach to reducing or eliminating cytosolic free fatty acid activation is to decrease the amount and/or activity of an acyl-CoA synthetase. For example, endogenous microbial genes encoding one or more acyl-CoA synthetases (e.g., yeast *FAA1, FAA4* and/or *FAT1* gene) can be disrupted or deleted in a host cell or microorganism to reduce or eliminate acyl-CoA synthetase activity in the host relative to a cell or microorganism in which the gene(s) have not been modified. Methods for decreasing the amount and/or activity of one or more acyl-CoA synthetases, such as acyl-CoA synthetases involved in activation of cytosolic free fatty acids, in a cell include, but are not limited to, modifying the amount of acyl-CoA synthetase protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous acyl-CoA synthetase gene in a cell or organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type acyl-CoA synthetase such that the encoded modified or substituted acyl-CoA synthetase protein has a reduced enzyme activity. For example, expression of a host acyl-CoA synthetase activity can be decreased or eliminated by disruption (e.g., knockout, insertion mutagenesis, the like and combinations thereof) of a host gene encoding the protein, or by decreasing the activity of the promoter (e.g., through addition of repressor sequences to the promoter or 5'UTR or replacing the promoter) that controls transcription of an acyl-CoA synthetase gene using recombinant molecular biology techniques known in the art and/or described herein. In one embodiment, a diploid yeast, such as, for example, a *Candida* yeast, when used as a host microorganism can be subjected to genetic modification in which one of the two alleles of an acyl-CoA synthetase gene is disrupted or deleted. In so doing, a single allele of the gene remains for a reduced amount of acyl-CoA synthetase expression in the microorganism and a reduced amount of the protein in the cell.

One method for disrupting an endogenous acyl-CoA synthetase gene is by recombinantly inserting a heterologous nucleic acid (e.g., a nucleotide sequence encoding a selectable marker such as an enzyme that restores an auxotrophic host organism to prototrophy) into the endogenous gene, thereby generating an engineered organism deficient in acyl-CoA synthetase activity. This can be done, for example, through homologous recombination in which a heterologous nucleic acid containing sequences of an endogenous acyl-CoA synthetase gene and a disrupting sequence (e.g., a knock-out gene cassette such as described herein) is introduced into a host cell or microorganism. Nucleic acids encoding an acyl-CoA synthetase

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1478 of 3077

can be obtained from a number of sources, including, for example, yeast cells. Genomic DNA from cell sources can be amplified using oligonucleotide primers based on the nucleotide sequence of an acyl-CoA synthetase encoding gene, including examples provided herein. Nucleotide sequences encoding (and the amino acid sequences of) *Candida viswanathii* acyl-CoA synthetase Faa1p and Fat1p are provided herein (nucleotide SEQ ID NOS: 91 and 98 and amino acid SEQ ID NOS: 35 and 44). Additional non-limiting examples of nucleic acids encoding acyl-CoA synthetases include *Saccharomyces cerevisiae FAA1* (Genbank accession numbers NM_001183737 and M96371), *Saccharomyces cerevisiae FAA4* (Genbank accession number NM_001182754), *Saccharomyces cerevisiae FAA2* (Genbank accession number NM_001178906), *Saccharomyces cerevisiae FAA3* (Genbank accession number NM_001179359), *Yarrowia lypolytica YAL1* (Genbank accession number XM_502959), *Yarrowia lypolytica FAT1* (Genbank accession number NC_006071), *Candida albicans FAA4* (Genbank accession number XM_714261), *Aspergillus nomius FAA4* (Genbank accession number XM_015551345), *Coccidioides immitis FAA4* (Genbank accession number XM_001240655) and *Aspergillus niger FAA4* (Genbank accession number XM_001397786).

In other embodiments, the amount and/or activity of a host acyl-CoA synthetase can be increased, for example, by increasing the number of copies of a gene encoding an acyl-CoA synthetase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the gene), by increasing the activity of a promoter that regulates transcription of a gene encoding an acyl-CoA synthetase, or by increasing the number of copies of a gene encoding an acyl-CoA synthetase and increasing the activity of a promoter that regulates transcription of a gene encoding an acyl-CoA synthetase. In some embodiments, an acyl-CoA synthetase is endogenous to the host cell or microorganism. In particular embodiments, the amount and/or activity of a host peroxisomal acyl-CoA synthetase is increased.

The presence, absence or amount of acyl-CoA synthetase activity can be detected by any suitable method known in the art. Non-limiting examples of suitable detection methods include enzymatic assays (e.g., Lageweg et al. (1991) *Anal. Biochem.* 197(2):384-388, Erland et al. (2001) *Anal. Biochem.* 295(1):38-44), PCR based assays (e.g., qPCR, RT-PCR), immunological detection methods (e.g., antibodies specific for acyl-CoA synthetase), the like and combinations thereof. Methods for determining acyl-CoA synthetase activities also include assays described by Trigatti et al. [(1992) *Biochem. Cell. Biol.* 70:76-80] and Kamiryo et al. [(1977) *Proc. Natl. Acad. Sci. USA* 74:4947-4950].

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1479 of 3077

Modification of acetyl-CoA carboxylase activity

Malonyl-CoA is a coenzyme A derivative of the dicarboxylic acid malonic acid that can serve as a precursor in the synthesis of numerous valuable organic molecules, including fatty acids and polyketides. In the cytosol, malonyl-CoA can be generated by carboxylation of acetyl-CoA through the addition of CO_2 (e.g., derived from bicarbonate) in a reaction catalyzed by the enzyme acetyl-CoA carboxylase (e.g., EC 6.4.1.2). Acetyl-CoA carboxylase sometimes is also referred to as "acetyl-CoA:carbon-dioxide ligase (ADP-forming)" and "acetyl coenzyme A carboxylase". In eukaryotes, acetyl-CoA carboxylase is a multifunctional polypeptide containing a biotin carrier protein domain, a biotin carboxylase domain and a carboxyl-transferase domain. Biotin joined to the biotin carrier protein is a co-factor in malonyl-CoA formation. It receives CO_2 which becomes attached to it at a biotin ring nitrogen in an ATP-dependent reaction catalyzed by the biotin carboxylase of the acetyl-CoA carboxylase. The activated CO_2 is then transferred from biotin to acetyl-CoA by the carboxyl-transferase domain to form malonyl-CoA. Malonyl-CoA can serve as a carbon donor in the synthesis of a fatty acid chain in repeated cycles of the addition of 2 carbon atoms per cycle to extend the chain and generate a fatty acid. The reactions of each cycle are catalyzed by fatty acid synthase (FAS) and continue until typically a 16-carbon fatty acid (palmitic acid) or 18-carbon fatty acid (stearic acid) is completed in the form of palmitoyl-CoA or stearoyl-CoA, respectively. Accordingly, a supply of malonyl-CoA, and/or precursors and enzymes (e.g., acetyl-CoA carboxylase) that generate malonyl-CoA, can be required for fatty acid synthesis.

In some embodiments of the cells, microorganisms, compositions and methods provided herein, the amount and/or activity of acetyl-CoA carboxylase in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase acetyl-CoA carboxylase and/or acetyl-CoA carboxylase activity, may be modified to decrease acetyl-CoA carboxylase and/or acetyl-CoA carboxylase activity, or may be modified to alternately increase and decrease acetyl-CoA carboxylase and/or acetyl-CoA carboxylase activity depending, for example, on the target molecule(s) produced and/or on the conditions in which the modified cell or microorganism is cultured.

A modification of cellular carbon flux that increases cytosolic acetyl-CoA alone may not be optimal for enhancing fatty acid or other target molecule production in an engineered, cell-based

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1480 of 3077

or microbial system if there is not a concurrent increase in conversion of acetyl-CoA to malonyl-CoA. To maximize production efficiencies, included in the cells, microorganisms, compositions and methods provided herein are cellular carbon flux modifications that increase the amount of cytosolic malonyl-CoA. In one embodiment, the amount and/or activity of acetyl-CoA carboxylase is increased in the cytosol to direct carbon flux towards generation of malonyl-CoA. The amount and/or activity of a host acetyl-CoA carboxylase can be increased, for example, by increasing the number of copies of a gene encoding an acetyl-CoA carboxylase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the gene), by increasing the activity of a promoter that regulates transcription of a gene encoding an acetyl-CoA carboxylase, or by increasing the number of copies of a gene encoding an acetyl-CoA carboxylase and increasing the activity of a promoter that regulates transcription of a gene encoding an acetyl-CoA carboxylase. In some embodiments, an acetyl-CoA carboxylase is endogenous to the host cell or microorganism. An acetyl-CoA carboxylase activity may be amplified by over-expression of an acetyl-CoA carboxylase gene by any suitable method. Non-limiting examples of methods suitable to amplify or over express a gene include amplifying the number of acetyl-CoA carboxylase genes in yeast, for example, following transformation with a high-copy number plasmid (e.g., such as one containing a 2u origin of replication), integration of multiple copies of the gene into the host genome, over-expression of the gene directed by a strong promoter, the like or combinations thereof. An acetyl-CoA carboxylase gene may be native to *Candida tropicalis* or *Candida viswanithii*, for example, or it may be obtained from a heterologous source. Examples of a *Candida viswanathii* acetyl-CoA carboxylase polypeptide amino acid sequence (Acc1p), and nucleotide sequence encoding it (*ACC1*), are provided herein (nucleotide SEQ ID NO: 74 and amino acid SEQ ID NOS: 18 and 19). Additional non-limiting examples of nucleic acids encoding an acetyl-CoA carboxylase include *Yarrowia lypolytica ACC1* (Genbank accession NC_006069), *Saccharomyces cerevisiae ACC1* (Genbank accession NM_001183193), *Candida tropicalis ACC* (Genbank accession number XM_002546179), *Candida albicans ACC1* (Genbank accession number XM_713531), *Aspergillus nidulans ACCA* (Genbank accession number Y15996), *Aspergillus niger ACCA* (Genbank accession number XM_001395439), *Aspergillus oryzae ACC* (Genbank accession number XM_001826359), *Schizosaccharomyces pombe ACC* (Genbank accession D78169), *Neurospora crassa ACC* (Genbank accession XM_957924), *Lipomyces starkeyi ACC1* (Genbank accession KJ948118), *Debaryomyces hansenii ACC1* (Genbank accession XM_457211), *Amylomyces rouxii ACC* (Genbank accession EF397565) and *Coccidioides immitis ACC* (Genbank accession number XM_001247055).

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1481 of 3077

The promoter used for regulating transcription of a heterologous nucleic acid encoding an acetyl-CoA carboxylase can also be modified. For example, the amount of an acetyl-CoA carboxylase protein expressed in a particular cellular location may be increased by including in the heterologous nucleic acid a strong heterologous promoter and/or a promoter that provides for a different pattern of expression in the cell or microorganism. An example of one such heterologous promoter is a fatty acid-inducible promoter that can provide for increased acetyl-CoA carboxylase expression, particularly when exposed to fatty acids as a carbon source. Such promoter elements include those that regulate expression of peroxisomal proteins and/or β-oxidation enzymes in microbes, e.g., a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene promoter. The nucleotide sequence of a *Candida viswanathii* HDE gene promoter is provided herein as are examples of additional fatty acid-inducible promoters.

The reverse activity (e.g., decarboxylation of malonyl-CoA) is carried out by a separate enzyme, malonyl-CoA decarboxylase. In some embodiments, to further increase carbon flux through a particular reaction or through a metabolic pathway, one or more reverse activities in the pathway can be altered to inhibit the back conversion of a desired product into its starting reactants. In certain embodiments, a malonyl-CoA decarboxylase activity is reduced or eliminated to further increase the carbon flux through an acetyl-CoA carboxylase activity in the direction of malonyl-CoA production.

Acetyl-CoA carboxylase is regulated by feedback inhibition of acyl-CoA (e.g., palmitoyl-CoA) and by phosphorylation. As such, increasing the copy number of acetyl-CoA carboxylase-encoding nucleic acids in a cell may not alone be sufficient in increasing the acetyl-CoA carboxylase activity in the cell. Because the dephosphorylated state is the active state of the enzyme, one approach for increasing the activity of acetyl-CoA carboxylase is to reduce or eliminate phosphorylation of the protein. Provided herein are modified acetyl-CoA carboxylase proteins (and mutant nucleic acids encoding the proteins) in which one or more phosphorylatable serine residues have been substituted with alanine residues thereby relieving the regulation by phosphorylation. In a particular embodiment, the modified acetyl-CoA carboxylase is a modified yeast enzyme. For example, as described herein, an endogenous *Candida viswanathii* acetyl-CoA carboxylase wild-type enzyme was modified to substitute alanine residues for one or more of the following serine amino acid residues: S652, S1131, S1138, S1153, S1158. The modified amino acid sequences are provided herein. Also provided

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1482 of 3077

herein are cells and microorganisms containing heterologous nucleic acid encoding a modified acetyl-CoA carboxylase protein and methods of increasing cytosolic malonyl-CoA generation and producing organic acids and other target products using the compositions provided herein. In another aspect of this embodiment, the amount and/or activity of acetyl-CoA carboxylase is increased in the cytosol of modified cells relative to an unmodified host cell by introducing multiple copies of the modified heterologous nucleic acid encoding acetyl-CoA carboxylase into a host cell to provide for increased acetyl-CoA carboxylase protein and/or by increasing the expression of the modified acetyl-CoA carboxylase in the cell through the use of a strong heterologous promoter.

In other aspects, the amount and/or activity of acetyl-CoA carboxylase in a cell is decreased. Methods for decreasing acetyl-CoA carboxylase activity in a cell include, but are not limited to, modifying the amount of acetyl-CoA carboxylase protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous acetyl-CoA carboxylase gene in an organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type acetyl-CoA carboxylase such that the encoded modified or substituted acetyl-CoA carboxylase protein has a reduced enzyme activity.

The presence, absence or amount of acyl-CoA carboxylase activity can be detected by any suitable method known in the art. Non-limiting examples of suitable detection methods include radioactive HCO_3^- incorporation assays and coupled enzyme assays (e.g., Diacovich et al. (2002) *J. Biol. Chem.* 277(34):31228-31236), PCR based assays (e.g., qPCR, RT-PCR), immunological detection methods (e.g., antibodies specific for acyl-CoA carboxylase), the like and combinations thereof.

Modification of thioesterase activity

A thioesterase is an enzyme that catalyzes the hydrolysis of a thioester bond between a carbonyl group and a sulfur atom. In cells, certain thioesterases (e.g., acyl-CoA thioesterase activity, acyl-ACP thioesterase activity) catalyze the removal of Coenzyme A or acyl carrier protein (e.g., ACP) from a fatty acid yielding a free fatty acid and unesterified carrier, e.g., Coenzyme A (CoASH). The reaction occurs in the presence of water, and Coenzyme A or acyl carrier protein is specifically removed at a thiol group. The released CoA can then be reused for other cellular activities. A non-limiting example of an enzyme with thioesterase activity is acyl-

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1483 of 3077

CoA hydrolase (e.g., EC 3.1.2.20; also referred to as acyl coenzyme A thioesterase, acyl-CoA thioesterase, acyl coenzyme A hydrolase, thioesterase B, thioesterase II, lecithinase B, lysophopholipase L1, acyl-CoA thioesterase 1, and acyl-CoA thioesterase). In eukaryotic microorganisms, acyl-CoA thioesterases are generally localized in peroxisomes but may also occur in mitochondria.

In some embodiments of the cells, microorganisms, compositions and methods provided herein, the amount and/or activity of a thioesterase in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase the amount of thioesterase and/or thioesterase activity, may be modified to decrease thioesterase and/or thioesterase activity, or may be modified to alternately increase and decrease thioesterase and/or thioesterase activity depending, for example, on the cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

Embodiments of cells and microorganisms provided herein in which carbon flux is modified to increase acetyl-CoA carboxylase and/or cytosolic malonyl-CoA, may further benefit from also modifying the amount of medium-to-long chain fatty acids present in the cytosol in the esterified form as acyl-CoA (e.g., palmitoyl-CoA). In some of these embodiments, the increased generation of malonyl-CoA can lead to increased fatty acid synthesis in the presence of cytosolic fatty acid synthase (FAS). The end-product of cytosolic fatty acid synthesis in yeast cells is typically an acyl-CoA, e.g., palmitoyl-CoA, which can then be used in cellular metabolic pathways other than desired engineered target molecule production processes. This represents a loss of the carbon atoms in the acyl-CoA which could have been incorporated into target products. Additionally, high levels of cytosolic acyl-CoA end products of fatty acid synthesis (e.g., palmitoyl-CoA) can inhibit acetyl-CoA carboxylase. Therefore, production efficiency may be enhanced in some embodiments by decreasing the amount of fatty acids present in the cytosol in the esterified form as acyl-CoA.

Included in the cells, microorganisms, compositions and methods provided herein are cellular carbon flux modifications that decrease the amount of fatty acids present in the cytosol in the esterified form as acyl-CoA. In one embodiment, the amount and/or activity of thioesterase is increased in the cytosol of cells or microorganisms. The amount and/or activity of a thioesterase can be increased, for example, by increasing the number of copies of a gene encoding a thioesterase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the gene),

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1484 of 3077

by increasing the activity of a promoter that regulates transcription of a gene encoding a thioesterase, or by increasing the number of copies of a gene encoding a thioesterase and increasing the activity of a promoter that regulates transcription of a gene encoding a thioesterase. In certain aspects, the amount and/or activity of a thioesterase in a cell and/or a particular cellular location is increased. For example, the amount and/or activity of a thioesterase in the cytosol of a cell can be increased. In some embodiments, the pattern of expression of a thioesterase can be modified such that the enzyme is produced in a cellular location where it is not produced in an unmodified cell and/or is no longer produced in a cellular location where it is produced in an unmodified cell.

In yeast, thioesterases are generally present in the peroxisomal compartment of the cells to ensure that free coenzyme A is available for beta-oxidation. Without being bound or limited by theory, this enzyme should not normally be present in the cytoplasm because producing fatty acyl-CoA via cytosolic fatty acid synthesis is an energy intensive process, and removing the CoA from the synthesized acyl-CoA would waste the energy put into the process. In order to provide for generation of free fatty acids in the cytoplasm of modified cells or organisms, peroxisomal thioesterase with activity on long chain fatty acids can be re-targeted to the cytoplasm. In one embodiment of the cells, microorganisms and methods provided herein, host cells are modified to express engineered thioesterase polypeptides that lack targeting signals that direct the enzyme to peroxisomes (i.e, PTS), thereby introducing, or increasing the amount of, cytosolic thioesterase in the cells. In a particular embodiment, a thioesterase lacking a PTS that has activity on medium and long chain fatty acids is heterologously expressed in the cytoplasm of modified cells. This can be accomplished by modifying nucleic acids encoding a peroxisomal thioesterase to delete the portion of the nucleic acid encoding the peroxisomal targeting signal at the C-terminus of the protein, and introducing the modified heterologous nucleic acid into host cells. An example of one such modified thioesterase protein is a yeast $Tes3p^{\Delta pts}$ which excludes a peroxisomal targeting signal (PTS). As described herein, generally, a yeast peroxisomal targeting sequence is a 3-amino acid consensus sequence (PTS1). In a particular embodiment, the thioesterase enzyme can be a *Candida* yeast protein. For example, *Candida* strain ATCC 20336 contains eight genes encoding peroxisomal thioesterases (*TES1-TES8*), each of which contains a C-terminal 3-amino acid PTS1 consensus sequence (SRL, ARL) or slight variant thereof (PKL, PKF). Nucleotide sequences encoding the 8 thioesterases, and the amino acid acid sequences of the thioesterases, are provided herein (nucleotide SEQ ID NOS: 78, 79, 80, 81, 82, 83, 84, 85, 86 and 87 and amino acid SEQ ID NOS: 22, 23, 24, 25,

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1485 of 3077

26, 27, 28, 29, 30 and 31). An example of a *Candida viswanathii* nucleotide sequence encoding a modified Tes3p lacking a peroxisomal-targeting sequence (*TES3*$^{\Delta pts}$), and the amino acid sequence of the modified Tes3p, are also provided herein (nucleotide SEQ ID NO: 88 and amino acid SEQ ID NO: 32). In some embodiments, the thioesterase polypeptide is from a different species than a host microorganism in which it is expressed.

Nucleic acid sequences encoding polypeptides conferring thioesterase activity can be obtained from a number of sources, including, for example, yeast (e.g., *Candida*, *Saccharomyces*, *Debaryomyces*, *Meyerozyma*, *Lodderomyces*, *Scheffersomyces*, *Clavispora*, *Yarrowia*, *Pichia, Kluyveromyces*, *Eremothecium*, *Zygosaccharomyces*, *Lachancea*, *Nakaseomyces*), animals (e.g., *Homo*, *Rattus*), bacteria (e.g., *Escherichia*, *Pseudomonas*, *Bacillus*), and plants (e.g., *Arabidopsis*, *Nictotania*, *Cuphea*). Examples of nucleotide sequences encoding polypeptides having thioesterase activity include, without limitation, *Saccharomyces cerevisiae PTE1* (Genbank accession no. AF124265), *Debaryomyces hansenii* (Genbank accession nos. XM_456353, XM_459767), *Aspergillus niger* (Genbank accession nos. XM_001392518, XM_011389712, XM_011395790), *Aspergillus fumigatus* (Genbank accession no. XM_742375), *Candida albicans* (Genbank accession nos. XM_705831, XM_705833), *Candida dubliniensis* (Genbank accession no. XM_002418475), *Candida orthopsilosis* (Genbank accession nos. XM_003866686, XM_003866684), *Neurospora crassa* (Genbank accession nos. XM_956915, XM_960627), *Rhodotorula toruloides* (Genbank accession no. XM_016414800), *Cryptococcus neoformans* (Genbank accession no. XM_012196078, XM_012195836), *Escherichia coli TesA* (Genbank accession no. L06182) and acyl-(ACP) thioesterase type B from *Cuphea lanceolata* (Genbank accession no. CAB60830).

The promoter used for regulating transcription of a heterologous nucleic acid encoding a thioesterase can also be modified. For example, the amount of a thioesterase protein expressed in a particular cellular location may be increased by including in the heterologous nucleic acid a strong heterologous promoter and/or a promoter that provides for a different pattern of expression in the cell or microorganism. An example of one such heterologous promoter is a fatty acid inducible promoter that can provide for increased thioesterase expression, particularly when exposed to fatty acids as a carbon source. Such promoter elements include those that regulate expression of peroxisomal proteins and/or β-oxidation enzymes in microbes, e.g., a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1486 of 3077

promoter. The nucleotide sequence of a *Candida viswanathii* HDE gene promoter is provided herein as are examples of additional fatty acid-inducible promoters.

In other aspects, the amount and/or activity of a thioesterase in a cell is decreased. Methods for decreasing thioesterase activity in a cell include, but are not limited to, modifying the amount of thioesterase protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous thioesterase gene in an organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type such that the encoded modified or substituted thioesterase protein has a reduced enzyme activity.

Presence, absence or amount of thioesterase activity can be detected by any suitable method known in the art or described herein (see, e.g., Jones et al. (1999) *J. Biol. Chem.* 274(14):9216-9223 and Chemistry and Biology 9: 981-988). Nucleic acid sequences representing native and/or modified thioesterase-encoding sequences also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or the amounts of the nucleic acids or encoded proteins can be assessed using quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cells or organisms exhibit increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism.

Methods of modifying cells and organisms

Provided herein are cells and organisms (including microorganisms) that have been modified in one or more aspects relative to the unmodified cell or organism (i.e., the cell or organism prior to the modification). For example, a cell or organism can be modified by altering one or more cellular activities and/or the sum total of a cell's or organism's activities. Thus, modifications can include, for example, alteration of cellular activities, addition of cellular activities and/or elimination of cellular activities. A cell or organism may be modified, for example, by altering the amount of one or more cellular compositions, e.g, polynucleotides and/or polypeptides. In some embodiments, an activity and/or amount of a composition can be altered by genetically modifying a host cell or microorganism which yields an engineered cell or microorganism having added, increased, reduced, decreased or removed activity or composition. Genetic

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1487 of 3077

modifications can be achieved in several ways, including, for example, introducing heterologous nucleic acids into host cells or organisms using molecular biological techniques known in the art and/or described herein.

Polynucleotides

The term "polynucleotide" is used herein interchangeably with the term "nucleic acid" and refers to an organic polymer composed of two or more monomers including nucleotides, nucleosides or analogs thereof, including but not limited to single stranded or double stranded, sense or antisense deoxyribonucleic acid (DNA) of any length and, where appropriate, single stranded or double stranded, sense or antisense ribonucleic acid (RNA) of any length, including siRNA. The term "nucleotide" refers to any of several compounds that consist of a ribose or deoxyribose sugar joined to a purine or a pyrimidine base (nitrogenous base) and to a phosphate group, and that are the basic structural units of nucleic acids. The term "nucleoside" refers to a compound (as guanosine or adenosine) that consists of a purine or pyrimidine base (nitrogenous base) combined with deoxyribose or ribose and is found especially in nucleic acids. The term "nucleotide analog" or "nucleoside analog" refers, respectively, to a nucleotide or nucleoside in which one or more individual atoms have been replaced with a different atom or with a different functional group. Accordingly, the term polynucleotide includes nucleic acids of any length, DNA, RNA, analogs and fragments thereof. A polynucleotide of three or more nucleotides is also called nucleotidic oligomer or oligonucleotide.

A nucleic acid (e.g., a nucleic acid reagent, target nucleic acid, target nucleotide sequence, nucleic acid sequence of interest or nucleic acid region of interest) can be from any source or composition, such as DNA, cDNA, gDNA (genomic DNA), RNA, siRNA (short inhibitory RNA), RNAi, tRNA or mRNA, for example, and can be in any form (e.g., linear, circular, supercoiled, single-stranded, double-stranded, and the like). A nucleic acid can also comprise DNA or RNA analogs (e.g., containing base analogs, sugar analogs and/or a non-native backbone and the like). It is understood that the term “nucleic acid” does not refer to or infer a specific length of the polynucleotide chain, thus polynucleotides and oligonucleotides are also included in the definition. Deoxyribonucleotides include deoxyadenosine, deoxycytidine, deoxyguanosine and deoxythymidine. For RNA, the uracil base is uridine.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1488 of 3077

The terms “endogenous polynucleotide,” “endogenous nucleic acid,” “native polynucleotide” and “native nucleic acid,” used interchangeably herein, refer to a polynucleotide of a cell or organism that exists, or is inherent, in the genetic composition of the cell or organism prior to modification.

The terms "heterologous polynucleotide," "heterologous nucleic acid," “exogenous polynucleotide,” “exogenous nucleic acid,” “foreign polynucleotide” and “foreign nucleic acid,” used interchangeably herein, refer to a polynucleotide as it relates to a particular reference cell or organism (e.g., a host cell or organism) and is one that is not present in the genetic composition of the reference cell or organism. A heterologous polynucleotide includes a polynucleotide that may be identical in nucleotide sequence to an endogenous polynucleotide present in a cell, but if introduced into the cell would alter the genetic composition of the cell by, for example, increasing the copy number of the polynucleotide in the cell, altering the position(s) of the polynucleotide in the cell genome, altering the expression of the polynucleotide in the cell, and the like. Thus, such a heterologous nucleic acid thereby genetically modifies the cell into which it is introduced. A heterologous polynucleotide in a host cell may exist in a nucleic acid autonomous of the host chromosome or may be inserted into a host chromosome. A heterologous polynucleotide can also be a polynucleotide with a different nucleotide sequence relative to any nucleic acid in a particular reference cell and can also be obtained from a different cell type or species of organism. A heterologous nucleic acid can also be generated by synthetic methods known in the art and/or described herein.

The term "expression" with respect to a nucleic acid sequence or protein refers to transcription of the nucleic acid and/or, as appropriate, translation of an mRNA transcript to a protein (protein synthesis). Thus, as will be clear from the context, expression of a protein results from transcription and translation of an open reading frame (ORF) sequence. The level of expression of a nucleic acid and/or protein in a cell may be determined, for example, on the basis of either the amount of RNA transcript of a nucleic acid that is present in the cell and/or the amount of the product encoded by the nucleic acid. For example, mRNA transcribed from a selected sequence can be quantitated by qRT-PCR or by Northern hybridization (see Sambrook et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press (1989)). Protein encoded by a nucleic acid can be quantitated by various methods, e.g., by ELISA, by assaying for the biological activity of the protein, or by employing assays that are independent of such activity, such as western blotting or radioimmunoassay, using antibodies that recognize and bind the protein (see Sambrook et al., 1989, supra).

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1489 of 3077

A nucleic acid sometimes is a plasmid, phage, autonomously replicating sequence (ARS), centromere, artificial chromosome, yeast artificial chromosome (e.g., YAC) or other nucleic acid able to replicate or be replicated in a host cell. In certain embodiments a nucleic acid can be from a library or can be obtained from enzymatically digested, sheared or sonicated genomic DNA (e.g., fragmented) from an organism of interest. Fragments can be generated by any suitable method in the art, and the average, mean or nominal length of nucleic acid fragments can be controlled by selecting an appropriate fragment-generating procedure. In some embodiments, the fragmented DNA can be size selected to obtain nucleic acid fragments of a particular size range. Nucleic acid can be fragmented by various methods known in the art, which include without limitation, physical, chemical and enzymic processes. Examples of such processes are described in U.S. Patent Application Publication No. 20050112590 (published on May 26, 2005, entitled "Fragmentation-based methods and systems for sequence variation detection and discovery," naming Van Den Boom et al.). Certain processes can be selected to generate non-specifically cleaved fragments or specifically cleaved fragments. Examples of processes that can generate non-specifically cleaved fragment sample nucleic acid include, without limitation, contacting sample nucleic acid with apparatus that expose nucleic acid to shearing force (e.g., passing nucleic acid through a syringe needle; use of a French press); exposing sample nucleic acid to irradiation (e.g., gamma, x-ray, UV irradiation; fragment sizes can be controlled by irradiation intensity); boiling nucleic acid in water (e.g., yields about 500 base pair fragments) and exposing nucleic acid to an acid and base hydrolysis process.

Nucleic acid may be specifically cleaved by contacting the nucleic acid with one or more specific cleavage agents. The term "specific cleavage agent" as used herein refers to an agent, sometimes a chemical or an enzyme that can cleave a nucleic acid at one or more specific sites. Specific cleavage agents often will cleave specifically according to a particular nucleotide sequence at a particular site. Examples of enzyme specific cleavage agents include without limitation endonucleases (e.g., DNase (e.g., DNase I, II); RNase (e.g., RNase E, F, H, P); Cleavase™ enzyme; Taq DNA polymerase; E. coli DNA polymerase I and eukaryotic structure-specific endonucleases; murine FEN-1 endonucleases; type I, II or III restriction endonucleases such as Acc I, Afl III, Alu I, Alw44 I, Apa I, Asn I, Ava I, Ava II, BamH I, Ban II, Bcl I, Bgl I. Bgl II, Bln I, Bsm I, BssH II, BstE II, Cfo I, Cla I, Dde I, Dpn I, Dra I, EclX I, EcoR I, EcoR I, EcoR II, EcoR V, Hae II, Hae II, Hind II, Hind III, Hpa I, Hpa II, Kpn I, Ksp I, Mlu I, MluN I, Msp I, Nci I, Nco I, Nde I, Nde II, Nhe I, Not I, Nru I, Nsi I, Pst I, Pvu I, Pvu II, Rsa I, Sac I, Sal I, Sau3A I,

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1490 of 3077

Sca I, ScrF I, Sfi I, Sma I, Spe I, Sph I, Ssp I, Stu I, Sty I, Swa I, Taq I, Xba I, Xho I); glycosylases (e.g., uracil-DNA glycolsylase (UDG), 3-methyladenine DNA glycosylase, 3-methyladenine DNA glycosylase II, pyrimidine hydrate-DNA glycosylase, FaPy-DNA glycosylase, thymine mismatch-DNA glycosylase, hypoxanthine-DNA glycosylase, 5-Hydroxymethyluracil DNA glycosylase (HmUDG), 5-Hydroxymethylcytosine DNA glycosylase, or 1,N6-etheno-adenine DNA glycosylase); exonucleases (e.g., exonuclease III); ribozymes, and DNAzymes. Nucleic acid may be treated with a chemical agent, or synthesized using modified nucleotides, and the modified nucleic acid may be cleaved. In non-limiting examples, nucleic acid may be treated with (i) alkylating agents such as methylnitrosourea that generate several alkylated bases, including N3-methyladenine and N3-methylguanine, which are recognized and cleaved by alkyl purine DNA-glycosylase; (ii) sodium bisulfite, which causes deamination of cytosine residues in DNA to form uracil residues that can be cleaved by uracil N-glycosylase; and (iii) a chemical agent that converts guanine to its oxidized form, 8-hydroxyguanine, which can be cleaved by formamidopyrimidine DNA N-glycosylase. Examples of chemical cleavage processes include without limitation alkylation, (e.g., alkylation of phosphorothioate-modified nucleic acid); cleavage of acid lability of P3'-N5'-phosphoroamidate-containing nucleic acid; and osmium tetroxide and piperidine treatment of nucleic acid.

As used herein, the term “complementary cleavage reactions” refers to cleavage reactions that are carried out on the same nucleic acid using different cleavage reagents or by altering the cleavage specificity of the same cleavage reagent such that alternate cleavage patterns of the same target or reference nucleic acid or protein are generated. In certain embodiments, nucleic acids of interest may be treated with one or more specific cleavage agents (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more specific cleavage agents) in one or more reaction vessels (e.g., nucleic acid of interest is treated with each specific cleavage agent in a separate vessel).

A nucleic acid suitable for use in the embodiments described herein sometimes is amplified by any amplification process known in the art (e.g., PCR, RT-PCR and the like). Nucleic acid amplification may be particularly beneficial when using organisms that are typically difficult to culture (e.g., slow growing, require specialize culture conditions and the like). The terms “amplify”, “amplification”, “amplification reaction”, or “amplifying” as used herein refer to any in vitro processes for multiplying the copies of a target sequence of nucleic acid. Amplification sometimes refers to an “exponential” increase in target nucleic acid. However, “amplifying” as used herein can also refer to linear increases in the numbers of a select target sequence of

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1491 of 3077

nucleic acid, but is different than a one-time, single primer extension step. In some embodiments, a limited amplification reaction, also known as pre-amplification, can be performed. Pre-amplification is a method in which a limited amount of amplification occurs due to a small number of cycles, for example 10 cycles, being performed. Pre-amplification can allow some amplification, but stops amplification prior to the exponential phase, and typically produces about 500 copies of the desired nucleotide sequence(s). Use of pre-amplification may also limit inaccuracies associated with depleted reactants in standard PCR reactions.

In some embodiments, a nucleic acid reagent sometimes is stably integrated into the chromosome of a host cell or organism, or a nucleic acid reagent can be a deletion of a portion of a host chromosome, in certain embodiments (e.g., genetically modified cells or organisms, where alteration of the host genome confers the ability to selectively or preferentially maintain the desired cell or organism carrying the genetic modification). Such nucleic acid reagents (e.g., nucleic acids or genetically modified cells or organisms whose altered genome confers a selectable trait to the cell or organism) can be selected for their ability to guide production of a desired protein or nucleic acid molecule. When desired, the nucleic acid reagent can be altered such that codons encode for (i) the same amino acid, using a different tRNA than that specified in the native sequence, or (ii) a different amino acid than is normal, including unconventional or unnatural amino acids (including detectably labeled amino acids).

A nucleic acid or nucleic acid reagent can comprise certain elements often selected according to the intended use of the nucleic acid. Any of the following elements can be included in or excluded from a nucleic acid reagent. A nucleic acid reagent, for example, may include one or more or all of the following nucleotide elements: one or more promoters, one or more 5' untranslated regions (5'UTRs), one or more regions into which a target nucleotide sequence may be inserted (an "insertion element"), one or more target nucleotide sequences, one or more terminator elements, one or more 3' untranslated regions (3'UTRs), and one or more selection elements. A nucleic acid reagent can be provided with one or more of such elements and other elements may be inserted into the nucleic acid before the nucleic acid is introduced into the desired cell or organism. In some embodiments, a provided nucleic acid reagent comprises a promoter, 5'UTR, optional 3'UTR and insertion element(s) by which a target nucleotide sequence is inserted (i.e., cloned) into the nucleotide acid reagent. In certain embodiments, a provided nucleic acid reagent comprises a promoter, insertion element(s) and optional 3'UTR, and a 5' UTR/target nucleotide sequence is inserted with an optional 3'UTR. The elements can

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1492 of 3077

be arranged in any order suitable for expression in the chosen expression system (e.g., expression in a chosen cell or organism, or expression in a cell free system, for example), and in some embodiments a nucleic acid reagent comprises the following elements in the 5' to 3' direction: (1) promoter, 5'UTR, and insertion element(s); (2) promoter, 5'UTR, and target nucleotide sequence; (3) promoter, 5'UTR, insertion element(s) and 3'UTR; and (4) promoter, 5'UTR, target nucleotide sequence and 3'UTR.

Promoters

A promoter typically is required for cellular DNA synthesis and/or RNA synthesis. A promoter often contains a region of DNA that can facilitate the transcription of a particular gene, by providing a start site for the synthesis of RNA corresponding to a gene. Promoters generally are located near the genes they regulate, are located upstream (i.e., 5') of the START codon of the structural gene, and are on the same strand of DNA as the sense strand of the gene, in some instances. Eukaryotic promoters generally include a core promoter element that may contain a TATA box, a proximal sequence and transcription enhancer sequences positioned farther upstream (referred to, e.g., with respect to yeast, as upstream activating sequences or UAS located several hundred to thousands of kilobases upstream from a transcriptional start site (TSS)). The types and combination of these elements can influence promoter strength (see, e.g., Hussain et al. (2016) *ACS Synth. Biol*. 5:213-223). As used herein, "promoter," "promoter sequence" and "promoter region" are used interchangeably to refer to nucleotide sequences that can regulate gene transcription. Such sequences can include, but are not limited to, core promoter (e.g., extending upstream from the transcription START site (TSS)) elements (e.g., TATA box, RNA polymerase binding site, CCAAT box), proximal cis-acting sequences that bind proteins and can facilitate binding of RNA polymerase to DNA, and distant cis-regulatory sequences (e.g, enhancers and silencers) that can bind transcription factors and influence (e.g., activate, increase, elevate, decrease, reduce) transcription.

In some embodiments, a promoter sequence can be isolated from a nucleic acid or cell or organism and combined in functional connection or operable linkage with a polynucleotide sequence to allow altered and/or regulated expression. A non-native promoter (e.g., promoter not normally associated with a given nucleic acid sequence) used for expression of a nucleic acid often is referred to as a heterologous promoter. In certain embodiments, a heterologous promoter and/or a 5'UTR can be combined in functional connection with a polynucleotide that

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1493 of 3077

encodes a polypeptide having a desired activity as described herein. The terms "operably linked" and "in functional connection with" as used herein with respect to promoters, refer to a relationship between a nucleic acid coding sequence and a promoter element. The promoter is operably linked or in functional connection with the coding sequence when expression from the coding sequence via transcription is regulated, or controlled by, the promoter element. The terms "operably linked" and "in functional connection with" are utilized interchangeably herein with respect to promoter elements.

A promoter often interacts with a RNA polymerase. A polymerase is an enzyme that catalyzes synthesis of nucleic acids using a preexisting nucleic acid reagent. When the template is a DNA template, an RNA molecule is transcribed before protein is synthesized. Enzymes having polymerase activity suitable for use in the present methods include any polymerase that is active in the chosen system with the chosen template to synthesize protein. In some embodiments, a promoter (e.g., a heterologous promoter), can be operably linked to a nucleotide sequence or an open reading frame (ORF). Transcription from the promoter element can catalyze the synthesis of an RNA corresponding to the nucleotide sequence or ORF sequence operably linked to the promoter, which in turn leads to synthesis of a desired peptide, polypeptide or protein.

There are generally several types of promoters, e.g., constitutive, repressible and inducible. Constitutive promoters can be considered as unregulated (i.e., regulated essentially only by RNA polymerase levels) and provide for consistent expression of a gene that is under the transcriptional control of the promoter. Repressible and inducible promoters are regulatable by various cellular conditions. A repressible promoter is one that can be silenced, or "turned off," by the binding of a repressor molecule to a particular nucleotide sequence which serves to inhibit the functional interaction of RNA polymerase with the promoter and inhibits transcription. This is referred to as negative control or regulation and is in contrast to positive control of transcription which can occur via activator molecules binding to DNA and increasing the rate of transcription. An inducible promoter is one in which transcription can be induced in the presence of an effector molecule that, for example, binds to a regulatory transcription factor and results in increased rates of transcription. As used herein, a "non-inducible" promoter is a promoter that does not exhibit increased activity, in terms of transcription activation of an operably linked nucleic acid, in response to the presence of an effector or inducing agent. A non-inducible promoter can be one that is not induced by one agent but is induced by another.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1494 of 3077

For example, the transcription-regulating activity of a non-fatty acid-inducible promoter is not detectably increased in the presence of a fatty acid, although there may be other agents that do induce the promoter and increase transcription of a nucleic acid operably linked to the promoter.

Promoters sometimes exhibit responsiveness to regulatory control. Promoters also sometimes can be regulated by a selective agent. That is, transcription from promoters sometimes can be turned on, turned off, up-regulated or down-regulated, in response to a change in environmental, nutritional or internal conditions or signals (e.g., heat inducible promoters, light regulated promoters, feedback regulated promoters, hormone influenced promoters, tissue specific promoters, oxygen and pH influenced promoters, promoters that are responsive to selective agents (e.g., kanamycin) and promoters responsive to certain carbon sources (e.g., fatty acids) and the like, for example). Promoters influenced by environmental, nutritional or internal signals frequently are influenced by a signal (direct or indirect) that binds at or near the promoter and increases or decreases expression of the target sequence under certain conditions.

The strength of a promoter sequence can be measured as the amount of transcription of a gene product initiated at the promoter relative to a reference or control. For example, a reference or control can be the amount of transcription of the same gene product (e.g., a reporter gene product) initiated from a standard or reference promoter under the same conditions. In assessing the strength of an inducible promoter, the amount of transcription of a gene product that occurs from the promoter in the absence (non-inducing conditions) and presence (inducing conditions) of an inducing factor, or environment or condition, can be compared to determine the degree of inducibility. The difference in those transcription amounts can also be compared to the difference in transcription amounts under the same non-inducing and inducing conditions of a reference or control promoter to determine relative strength and inducibility. Methods for evaluating promoter strength using quantitative techniques for measuring gene product expression include, for example, RT-qPCR, northern blot techniques, and reporter gene product expression assays (see, e.g., Teste et al. (2009) *BMC Molecular Biology* 10:99; Wang et al. (2016) *Yeast* 33:99-106; Peng et al. (2015) Microb. Cell Fact. 14:91). For example, transcription (e.g., measured in ways known in the art) can sometimes be increased by at least about the following percentages when an inducible promoter controlling transcription of a nucleic acid is subjected to inducing conditions as compared to transcription under non-inducing conditions: by at least about 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 11%, 12%, 13%,

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1495 of 3077

14%, 15%, 16%, 17%, 18%, 19%, 20%, 21%, 22%, 23%, 24%, 25%, 26%, 27%, 28%, 29%, 30%, 31%, 32%, 33%, 34%, 35%, 36%, 37%, 38%, 39%, 40%, 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 49%, 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 100% or more. In some instances, transcription can sometimes be increased by at least about the following fold when an inducible promoter controlling transcription of a nucleic acid is subjected to inducing conditions as compared to transcription under non-inducing conditions: at least about 1.5-fold, 2-fold, 2.5-fold, 3-fold, 3.5-fold, 4-fold, 4.5-fold, 5-fold, 5.5-fold, 6-fold, 6.5-fold, 7-fold, 7.5-fold, 8-fold, 8.5-fold, 9-fold, 9.5-fold,10-fold, 11-fold, 12-fold, 13-fold, 14-fold, 15-fold, 20-fold or more.

In some instances, the use of a stronger heterologous inducible promoter to control transcription can increase the amount of induced transcription of a product-encoding nucleic acid by at least about the following percentages over the amount of induced transcription of the same nucleic acid controlled by a weaker inducible promoter: by at least about 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19%, 20%, 21%, 22%, 23%, 24%, 25%, 26%, 27%, 28%, 29%, 30%, 31%, 32%, 33%, 34%, 35%, 36%, 37%, 38%, 39%, 40%, 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 49%, 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 100% or more. In some instances, the use of a stronger heterologous inducible promoter to control transcription can increase the amount of induced transcription of a product-encoding nucleic acid by at least about the following fold over the amount of induced transcription of the same nucleic acid controlled by a weaker inducible promoter: by at least about 1.5-fold, 2-fold, 2.5-fold, 3-fold, 3.5-fold, 4-fold, 4.5-fold, 5-fold, 5.5-fold, 6-fold, 6.5-fold, 7-fold, 7.5-fold, 8-fold, 8.5-fold, 9-fold, 9.5-fold,10-fold, 11-fold, 12-fold, 13-fold, 14-fold, 15-fold, 20-fold or more.

In some embodiments, regulation of a promoter can be used to alter (e.g., increase, add, decrease or substantially eliminate) the activity of a peptide, polypeptide or protein (e.g., enzyme activity). For example, a cell or microorganism can be engineered by genetic modification to express a nucleic acid reagent that can add a novel activity (e.g., an activity not normally found in the host cell or organism) or increase the expression of an existing activity by

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1496 of 3077

increasing transcription from a homologous or heterologous promoter operably linked to a nucleotide sequence of interest (e.g., heterologous nucleotide sequence of interest), in certain embodiments. In some embodiments, a cell or microorganism can be engineered by genetic modification to express a nucleic acid reagent that can decrease expression of an activity by decreasing or substantially eliminating transcription from a homologous or heterologous promoter operably linked to a nucleotide sequence of interest, in certain embodiments. In some embodiments, an inducible heterologous promoter can be used to regulate transcription of a protein-encoding nucleic acid that is a stronger, or more strongly inducible, promoter than an endogenous inducible promoter that regulates expression of the protein-encoding nucleic acid in its endogenous state.

In some embodiments the activity can be altered using recombinant DNA and genetic techniques known to the artisan. Methods for engineering cells and microorganisms are further described herein. Also provided herein are non-limiting examples of regulated promoters, e.g., promoters that are up-regulated by oxygen, promoters that are down-regulated by oxygen, promoters that are repressed in the presence of certain carbon sources (e.g, glucose), promoters that are de-repressed under certain carbon source conditions (e.g., limited or depleted glucose and/or non-fermentable carbon sources), promoters that are induced in the presence of certain carbon sources (e.g., fatty acids), transcriptional repressors and their associated genes, DNA binding motifs as determined using the MEME sequence analysis software. Potential regulator binding motifs can be identified using the program MEME to search intergenic regions bound by regulators for overrepresented sequences. For each regulator, the sequences of intergenic regions bound with p-values less than 0.001 are extracted to use as input for motif discovery. The MEME software can be run, for example, using the following settings: a motif width ranging from 6 to 18 bases, the “zoops” distribution model, a 6th order Markov background model and a discovery limit of 20 motifs. The discovered sequence motifs can be scored for significance by two criteria: an E-value calculated by MEME and a specificity score. The motif with the best score using each metric is shown for each regulator.

Carbon source-dependent gene regulatory elements

Many cells and organisms, including, for example, many yeast species, preferentially use glucose over other carbon sources. Often, cell and organism growth is maximal in the presence

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1497 of 3077

of glucose. However, some cells and organisms are able to use alternative carbon sources for the production of metabolic energy and cellular biomass. In doing so, cellular metabolism can undergo substantial changes as certain pathways (such as, for example, oxidative metabolism, the TCA cycle, glyoxylate cycle and gluconeogenesis) required for utilizing non-glucose carbon sources are activated. Genes encoding such pathway-specific components can be subject to carbon source regulation of transcription. When glucose is present as a carbon source, some components (e.g. enzymes) of these other pathways may not be expressed, or are less expressed, because the pathways are not essential, or are used to a lesser extent, in the presence of glucose. This is referred to as glucose repression. Thus, in contrast to unregulated constitutive promoters, transcription regulatory elements for genes such as these are repressed, derepressible and/or inducible by varying carbon sources. When glucose is depleted, genes that were subject to glucose repression are then transcribed in a process referred to as glucose derepression. For some of these genes, this increase in transcription due to derepression represents the extent to which the genes will be expressed because they are not subject to induction and further increased transcription. For others of these genes, transcription may be increased (e.g., several-fold) over the derepressed level upon induction by, for example, certain carbon sources. Examples of such carbon sources include fatty acids (e.g, oleic acid) and n-alkanes. Some genes encoding peroxisomal proteins (including enzymes involved in fatty acid catabolism) are subject to glucose repression/derepression. *Cis*-acting regulatory elements have been identified for some of these genes. For example, sequences located upstream of the TATA boxes in the *Saccharomyces cerevisiae FOX1* gene encoding an acyl-CoA oxidase and *FOX3* gene encoding 3-oxoacyl-CoA thiolase have been reported as glucose response elements (see, e.g., Wang et al. (1992) *Nucleic Acids Res*. 20:3495; Wang et al. (1994) *J. Biol. Chem*. 269:24480; and Einerhand et al. (1991) *Eur. J. Biochem*. 200:113).

As described herein, engineered alteration of carbon flux in cells and organisms can involve directing internalized carbon sources toward particular cellular processing pathways and/or away from particular pathways. Some host cell modifications made in engineered bioproduction systems described herein can depend in part on the carbon source or sources used and the target molecule being produced. For example, in some embodiments provided herein, cells or organisms are modified to enhance carbon flux through oxidative metabolism pathways (e.g., β-oxidation and/or ω-oxidation) and/or fatty acid synthesis for production of organic acid, polyketide, terpene and/or other target molecules. In particular embodiments, the modified cells or organisms are provided with non-fermentable carbon sources (e.g., fatty acids, alkanes)

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1498 of 3077

and/or limited amounts of, or no, glucose for production of target molecule production. In these and other embodiments described herein, genetic modifications may be made to the cells or organisms to, for example, modify the amount and/or activity of one or more enzymes (e.g., acetyl-CoA carboxylase, ATP citrate lyase, carnitine acetyltransferase, acyl-CoA thioesterase, acetyl-CoA hydrolase, acetyl-CoA synthetase) in carbon-processing pathways. As also described herein, in some of these genetic modifications, it may be beneficial to use heterologous transcription-regulating nucleic acid elements that are differentially responsive to certain carbon sources for controlling expression of the enzyme(s). For example, promoters and other regulatory nucleic acid elements that are repressed when glucose is present, derepressed in glucose-limiting, or depleted, conditions and/or induced in the presence of alternative carbon sources can provide for optimized and regulatable production of target molecules, such as in embodiments involving use of non-glucose carbon sources. This is particularly useful in instances where target molecules may be toxic to cells or organisms in high levels. In this example, modified cells could initially be cultured in the presence of glucose, if desired to build up cell mass during a growth phase, and then switched to an alternative carbon source for target molecule production through engineered pathways during which time expression of modified enzymes would be derepressed and/or induced.

In some embodiments of the engineered cells and organisms provided herein, carbon flux alterations may include diversion of carbon atoms (e.g., acetyl groups) away from particular cellular pathways (e.g. the TCA cycle) to minimize carbon atom loss to those pathways at the expense of target molecule-producing pathways. In these instances, it may be beneficial to modify and/or replace promoters and other transcription regulatory elements that control expression of components (e.g., mitochondrial proteins such as carnitine acetyltransferase and carnitine transporters) of pathways not involved in target molecule production such that the components are not expressed, or are expressed at reduced levels, under glucose-limiting conditions and/or in the presence of alternative carbon sources. Heterologous transcription-regulatory nucleic acid elements suitable in achieving such control include, but are not limited to, weak, constitutive promoters and promoters that are repressed when non-glucose (or non-fermentable) carbon sources are available, derepressed when glucose is present and/or induced in the presence of glucose. Examples of such transcriptional control elements include, but are not limited to, promoter sequences regulating transcription of genes encoding phosphoglycerate kinase (PGK), glyceraldehyde-3-phosphate dehydrogenase (GPD), translation elongation factor (TEF) and glucose-6-phosphate isomerase (G6PI; also referred to

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1499 of 3077

as phosphoglucose isomerase or PGI). Nucleotide sequences for promoters of the *PGK* (SEQ ID NO: 114), *GPD* (SEQ ID NO: 119), *TEF* (SEQ ID NO: 120) and *G6PI* (SEQ ID NO: 118) genes of *Candida* strain ATCC 20336 are provided herein. Additional examples of sources of PGK, GPD, TEF and G6PI gene sequences include, but are not limited to: *Saccharomyces cerevisiae PGI1* (Genbank accession no. X13977), *Aspergillus oryzae PGK* (Genbank accession no. E04898), *Yarrowia lipolytica PGK* (Genbank accession no. M91598), *Candida albicans PGK* (Genbank accession no. U25180), *Candida maltosa C-PGK1* (Genbank accession no. D12474), *Saccharomyces cerevisiae GPD* (Genbank accession no. M13807), *Cyberlindnera jadinii GAP* (Genbank accession no. FJ664342), *Yarrowia lipolytica TEF1* (Genbank accession no. AF054508), *Debaryomyces hansenii TEF1* (Genbank accession no. AM600962).

Genes that are transcriptionally regulated by carbon source availability to cells (see, e.g., Turcotte et al. (2010) *FEMS Yeast Res.* 10:2-13; Weinhandl et al. (2014) *Microbial Cell Factories* 13(5):1-17) are possible sources of carbon source-dependent heterologous promoters for use in modification of cells and organisms as described herein. For example, genes encoding glycerol kinase and glycerol-3-phosphate dehydrogenase, such as the *GUT1* and *GUT2* genes of yeast, respectively, can be repressed in the presence of fermentable carbon sources such as glucose and expressed in the presence of non-fermentable carbon sources, e.g., glycerol or ethanol. When glucose is present, repression of the *Saccharomyces cerevisiae GUT1* and *GUT2* genes occurs in cells and is mediated by the negative regulator Opi1. The promoter region of the *Saccharomyces cerevisiae GUT1* gene contains two upstream transcription activation sequences, UAS_{ADR1} and UAS_{INO}, that can be binding sites for Adr1p (a zinc finger transcription factor) and Ino2p/Ino4p (basic helix-loop-helix factors), respectively, which are responsible for about 90% of the *GUT1* gene expression in the presence of glycerol (see, e.g., Grauslund et al. (1999) *Nucleic Acids Res.* 27(22):4391-4398). Similarly, the promoter region of the *S. cerevisiae GUT2* gene contains a UAS_{HAP} upstream sequence that can be a binding site for the Hap2/3/4/5 protein complex which activates transcription of several genes with mitochondrial functions (see, e.g., Grauslund and Ronnow (2000) *Can. J. Microbiol.* 46:1096-1100). The UAS_{HAP} element is required for full expression of the *GUT2* gene in the presence of glycerol or ethanol.

Additional examples of carbon source-dependent promoters include regulatory nucleic acid sequences controlling the transcription of genes encoding some polypeptides involved in fatty

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1500 of 3077

acid metabolism, peroxisomal transport/biogenesis and/or the glyoxylate cycle. Some of these genes undergo significant induction of transcription in the presence of fatty acids (e.g., oleic acid) and/or n-alkanes. This phenomenon is referred to as fatty acid or oleic acid (or oleate) induction or alkane induction. In yeast, some of the genes subject to fatty acid induction are transcriptionally controlled by promoter regions containing an oleate response element (ORE). In *Saccharomyces cerevisiae*, for example, the ORE of fatty acid-inducible gene promoters binds the positive transcription factor of zinc cluster proteins, Pip2p-Oaf1p. The promoter region of such genes typically contains a palindrome sequence of two CGG triplets with a sequence of 15-18 nucleotides between them that includes at least one half site containing a TNA triplet (where "N" represents any nucleotide) and thus has the sequence 5'-CGGNNNTNA(N_{9-12})CCG-3' (see, e.g., Gurvitz and Rottensteiner (2006) *Biochim. Biophys. Acta* 1763:1392-1402). Variants of this sequence in fatty acid-inducible *S. cerevisiae* gene promoters have been identified in connection with the *ANT1* and *PEX25* genes leading to the following sequence as being considered the minimal ORE: CGGNNNTN$^{A}/_{R}$(N_{8-12})CCG (see, e.g., Rottensteiner et al. (2003) *Eur. J. Biochem.* 270:2013-2022). Some of the promoter regions of fatty acid-inducible *S. cerevisiae* genes also include a type 1 upstream activation sequence (UAS1) having a consensus sequence of CYCCR(A/T/G)N_{4-36}(T/A/C)YGGRG that binds the Adr1 transcription factor and directly regulates some *S. cerevisiae* genes including *SPS19*, *POX1*, *CTA1*, *PEX11*, *PIP2* which encode peroxisomal proteins (e.g., peroxisomal 2,4-dienoyl-CoA reductase, or SPS19p, and Pex11p) and/or proteins involved in β-oxidation (e.g., acyl-CoA oxidase or Pox1p) or involved in the regulation of genes associated with fatty acid metabolism (see, e.g., Gurvitz et al. (2000) *Mol. Cell. Biol. Res. Commun.* 4:81-89; Gurvitz et al. (2001) *J. Biol. Chem.* 276:31825-31830; Rottensteiner et al. (2003) *J. Biol. Chem.* 278:27605-276110). Additional *S. cerevisiae* gene promoter regions containing an ORE include those controlling transcription of *MDH3* (peroxisomal malate dehydrogenase), *YCAT* (peroxisomal and mitochondrial carnitine acetyltransferase), *CRC1* (mitochondrial carnitine transporter) and *TES1* (peroxisomal thioesterase) (see, e.g., Karpichev and Small (1998) *Mol. Cell. Bio.* 18:6560-6570).

There are numerous DNA-binding factors and regulatory proteins involved in transcriptional regulation associated with carbon source utilization. For example, in yeast such as *S. cerevisiae*, glucose repression is mediated by repressors such as, for example, members of the Mig family of C_2H_2-zinc-finger DNA-binding proteins, and some zinc cluster proteins, e.g.,Oaf3. The promoter regions of genes subject to glucose repression typically contain a GC-rich recognition site (e.g., SYGGGG) to which a repressor, e.g., Mig1, binds in the presence of high

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1501 of 3077

levels of glucose (see, e.g., Gancedo (1998) *Microbiol. Mol. Biol. Rev*. 62(2):334-361). The repressor recruits a repressor complex, e.g., Ssn6-Tup1, resulting in conformational changes in the chromatin structure that prevent transcription initiation factors (e.g., the Sip4 and Adr1 activators of genes encoding, for example, gluconeogenic and glycerol or ethanol utilization proteins) from binding to sites in the DNA. Derepression occurring when glucose is depleted can result in activation of a protein kinase, Snf1, which participates in phosphorylation and release of the repressor complex thereby allowing for the activator to bind DNA in the promoter region. A shift from glucose-repressing to derepressing conditions typically results in an increased binding of Oaf1-Pip2 to oleate-response elements in promoters of fatty acid-inducible genes; however, in the presence of inducer (e.g., oleic acid), there may be only a marginal increase in this binding. This is because under derepressed conditions, Oaf1-Pip2 may be constitutively bound to target gene promoters. Activation of Oaf1-Pip2 involves binding of oleate to Oaf1 which is hyper-phosphorylated in the presence of oleate. In the activation of some fatty acid-inducible genes, Adr1 may also be involved. For example, promoter regions of genes encoding peroxisomal proteins often include a UAS1 that binds Adr1. DNA motifs for regulator protein (e.g., Adr1p, Hap2, Mig1) binding in carbon source-dependent promoters have been identified (see e.g., Weinhandl et al. (2014) *Microbial Cell Factories* 13(5):1-17) as have entire carbon source-dependent promoter nucleic acid sequences.

In other fungi, there can be different DNA-binding factors and regulatory proteins involved in transcriptional regulation associated with carbon source utilization. For example, in the filamentous fungus *Aspergillus nidulans*, glucose repression is mediated by the CreA repressor. In order to grow on fatty acids as a sole carbon source, this fungus typically requires two Zn_2Cys_6 proteins, FarAp and FarBp. These proteins are transcription factors that bind to a CCTCGG motif contained in the promoter region of genes encoding proteins involved in β-oxidation, peroxisomal functions and the glyoxylate cycle in this fungus. Specifically, FarAp is required for oxidation of short- and long-chain fatty acids and FarBp is required for oxidation of short-chain fatty acids. A homolog of FarA/FarB in *Candida albicans*, referred to as Ctf1p, is typically required for growth of *C. albicans* on fatty acids and regulates expression of some of the genes encoding proteins involved in β-oxidation.

Fatty acid- and/or alkane-inducible promoters from other organisms include, but are not limited to, those regulating transcription of the following genes: *Yarrowia lipolytica POX2* (acyl-CoA oxidase; Genbank accession no. AJ001300), *Yarrowia lipolytica POT1* (3-oxo-acyl-CoA

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1502 of 3077

thiolase; Genbank accession no. X69988), *Yarrowia lipolytica ICL1* (isocitrate lyase; Genbank accession no. CQ771439) and *Candida tropicalis HDE* (hydratase-dehydrogenase-epimerase; Genbank accession no. X57854), *Candida tropicalis POX4* (acyl-CoA oxidase; Genbank accession no. AB031271), *Candida tropicalis POX18* (peroxisomal 18-kDa protein; Genbank accession no. X53633), *Candida tropicalis SPS19* (2,4-dienoyl-CoA reductase; Genbank accession no. XM_002545237), *Candida albicans PEX11* (peroxisomal protein; Genbank accession no. XM_707009), *Candida tropicalis P450alk* (alkane-inducible cytochrome P450; Genbank accession no. M24894), and *Candida tropicalis CATL* (catalase; Genbank accession no. AB181391) (see, e.g., Hussain et al. (2016) *ACS Synth. Biol*. 5:213-223 and Sloots et al. (1991) *Gene* 105:129-134). Sequences of promoter elements of fatty acid-inducible genes (e.g., *HDE*, *POX4*, *PEX11*) from *Candida* strain ATCC 20336 are also provided herein (SEQ ID NOS: 113, 117 and 121).

The promoter region controlling transcription of the *Candida tropicalis* peroxisomal *HDE* gene includes a sequence similar to, but with deviations from, the *S. cerevisiae* ORE consensus sequence, and is as follows: $CGGNNNTTAN_{12}CAG$. This sequence, located in a region between nucleotides -393 and -341 (relative to the A nucleotide of the translation START codon), contains a 3' triplet of CAG in contrast to the CCG 3' triplet of the *S. cerevisiae* ORE consensus sequence. Specific nucleotides of the *C. tropicalis HDE* gene promoter ORE are CGGTTATTACGCCTGGGGGGGCAG. Similar sequences occur in the upstream promoter regions of *C. tropicalis* genes *POX4*, *POX18*, *P450alk* and *CATL* (see Sloots et al. (1991) *Gene* 105:129-134). The promoter regions for these genes (and the *HDE* gene) can also contain sequences similar to a 7-nucleotide consensus sequence ($ATTTCC_TC_T$) for regulation of the *S. cerevisiae SUC2* gene by glucose. This glucose-responsive region of the *C. tropicalis* HDE gene is located between nucleotides -526 and -393.

Alkane-assimilating organisms, such as, for example, *Candida tropicalis*, *Candida maltosa*, *Candida albicans*, *Candida bombicola*, *Candida parapsilosis*, *Yarrowia lipolytica*, *Pichia stipitis* and *Debaryomyces hansenii*, can utilize alkanes by first converting them to fatty alcohols through oxidation catalyzed by cytochrome P450. The fatty alcohols are then oxidized to fatty aldehydes which are in turn oxidized to fatty acids. Promoters for some of the genes in alkane-assimilating yeast have been shown to contain alkane-responsive elements. For example, an upstream activating sequence referred to as ARE1 and having a sequence $CTTGTGN_xCATGTG$ (where N represents any nucleotide and x refers to the number of

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1503 of 3077

nucleotides) has been identified as an alkane-responsive element present in the promoter of the *Yarrowia lipolytica ALK1* gene (cytochrome P450; Genbank accession no. AB010388) (see, e.g., Sumita et al (2002) *Biochem. Biophys. Res. Commun*. 294:1071-1078). Similar ARE1-like sequences (and/or conserved repeating motif: TGTG, or the CACA complement) occur in promoters of other genes encoding enzymes involved in alkane degradation, including, for example, cytochrome P450 genes of *Candida tropicalis*, (see, e.g., Seghezzi et al. (1992) *DNA Cell Biol*. 11:767-780), *Candida maltosa* (Genbank accession no. X55881), *Debaryomces hansenii* (Genbank accession no. AF103948) and also thiolase genes such as the acetoacetyl-CoA thiolase encoded by the *Y. lipolytica PAT1* gene (Genbank accession no. AB1020846) and the peroxisomal 3-ketoacyl-CoA thiolase encoded by the *C. tropicalis CT-T3A* gene (Genbank accession no. AB025647).

Possible additional candidate fatty acid- and/or alkane-inducible promoter sequences may be identified by searching genome databases for ORE consensus sequences located within about 500-1000 bp upstream of the START codon of an ORF and operably linking identified sequences with a reporter protein-encoding nucleic acid sequence for introduction into a host cell and analysis of reporter protein expression in the presence of varying carbon sources (such as fermentable and non-fermentable carbon sources and, in particular, fatty acids). Induced reporter protein expression in the presence of fatty acids and/or alkanes is indicative of a regulable, fatty acid- and/or alkane-inducible promoter sequence being linked to the reporter protein-encoding nucleic acid. Computer-assisted bioinformatics search programs are also available for use in identifying candidate transcription regulatory elements for genes (see, e.g., Worldwide Web uniform Resource Locator (URL) yeastract.com; Worldwide Web uniform Resource Locator (URL) pepper.molgenrug.nl/; Worldwide Web uniform Resource Locator (URL) rulai.cshl.edu/SCPD/; Worldwide Web uniform Resource Locator (URL) bimas.cit.nih.gov/molbio/proscan/; Worldwide Web uniform Resource Locator (URL) bioit.dmbr.ugent.be/contrav2/index.php).

Untranslated regions (UTR)

Nucleic acid reagents may also contain one or more 5' UTRs, and one or more 3'UTRs. Untranslated regions of a gene are sequences that are transcribed but are not translated into protein. A 5' UTR generally extends from the transcription start initiation site up to the first nucleotide of the translation START codon. A 3' UTR generally extends from the translation

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1504 of 3077

STOP codon to the polyA tail. Untranslated sequences can play important roles in post-transcriptional gene expression, including, for example, transport of a transcript out of the nucleus, translation efficiency, subcellular localization and mRNA stability.

A 5' UTR used in a nucleic acid reagent in genetically modifying cells may include one or more elements that are associated with it in an endogenous state, e.g., in a cell from which it originates, and sometimes includes one or more exogenous elements. A 5' UTR can originate from any suitable nucleic acid, such as genomic DNA, plasmid DNA, RNA or mRNA, for example, from any suitable organism (e.g., virus, bacterium, yeast, fungi, plant, insect or mammal). Appropriate elements for the 5' UTR can be selected based upon the chosen expression system (e.g., expression in a chosen organism, or expression in a cell free system, for example). A 5' UTR sometimes contains one or more of the following elements: enhancer sequences (e.g., translational), translation regulation site, translation initiation site, translation factor binding site, accessory protein binding site, feedback regulation agent binding sites, ribosome binding site, replicon, internal ribosome entry site (IRES), silencer element and the like. In some embodiments, a promoter element may be isolated such that all 5' UTR elements necessary for proper conditional regulation are contained in the promoter element fragment, or within a functional subsequence of a promoter element fragment.

A 5 'UTR in the nucleic acid reagent can include a translational enhancer nucleotide sequence. A translational enhancer nucleotide sequence often is located between the promoter and the target nucleotide sequence in a nucleic acid reagent. A translational enhancer sequence often binds to a ribosome, sometimes is an 18S rRNA-binding ribonucleotide sequence (i.e., a 40S ribosome binding sequence) and sometimes is an internal ribosome entry sequence (IRES). An IRES generally forms an RNA scaffold with precisely placed RNA tertiary structures that contact a 40S ribosomal subunit via a number of specific intermolecular interactions. Examples of ribosomal enhancer sequences are known and can be identified by the artisan (e.g., Mignone et al. (2005) *Nucleic Acids Research 33*: D141-D146; Paulous et al. (2003) *Nucleic Acids Research 31*:722-733; Akbergenov et al. (2004) *Nucleic Acids Research 32*:239-247; Mignone et al. (2002) *Genome Biology 3*(3): reviews0004.1-0001.10; Gallie (2002) *Nucleic Acids Research 30*:3401-3411; Shaloiko et al., Worldwide Web uniform Resource Locator (URL) interscience.wiley.com, DOI: 10.1002/bit.20267; and Gallie et al. (1987) *Nucleic Acids Research 15*:3257-3273).

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1505 of 3077

A translational enhancer sequence sometimes is a eukaryotic sequence, such as a Kozak consensus sequence or other sequence (e.g., hydroid polyp sequence, GenBank accession no. U07128). A translational enhancer sequence sometimes is a prokaryotic sequence, such as a Shine-Dalgarno consensus sequence. In certain embodiments, the translational enhancer sequence is a viral nucleotide sequence. A translational enhancer sequence sometimes is from a 5' UTR of a plant virus, such as Tobacco Mosaic Virus (TMV), Alfalfa Mosaic Virus (AMV); Tobacco Etch Virus (ETV); Potato Virus Y (PVY); Turnip Mosaic (poty) Virus and Pea Seed Borne Mosaic Virus, for example. In certain embodiments, an omega sequence about 67 bases in length from TMV is included in the nucleic acid reagent as a translational enhancer sequence (e.g., devoid of guanosine nucleotides and includes a 25 nucleotide long poly (CAA) central region).

A 3' UTR used in a nucleic acid reagent in genetically modifying cells may include one or more elements that are associated with it in an endogenous state, e.g., in a cell from which it originates, and sometimes includes one or more exogenous elements. A 3' UTR may originate from any suitable nucleic acid, such as genomic DNA, plasmid DNA, RNA or mRNA, for example, from any suitable organism (e.g., a virus, bacterium, yeast, fungi, plant, insect or mammal). Appropriate elements for the 3' UTR can be selected based upon the chosen expression system (e.g., expression in a chosen organism, for example). A 3' UTR sometimes comprises one or more of the following elements: translation regulation site, translation termination site, translation initiation site, translation factor binding site, ribosome binding site, replicon, enhancer element, silencer element and polyadenosine tail. A 3' UTR often includes a polyadenosine tail and sometimes does not, and if a polyadenosine tail is present, one or more adenosine moieties may be added or deleted from it (e.g., about 5, about 10, about 15, about 20, about 25, about 30, about 35, about 40, about 45 or about 50 adenosine moieties may be added or subtracted).

In some embodiments, modification of a 5' UTR and/or a 3' UTR can be used to alter (e.g., increase, add, decrease or substantially eliminate) gene expression activity. This can in turn alter the activity of a peptide, polypeptide or protein (e.g., enzyme activity for example), by a change in transcription of the nucleotide sequence(s) of interest from an operably linked promoter element comprising the modified 5' or 3' UTR. For example, a microorganism can be engineered by genetic modification to express a nucleic acid reagent comprising a modified 5' or 3' UTR that can add a novel activity (e.g., an activity not normally found in the host organism)

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1506 of 3077

or increase the expression of an existing activity by increasing transcription from a homologous or heterologous promoter operably linked to a nucleotide sequence of interest (e.g., homologous or heterologous nucleotide sequence of interest), in certain embodiments. In some embodiments, a microorganism can be engineered by genetic modification to express a nucleic acid reagent comprising a modified 5' or 3' UTR that can decrease the expression of an activity by decreasing or substantially eliminating transcription from a homologous or heterologous promoter operably linked to a nucleotide sequence of interest, in certain embodiments.

Nucleic acid or protein similarity

In addition to the nucleotide and amino acid sequences provided herein, a polynucleotide or polypeptide sequence may also be one that is substantially similar to those provided herein, including, but not limited to, promoter sequences, regulatory sequences, coding polynucleotides, amino acid signal sequences and amino acid protein sequences provided herein. Similarity between two nucleic acids or polypeptides refers to the relatedness between nucleotide sequences or amino acid sequences. Similarity can be based on the degree of identity and/or homology of sequences and the residues contained therein. Methods of assessing the degree of similarity between nucleic acids or proteins are known to those of skill in the art. For example, in one method of assessing sequence similarity, two nucleotide or amino acid sequences are aligned in a manner that yields a maximal level of identity between the sequences. Identity refers to the extent to which the sequences are invariant. Alignment of amino acid sequences, and, to some extent, nucleotide sequences, also can take into account conservative differences and/or frequent substitutions in amino acids (or nucleotides). Conservative differences are those that conserve the physico-chemical properties of the residues involved. Alignments can be global (alignment of the compared sequences over the entire length of the sequences and including all residues) or local (alignment of a portion of the compared sequences e.g., a portion or portions that includes only the most similar region or regions). Homology, with reference to polynucleotide or polypeptide sequences, refers to nucleotide or amino acid sequence similarity that takes into account identical residues and residues that can substitute for one another.

Percent identity and/or homology may be determined, for example, by comparing sequence information using any of a number of computer algorithms known in the art. In one example, calculations of sequence identity can be performed as follows. Sequences are aligned for

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1507 of 3077

optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment, and non-homologous sequences can be disregarded for comparison purposes). The length of a reference sequence aligned for comparison purposes is sometimes 30% or more, 40% or more, 50% or more, often 60% or more, and more often 70% or more, 80% or more, 90% or more, or 100% of the length of the reference sequence. The nucleotides or amino acids at corresponding nucleotide or polypeptide positions, respectively, are then compared among the two sequences. When a position in the first sequence is occupied by the same nucleotide or amino acid as the corresponding position in the second sequence, the nucleotides or amino acids are deemed to be identical at that position. The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, introduced for optimal alignment of the two sequences. Examples of sequence alignment and analysis software that can be used to calculate sequence identity include BLAST (Worldwide Web uniform Resource Locator (URL) blast.ncbi.nlm.nih.gov/Blast.cgi), MUSCLE (Worldwide Web uniform Resource Locator (URL) ebi.ac.uk/Tools/msa/muscle/ and Worldwide Web uniform Resource Locator (URL) drive5.com/muscle/) and MAFFT (Worldwide Web uniform Resource Locator (URL) mafft.cbrc.jp/alignment/server/ and Worldwide Web uniform Resource Locator (URL) ebi.ac.uk/Tools/msa/mafft/) for comparing nucleotide sequences and SIM (Worldwide Web uniform Resource Locator (URL) web.expasy.org/sim/) and BLAST for comparison of amino acid sequences. Nucleic acid sequence identity can also be determined by hybridization assays conducted under stringent conditions. As used herein, the term “stringent conditions” refers to conditions for hybridization and washing. Stringent conditions are known to those skilled in the art and can be found, for example, in Current Protocols in Molecular Biology, John Wiley & Sons, N.Y., 6.3.1-6.3.6 (1989). Aqueous and non-aqueous methods are described in that reference and either can be used. An example of stringent hybridization conditions is hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45ºC, followed by one or more washes in 0.2X SSC, 0.1% SDS at 50ºC. Another example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45ºC, followed by one or more washes in 0.2X SSC, 0.1% SDS at 55ºC. A further example of stringent hybridization conditions is hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45ºC, followed by one or more washes in 0.2X SSC, 0.1% SDS at 60ºC. Often, stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45ºC, followed by one or more washes in 0.2X SSC, 0.1% SDS at 65ºC. More often, stringency

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1508 of 3077

conditions are 0.5M sodium phosphate, 7% SDS at 65ºC, followed by one or more washes at 0.2X SSC, 1% SDS at 65ºC.

A nucleic acid or polypeptide for use in developing cells and organisms and/or methods described herein may be, for example, a polynucleotide or amino acid sequence that is homologous or identical to a nucleotide sequence (or complement thereof) or amino acid sequence provided herein over at least about 75%, or at least about 77%, or at least about 78%, or at least about 79%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 95% or more of the provided reference sequence. For example, a polynucleotide or polypeptide can be one that is at least about 50%, or at least about 51%, or at least about 52%, or at least about 54%, or at least about 55%, or at least about 58%, or at least about 60%, or at least about 62%, or at least about 65%, or at least about 70%, or at least about 75% or at least about 80% or more homologous or identical to a nucleic acid (or complement thereof) or polypeptide provided herein over the specified extent of a nucleic acid or polypeptide provided herein. In another embodiment, a nucleic acid or polypeptide can be one that is homologous or identical to a nucleic acid (or complement thereof) or polypeptide provided herein over at least about 86%, or at least about 87%, or at least about 88%, or at least about 89%, or at least about 92%, or at least about 97% or more of the provided reference nucleic acid (or complement thereof) or polypeptide. For example, the protein can be one that is at least about 67%, or at least about 68%, or at least about 69%, or at least about 72%, or at least about 77%, or at least about 82%, or at least about 87%, or at least about 90%, or at least about 95% or at least about 96%, or at least about 97%, or at least about 98%, or at least about 99% or more homologous or identical to a reference nucleic acid (or complement thereof) or polypeptide provided herein over the specified extent of the nucleic acid (or complement thereof) or polypeptide.

In some embodiments, a nucleotide or amino acid sequence that is at least 80% or more, 81% or more, 82% or more, 83% or more, 84% or more, 85% or more, 86% or more, 87% or more, 88% or more, 89% or more, 90% or more, 91% or more, 92% or more, 93% or more, 94% or more, 95% or more, 96% or more, 97% or more, 98% or more, or 99% or more identical to a nucleotide sequence (or complement thereof) or amino acid sequence described herein can be utilized. The term “identical” as used herein refers to two or more nucleotide or amino acid sequences having substantially the same nucleotide or amino acid sequence when compared to each other. One test for determining whether two nucleotide sequences or amino acids

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1509 of 3077

sequences are substantially identical is to determine the percent of identical nucleotide sequences or amino acid sequences shared.

Target nucleotide sequence

A nucleic acid reagent sometimes can comprise a target nucleotide sequence. A "target nucleotide sequence" as used herein encodes a nucleic acid, peptide, polypeptide or protein of interest, and may be a ribonucleotide sequence or a deoxyribonucleotide sequence. A target nucleic acid sometimes is an untranslated ribonucleic acid and sometimes is a translatable ribonucleic acid. An untranslated ribonucleic acid may include, but is not limited to, a small interfering ribonucleic acid (siRNA), a short hairpin ribonucleic acid (shRNA), other ribonucleic acid capable of RNA interference (RNAi), an antisense ribonucleic acid, or a ribozyme. A translatable target nucleotide sequence (e.g., a target ribonucleotide sequence) sometimes encodes a peptide, polypeptide or protein, which are sometimes referred to herein as "target peptides," "target polypeptides" or "target proteins."

Any peptides, polypeptides or proteins, or an activity catalyzed by one or more peptides, polypeptides or proteins, may be encoded by a target nucleotide sequence and may be selected by a user. Representative proteins include enzymes (e.g., acetyl-CoA carboxylase, acyl-CoA oxidase, thioesterase, monooxygenase, monooxygenase reductase, fatty alcohol oxidase, acyltransferase and the like, for example), antibodies, serum proteins (e.g., albumin), membrane bound proteins, hormones (e.g., growth hormone, erythropoietin, insulin, and the like), cytokines, and the like, and include both naturally occurring and exogenously expressed polypeptides. Representative activities (e.g., enzymes or combinations of enzymes which are functionally associated to provide an activity) include thioesterase activity, monooxygenase activity, monooxygenase reductase activity, acetyltransferase activity, omega hydroxyl fatty acid dehydrogenase activity, beta-oxidation activity, omega-oxidation activity and the like, for example. The term "enzyme" as used herein refers to a protein which can act as a catalyst to induce a chemical change in other compounds, thereby producing one or more products from one or more substrates.

Specific polypeptides (e.g., enzymes) useful for embodiments described herein are listed herein. The term "protein" as used herein refers to a molecule having a sequence of amino acids linked by peptide bonds. This term includes fusion proteins, oligopeptides, peptides,

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1510 of 3077

cyclic peptides, polypeptides and polypeptide derivatives, whether native or recombinant, and also includes fragments, derivatives, homologs, and variants thereof. A protein or polypeptide sometimes is of intracellular origin (e.g., located in the nucleus, cytosol, organelle (e.g., mitochondria or peroxisome) or interstitial space of host cells in vivo) and sometimes is a cell membrane protein in vivo. In some embodiments (described in further detail herein), a genetic modification can result in a modification (e.g., increase, substantially increase, decrease or substantially decrease) of a target activity and/or in a modification of a cellular location for a protein.

A translatable nucleotide sequence generally is located between a start codon (AUG in ribonucleic acids and ATG in deoxyribonucleic acids) and a stop codon (e.g., UAA (ochre), UAG (amber) or UGA (opal) in ribonucleic acids and TAA, TAG or TGA in deoxyribonucleic acids), and sometimes is referred to herein as an "open reading frame" (ORF). A translatable nucleotide sequence (e.g., ORF) sometimes is encoded differently in one organism (e.g., most organisms encode CTG as leucine) than in another organism (e.g., *Candida tropicalis* and *Candida viswanathii* encode CTG as serine). In some embodiments, a translatable nucleotide sequence is altered to correct alternate genetic code (e.g., codon usage) differences between a nucleotide donor organism and a nucleotide recipient organism (e.g., engineered organism). In certain embodiments, a translatable nucleotide sequence is altered to improve; (i) codon usage, (ii) transcriptional efficiency, (iii) translational efficiency, (iv) the like, and combinations thereof.

Nucleic acid reagents and tools

A nucleic acid reagent sometimes includes one or more ORFs. An ORF may be from any suitable source, sometimes from genomic DNA, mRNA, reverse transcribed RNA or complementary DNA (cDNA) or a nucleic acid library comprising one or more of the foregoing, and is from any organism species that contains a nucleic acid sequence of interest, protein of interest, or activity of interest. Non-limiting examples of organisms from which an ORF can be obtained include bacteria, yeast, fungi, plant, human, insect, nematode, bovine, equine, canine, feline, rat or mouse, for example.

A nucleic acid reagent sometimes contains a nucleotide sequence adjacent to an ORF that is translated in conjunction with the ORF and encodes an amino acid tag. The tag-encoding nucleotide sequence can be located 3' and/or 5' of an ORF in the nucleic acid reagent, thereby

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1511 of 3077

encoding a tag at the C-terminus or N-terminus of the protein or peptide encoded by the ORF. Any tag that does not abrogate in vitro transcription and/or translation may be utilized and may be appropriately selected by the artisan. Tags may facilitate isolation and/or purification of the desired ORF product from culture or fermentation media.

A tag sometimes specifically binds a molecule or moiety of a solid phase or a detectable label, for example, thereby having utility for isolating, purifying and/or detecting a protein or peptide encoded by the ORF. In some embodiments, a tag includes one or more of the following elements: FLAG (e.g., DYKDDDDKG), V5 (e.g., GKPIPNPLLGLDST), c-MYC (e.g., EQKLISEEDL), HSV (e.g., QPELAPEDPED), influenza hemaglutinin, HA (e.g., YPYDVPDYA), VSV-G (e.g., YTDIEMNRLGK), bacterial glutathione-S-transferase, maltose binding protein, a streptavidin- or avidin-binding tag (e.g., pcDNA™6 BioEase™ Gateway® Biotinylation System (ThermoFisher Scientific)), thioredoxin, β-galactosidase, VSV-glycoprotein, a fluorescent protein (e.g., green fluorescent protein or one of its many color variants (e.g., yellow, red, blue)), a polylysine or polyarginine sequence, a polyhistidine sequence (e.g., His6) or other sequence that chelates a metal (e.g., cobalt, zinc, copper), and/or a cysteine-rich sequence that binds to an arsenic-containing molecule. In certain embodiments, a cysteine-rich tag comprises the amino acid sequence CC-Xn-CC, where X is any amino acid and n is 1 to 3, and the cysteine-rich sequence sometimes is CCPGCC. In certain embodiments, the tag contains a cysteine-rich element and a polyhistidine element (e.g., CCPGCC and His6).

A tag often conveniently binds to a binding partner. For example, some tags bind to an antibody (e.g., FLAG) and sometimes specifically bind to a small molecule. For example, a polyhistidine tag specifically chelates a bivalent metal, such as copper, zinc and cobalt; a polylysine or polyarginine tag specifically binds to a zinc finger; a glutathione S-transferase tag binds to glutathione; and a cysteine-rich tag specifically binds to an arsenic-containing molecule. Arsenic-containing molecules include LUMIO™ agents (ThermoFisher Scientific), such as FlAsH™ (EDT2[4',5'-bis(1,3,2-dithioarsolan-2-yl)fluorescein-(1,2-ethanedithiol)2]) and ReAsH reagents (e.g., U.S. Patent 5,932,474 to Tsien et al., entitled “Target Sequences for Synthetic Molecules;” U.S. Patent 6,054,271 to Tsien et al., entitled “Methods of Using Synthetic Molecules and Target Sequences;” U.S. Patents 6,451,569 and 6,008,378; published U.S. Patent Application 2003/0083373, and published PCT Patent Application WO 99/21013, all to Tsien et al. and all entitled “Synthetic Molecules that Specifically React with Target

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1512 of 3077

Sequences"). Such antibodies and small molecules sometimes are linked to a solid phase for convenient isolation of the target protein or target peptide.

A tag sometimes includes a sequence that localizes a translated protein or peptide to a component in a system, which may be referred to as a "signal sequence," "targeting sequence" or "localization signal sequence" herein. A signal sequence often is incorporated at the N-terminus of a target protein or target peptide, and sometimes is incorporated at the C-terminus. Examples of signal sequences are known to the artisan, are readily incorporated into a nucleic acid reagent, and often are selected according to the organism in which expression of the nucleic acid reagent is performed. A signal sequence in some embodiments localizes a translated protein or peptide to a cell membrane. Examples of signal sequences include, but are not limited to, a nucleus targeting signal (e.g., steroid receptor sequence and N-terminal sequence of SV40 virus large T antigen); mitochondrial targeting signal (e.g., amino acid sequence that forms an amphipathic helix); peroxisome targeting signal (e.g., C-terminal sequence in YFG from *S.cerevisiae*); and a secretion signal (e.g., N-terminal sequences from invertase, mating factor alpha, PHO5 and SUC2 in *S.cerevisiae*; multiple N-terminal sequences of *B. subtilis* proteins (e.g., Tjalsma et al., *Microbiol.Molec. Biol. Rev*. 64: 515-547 (2000)); alpha amylase signal sequence (e.g., U.S. Patent No. 6,288,302); pectate lyase signal sequence (e.g., U.S. Patent No. 5,846,818); precollagen signal sequence (e.g., U.S. Patent No. 5,712,114); OmpA signal sequence (e.g., U.S. Patent No. 5,470,719); Iam beta signal sequence (e.g., U.S. Patent No. 5,389,529); *B. brevis* signal sequence (e.g., U.S. Patent No. 5,232,841); and *P. pastoris* signal sequence (e.g., U.S. Patent No. 5,268,273)).

A tag sometimes is directly adjacent to the amino acid sequence encoded by an ORF (i.e., there is no intervening sequence) and sometimes a tag is substantially adjacent to an ORF encoded amino acid sequence (e.g., an intervening sequence is present). An intervening sequence sometimes includes a recognition site for a protease, which is useful for cleaving a tag from a target protein or peptide. In some embodiments, the intervening sequence is cleaved by Factor Xa (e.g., recognition site I (E/D)GR), thrombin (e.g., recognition site LVPRGS), enterokinase (e.g., recognition site DDDDK), TEV protease (e.g., recognition site ENLYFQG) or PreScission™ protease (e.g., recognition site LEVLFQGP), for example.

An intervening sequence sometimes is referred to herein as a "linker sequence," and may be of any suitable length selected by the artisan. A linker sequence sometimes is about 1 to about 20

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1513 of 3077

amino acids in length, and sometimes about 5 to about 10 amino acids in length. The linker length can be selected to substantially preserve target protein or peptide function (e.g., a tag may reduce target protein or peptide function unless separated by a linker), to enhance disassociation of a tag from a target protein or peptide when a protease cleavage site is present (e.g., cleavage may be enhanced when a linker is present), and to enhance interaction of a tag/target protein product with a solid phase. A linker can be of any suitable amino acid content, and often comprises a higher proportion of amino acids having relatively short side chains (e.g., glycine, alanine, serine and threonine).

A nucleic acid reagent sometimes includes a stop codon between a tag element and an insertion element or ORF, which can be useful for translating an ORF with or without the tag. Mutant tRNA molecules that recognize stop codons suppress translation termination and thereby are designated “suppressor tRNAs.” Suppressor tRNAs can result in the insertion of amino acids and continuation of translation past stop codons (e.g., U.S. Patent Application No. 60/587,583, filed July 14, 2004, entitled "Production of Fusion Proteins by Cell-Free Protein Synthesis,"; Eggertsson, et al., (1988) *Microbiological Review* 52(3):354-374, and Engleerg-Kukla, et al. (1996) in *Escherichia coli* and *Salmonella* Cellular and Molecular Biology, Chapter 60, pps 909-921, Neidhardt, et al. eds., ASM Press, Washington, DC). A number of suppressor tRNAs are known, including but not limited to, supE, supP, supD, supF and supZ suppressors, which suppress the termination of translation of the amber stop codon; supB, glT, supL, supN, supC and supM suppressors, which suppress the function of the ochre stop codon and glyT, trpT and Su-9 suppressors, which suppress the function of the opal stop codon. In general, suppressor tRNAs contain one or more mutations in the anti-codon loop of the tRNA that allows the tRNA to base pair with a codon that ordinarily functions as a stop codon. The mutant tRNA is charged with its cognate amino acid residue and the cognate amino acid residue is inserted into the translating polypeptide when the stop codon is encountered. Mutations that enhance the efficiency of termination suppressors (i.e., increase stop codon read-through) have been identified. These include, but are not limited to, mutations in the uar gene (also known as the prfA gene), mutations in the ups gene, mutations in the sueA, sueB and sueC genes, mutations in the rpsD (ramA) and rpsE (spcA) genes and mutations in the rplL gene.

Thus, a nucleic acid reagent containing a stop codon located between an ORF and a tag can yield a translated ORF alone when no suppressor tRNA is present in the translation system, and can yield a translated ORF-tag fusion when a suppressor tRNA is present in the system.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1514 of 3077

Suppressor tRNA can be generated in cells transfected with a nucleic acid encoding the tRNA (e.g., a replication incompetent adenovirus containing the human tRNA-Ser suppressor gene can be transfected into cells, or a YAC containing a yeast or bacterial tRNA suppressor gene can be transfected into yeast cells, for example). Vectors for synthesizing suppressor tRNA and for translating ORFs with or without a tag are available to the artisan (e.g., Tag-On-Demand™ kit (ThermoFisher Scientific); Tag-On-Demand™ Suppressor Supernatant Instruction Manual, Version C, 31 October 2010, World Wide Web Uniform Resource Locator (URL) tools.thermofisher.com/content/sfs/manuals/tagondemand_supernatant_man.pdf; Tag-On-Demand™ Gateway® Vector Instruction Manual, Version D, 31 October 2010 World Wide Web Uniform Resource Locator (URL) tools.thermofisher.com/content/sfs/manuals/tagondemand_vectors_man.pdf; and Capone et al. (1985) Amber, ochre and opal suppressor tRNA genes derived from a human serine tRNA gene. *EMBO J.* 4:213).

Any convenient cloning strategy known in the art may be utilized to incorporate an element, such as an ORF, into a nucleic acid reagent. Known methods can be utilized to insert an element into the template independent of an insertion element, such as (1) cleaving the template at one or more existing restriction enzyme sites and ligating an element of interest and (2) adding restriction enzyme sites to the template by hybridizing oligonucleotide primers that include one or more suitable restriction enzyme sites and amplifying by polymerase chain reaction. Other cloning strategies take advantage of one or more insertion sites present or inserted into the nucleic acid reagent, such as an oligonucleotide primer hybridization site for PCR, for example, and others described herein. In some embodiments, a cloning strategy can be combined with genetic manipulation such as recombination (e.g., recombination of a nucleic acid reagent with a nucleic acid sequence of interest into the genome of the organism that is modified, as described further herein). In some embodiments, the cloned ORF(s) can produce (directly or indirectly), for example, a fatty acid or dicarboxylic acid (e.g., adipic acid, octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid), 3-hydroxyproprionic acid, triacetic acid lactone, by engineering a cell or microorganism with one or more ORFs of interest, which cell or microorganism may include one or more altered activities such as, for example, carnitine acetyltransferase activity, acetyl-CoA carboxylase activity, ATP citrate lyase activity, acetyl-CoA synthetase activity, cytochrome P450 reductase activity, acetyl-CoA hydrolase activity, 6-oxohexanoic acid dehydrogenase activity, 6-hydroxyhexanoic acid dehydrogenase activity,

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1515 of 3077

glucose-6-phosphate dehydrogenase activity, hexanoate synthase activity, lipase activity, fatty acid synthase activity, omega hydroxyl fatty acid dehydrogenase activity, acyl-CoA oxidase activity, acyltransferase activity, thioesterase activity, monooxygenase activity and monooxygenase reductase activity.

In some embodiments, a nucleic acid reagent includes one or more recombinase insertion sites. A recombinase insertion site is a recognition sequence on a nucleic acid molecule that participates in an integration/recombination reaction by recombination proteins. For example, the recombination site for Cre recombinase is loxP, which is a 34 base pair sequence comprised of two 13 base pair inverted repeats (serving as the recombinase binding sites) flanking an 8 base pair core sequence (e.g., Figure 1 of Sauer, B., *Curr. Opin. Biotech. 5*:521-527 (1994)). Other examples of recombination sites include attB, attP, attL, and attR sequences, and mutants, fragments, variants and derivatives thereof, which are recognized by the recombination protein λ Int and by the auxiliary proteins integration host factor (IHF), FIS and excisionase (Xis) (e.g., U.S. Patent Nos. 5,888,732; 6,143,557; 6,171,861; 6,270,969; 6,277,608; and 6,720,140; U.S. Patent Appln. Nos. 09/517,466, filed March 2, 2000, and 09/732,914, filed August 14, 2003, and in U.S. patent publication no. 2002-0007051-A1; Landy, Curr. Opin. Biotech. 3:699-707 (1993)).

Examples of recombinase cloning nucleic acids are in Gateway® systems (ThermoFisher Scientific), which include at least one recombination site for cloning a desired nucleic acid molecules in vivo or in vitro. In some embodiments, the system utilizes vectors that contain at least two different site-specific recombination sites, often based on the bacteriophage lambda system (e.g., att1 and att2), and are mutated from the wild-type (att0) sites. Each mutated site has a unique specificity for its cognate partner att site (i.e., its binding partner recombination site) of the same type (for example attB1 with attP1, or attL1 with attR1) and will not cross-react with recombination sites of the other mutant type or with the wild-type att0 site. Different site specificities allow directional cloning or linkage of desired molecules thus providing desired orientation of the cloned molecules. Nucleic acid fragments flanked by recombination sites are cloned and subcloned using the Gateway® system by replacing a selectable marker (for example, ccdB) flanked by att sites on the recipient plasmid molecule, sometimes termed the Destination Vector. Desired clones are then selected by transformation of a ccdB sensitive host strain and positive selection for a marker on the recipient molecule. Similar strategies for

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1516 of 3077

negative selection (e.g., use of toxic genes) can be used in other organisms such as thymidine kinase (TK) in mammals and insects.

A recombination system useful for engineering yeast is outlined briefly. The system makes use of the *URA3* gene (e.g., for *Candida tropicalis*, *Candida viswanathii*, *Saccharomyces cerevisiae* and *Candida albicans*, for example) or *URA4* and *URA5* genes (e.g., for *S. pombe*, for example) and toxicity of the nucleotide analogue 5-Fluoroorotic acid (5-FOA). The *URA3* or *URA4* and *URA5* genes encode orotidine-5'-monophosphate (OMP) dicarboxylase. Yeast with an active *URA3* or *URA4* and *URA5* gene (phenotypically Ura+) convert 5-FOA to fluorodeoxyuridine, which is toxic to yeast cells. Yeast carrying a mutation in the appropriate gene(s) or having a knock out of the appropriate gene(s) can grow in the presence of 5-FOA, if the media is also supplemented with uracil.

A nucleic acid engineering construct can be made which may contain the *URA3* gene or cassette (for *C. tropicalis*, *C. viswanathii* or *S. cerevisiae*, for example), flanked on either side by the same nucleotide sequence in the same orientation. The *URA3* cassette can include a promoter, the *URA3* gene and a functional transcription terminator. Target sequences which direct the construct to a particular nucleic acid region of interest in the cell or organism to be engineered are added such that the target sequences are adjacent to and abut the flanking sequences on either side of the *URA3* cassette. Yeast can be transformed with the engineered construct and plated on minimal media without uracil. Colonies can be screened by PCR to determine those transformants that have the engineering construct inserted in the proper location in the genome. Checking insertion location prior to selecting for recombination of the ura3 cassette may reduce the number of incorrect clones carried through to later stages of the procedure. Correctly inserted transformants can then be grown and plated on minimal media containing 5-FOA to select for recombination of the URA3 cassette out of the construct, leaving a disrupted gene and an identifiable footprint (e.g., nucleic acid sequence) that can be used to verify the presence of the disrupted gene. The technique described is useful for disrupting or "knocking out" gene function, but also can be used to insert genes or constructs into a host cell genome in a targeted, sequence specific manner.

In certain embodiments, a nucleic acid reagent includes one or more topoisomerase insertion sites. A topoisomerase insertion site is a defined nucleotide sequence recognized and bound by a site-specific topoisomerase. For example, the nucleotide sequence 5'-(C/T)CCTT-3' is a

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1517 of 3077

topoisomerase recognition site bound specifically by most poxvirus topoisomerases, including vaccinia virus DNA topoisomerase I. After binding to the recognition sequence, the topoisomerase cleaves the strand at the 3'-most thymidine of the recognition site to produce a nucleotide sequence comprising 5'-(C/T)CCTT-PO4-TOPO, a complex of the topoisomerase covalently bound to the 3' phosphate via a tyrosine in the topoisomerase (e.g., Shuman (1991) *J. Biol. Chem. 266*:11372-11379; Sekiguchi and Shuman (1994) *Nucl. Acids Res*. 22:5360-5365; U.S. Pat. No. 5,766,891; PCT/US95/16099; and PCT/US98/12372). In comparison, the nucleotide sequence 5'-GCAACTT-3' is a topoisomerase recognition site for type IA *E. coli* topoisomerase III. An element that is inserted often is combined with topoisomerase-reacted template and thereby incorporated into the nucleic acid reagent (e.g., World Wide Web Uniform Resource Locator (URL) tools.thermofisher.com/downloads/F-13512_Topo_Flyer.pdf; World Wide Web Uniform Resource Locator (URL) tools.thermofisher.com/content/sfs/brochures/topo-pcr-cloning-brochure.pdf; TOPO TA Cloning® Kit and Zero Blunt® TOPO® Cloning Kit product information).

A nucleic acid reagent sometimes contains one or more origin of replication (ORI) elements. In some embodiments, a template comprises two or more ORIs, where one functions efficiently in one organism (e.g., a bacterium) and another functions efficiently in another organism (e.g., a eukaryote, like yeast for example). In some embodiments, an ORI may function efficiently in one species (e.g., *S. cerevisiae*, for example) and another ORI may function efficiently in a different species (e.g., *S. pombe*, for example). A nucleic acid reagent also sometimes includes one or more transcription regulation sites.

A nucleic acid reagent can include one or more selection elements (e.g., elements for selection of the presence of the nucleic acid reagent, and not for activation of a promoter element which can be selectively regulated). Selection elements often are utilized using known processes to determine whether a nucleic acid reagent is included in a cell. In some embodiments, a nucleic acid reagent includes two or more selection elements, where one functions efficiently in one organism and another functions efficiently in another organism. Examples of selection elements include, but are not limited to, (1) nucleic acid segments that encode products that provide resistance against otherwise toxic compounds (e.g., antibiotics); (2) nucleic acid segments that encode products that are otherwise lacking in the recipient cell (e.g., essential products, tRNA genes, auxotrophic markers); (3) nucleic acid segments that encode products that suppress the activity of a gene product; (4) nucleic acid segments that encode products that can be readily

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1518 of 3077

identified (e.g., phenotypic markers such as antibiotics (e.g., β-lactamase), β-galactosidase, green fluorescent protein (GFP), yellow fluorescent protein (YFP), red fluorescent protein (RFP), cyan fluorescent protein (CFP), and cell surface proteins); (5) nucleic acid segments that bind products that are otherwise detrimental to cell survival and/or function; (6) nucleic acid segments that otherwise inhibit the activity of any of the nucleic acid segments described in Nos. 1-5 above (e.g., antisense oligonucleotides); (7) nucleic acid segments that bind products that modify a substrate (e.g., restriction endonucleases); (8) nucleic acid segments that can be used to isolate or identify a desired molecule (e.g., specific protein binding sites); (9) nucleic acid segments that encode a specific nucleotide sequence that can be otherwise non-functional (e.g., for PCR amplification of subpopulations of molecules); (10) nucleic acid segments that, when absent, directly or indirectly confer resistance or sensitivity to particular compounds; (11) nucleic acid segments that encode products that either are toxic or convert a relatively non-toxic compound to a toxic compound (e.g., Herpes simplex thymidine kinase, cytosine deaminase) in recipient cells; (12) nucleic acid segments that inhibit replication, partition or heritability of nucleic acid molecules that contain them; and/or (13) nucleic acid segments that encode conditional replication functions, e.g., replication in certain hosts or host cell strains or under certain environmental conditions (e.g., temperature, nutritional conditions, and the like). In some embodiments, the regulatory or selective agent can be added to change the existing growth conditions to which a cell or organism is subjected (e.g., growth in liquid culture, growth in a fermenter, growth on solid nutrient plates and the like for example).

A nucleic acid reagent can sometimes be of any form useful for in vivo transcription and/or translation. A nucleic acid sometimes is a plasmid, such as a supercoiled plasmid, sometimes is a yeast artificial chromosome (e.g., YAC), sometimes is a linear nucleic acid (e.g., a linear nucleic acid produced by PCR or by restriction digest), sometimes is single-stranded and sometimes is double-stranded. A nucleic acid reagent sometimes is prepared by an amplification process, such as a polymerase chain reaction (PCR) process or transcription-mediated amplification process (TMA). In TMA, two enzymes are used in an isothermal reaction to produce amplification products detected by light emission (see, e.g., *Biochemistry* 1996 Jun 25;35(25):8429-38). Standard PCR processes are known (e.g., U. S. Patent Nos. 4,683,202; 4,683,195; 4,965,188; and 5,656,493), and generally are performed in cycles. Each cycle includes heat denaturation, in which hybrid nucleic acids dissociate; cooling, in which primer oligonucleotides hybridize; and extension of the oligonucleotides by a polymerase (i.e., Taq polymerase). An example of a PCR cyclical process is treating the sample at 95°C for 5

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1519 of 3077

minutes; repeating forty-five cycles of 95°C for 1 minute, 59°C for 1 minute, 10 seconds, and 72°C for 1 minute 30 seconds; and then treating the sample at 72°C for 5 minutes. Multiple cycles frequently are performed using a commercially available thermal cycler. PCR amplification products sometimes are stored for a time at a lower temperature (e.g., at 4°C) and sometimes are frozen (e.g., at –20°C) before analysis.

In some embodiments, a nucleic acid reagent, protein reagent, protein fragment reagent or other reagent described herein is isolated or purified. The term “isolated” as used herein refers to material removed from its original environment (e.g., the natural environment if it is naturally occurring, or a host cell if expressed exogenously), and thus is altered "by the hand of man" from its original environment. The term “purified” as used herein with reference to molecules does not refer to absolute purity. Rather, “purified” refers to a substance in a composition that contains fewer substance species in the same class (e.g., nucleic acid or protein species) other than the substance of interest in comparison to the sample from which it originated. “Purified,” if a nucleic acid or protein for example, refers to a substance in a composition that contains fewer nucleic acid species or protein species other than the nucleic acid or protein of interest in comparison to the sample from which it originated. Sometimes, a protein or nucleic acid is “substantially pure,” indicating that the protein or nucleic acid represents at least 50% of protein or nucleic acid on a mass basis of the composition. Often, a substantially pure protein or nucleic acid is at least 75% on a mass basis of the composition, and sometimes at least 95% on a mass basis of the composition.

Genetic engineering methods

Methods and compositions (e.g., nucleic acid reagents) described herein can be used to generate modified or engineered cells or organisms. For example, a cell or organism can be modified by altering one or more cellular activities and/or the sum total of a cell’s or organism’s activities. Modifications can be, for example, any alteration of cellular activities, including addition of cellular activities and/or elimination of cellular activities. The term "altered activity" as used herein refers to an activity in an engineered cell or microorganism that is added, removed or modified in any way relative to the host cell or microorganism (e.g., added, increased, reduced, decreased, inhibited, removed or redirected activity). In some embodiments, the methods and nucleic acid reagents described herein can be used to generate genetically modified cells and organisms with altered activities in cellular carbon processing.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1520 of 3077

For example, the methods of genetic modification can be used to alter fatty acid (e.g., oleic acid, adipic acid, sebacic acid, suberic acid, octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid) synthesis and/or catabolism. In some embodiments, an engineered cell or organism described herein may include an increased number of copies of one or more polynucleotides encoding one or more polypeptides having carnitine acetyltransferase, acetyl-CoA carboxylase, ATP citrate lyase, thioesterase, acetyl-CoA hydrolase, acetyl-CoA synthetase, acyl-CoA oxidase, cytochrome P450 reductase, monooxygenase, peroxisomal biogenesis factor, alcohol dehydrogenase, alcohol oxidase, aldehyde dehydrogenase, 3-ketoacyl-CoA thiolase, and/or multifunctional enzyme (e.g., enoyl-CoA hydratase and/or 3-hydroxyacyl-CoA dehydrogenase) activity. In certain embodiments, an engineered cell or microorganism described herein may include one or more genetic modifications that reduce one or more of the following activities: carnitine acetyltransferase (e.g., mitochondrial), acetyl-carnitine translocase (e.g., mitochondrial), acyl-CoA synthetase, acyl-CoA oxidase and peroxisomal transporter activity.

In some embodiments, the engineered cell or organism can be a prokaryote. In certain embodiments, the prokaryote can be a bacterium, e.g., *Escherichia coli*. In some embodiments, the engineered cell or organism can be a eukaryote. In some embodiments, the eukaryote may be a fungus. In certain embodiments, the eukaryote can be a yeast. In certain embodiments, the yeast can be a *Candida* yeast. In some embodiments, the *Candida* yeast may be *C. viswanathii* or *C. troplicalis*. In certain embodiments, the fungus can be a *Yarrowia* fungus. In some embodiments the *Yarrowia* fungus may be *Y. lipolytica*. In some embodiments, the fungus can be a *Blastobotrys* yeast, e.g., *B. adeninivorans*. In certain embodiments, the fungus can be an *Aspergillus* fungus. In some embodiments, the *Aspergillus* fungus may be *A. parasiticus* or A. *nidulans*.

In some embodiments, an activity and/or amount of a composition can be altered by genetically modifying a host cell or organism which yields an engineered cell or organism having added, increased, reduced, decreased, inhibited, redirected, removed and/or otherwise modified activity or composition. A cell or organism may be modified, for example, by altering the amount of one or more cellular compositions, e.g, polynucleotides and/or polypeptides. Engineered cells or organisms typically arise as a result of a genetic modification, usually introduced by one of skill in the art using readily available techniques. Such cells or organisms are referred to herein as genetically modified or genetically engineered cells, microorganisms or organisms. The term

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1521 of 3077

"genetic modification" as used herein refers to any alteration in the genetic make-up of a cell or organism, including, for example, any nucleic acid addition, removal or alteration. Genetic modifications include, without limitation, insertion of one or more nucleotides in an endogenous nucleic acid of a host cell or organism in one or more locations, deletion of one or more nucleotides in an endogenous nucleic acid of a host cell or organism in one or more locations, modification or substitution of one or more nucleotides in an endogenous nucleic acid of a host cell or organism in one or more locations. In some embodiments, a portion of a host genome can be replaced with a heterologous nucleic acid. A genetic modification can also be insertion of a nucleic acid into a host cell organism that is distinct from the host endogenous genome (e.g., insertion of an autonomously replicating vector), and removal of a nucleic acid that is distinct from the endogenous host genome (e.g., removal of a vector).

Non-limiting examples of methods useful for genetically modifying a cell or organism include, introducing a heterologous polynucleotide (e.g., nucleic acid or gene integration, also referred to as “knock in”), removing an endogenous polynucleotide, altering the sequence of an existing endogenous nucleic acid sequence (e.g., site-directed mutagenesis), disruption of an existing endogenous nucleic acid sequence (e.g., knock outs and transposon or insertion element mediated mutagenesis), selection for an altered activity where the selection causes a change in a naturally occurring activity that can be stably inherited (e.g., causes a change in a nucleic acid sequence in the genome of the cell or organism or in an epigenetic nucleic acid that is replicated and passed on to daughter cells), PCR-based mutagenesis, and the like. The terms “mutant” and “mutagenesis” as used herein refer to any modification to a nucleic acid (e.g., nucleic acid reagent or host chromosome) and/or polypeptide which results in an altered nucleic acid and/or polypeptide. Non-limiting examples of mutagenesis include, deletion, insertion, substitution, rearrangement, point mutations, suppressor mutations and the like of a single or multiple residues in a polynucleotide. Mutagenesis methods are known in the art and are readily available to the artisan. Non-limiting examples of mutagenesis methods are described herein and can also be found in Maniatis, T., E. F. Fritsch and J. Sambrook (1982) *Molecular Cloning: a Laboratory Manual*; Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. Another non-limiting example of mutagenesis can be conducted using an Agilent (Santa Clara, CA) “QuickChange” kit according to the manufacturer’s instructions.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1522 of 3077

Decreasing an amount of a composition and/or activity in a cell

An altered activity or composition sometimes is an activity or composition detectable in a host cell or organism and that is reduced, decreased, inhibited or removed (i.e., not detectable) in an engineered cell or organism. For example, a genetic modification that disrupts cellular synthesis of a composition (e.g., acyl-CoA synthetase protein) and/or or disrupts an activity, such as activation of fatty acids, or disrupts a polynucleotide that encodes a polypeptide that carries out a forward reaction in the activity (e.g., acyl-CoA synthetase activity), may render the composition (e.g., acyl-CoA synthetase protein) or activity, such as fatty acid activation, undetectable. The term "undetectable" as used herein refers to an amount of an analyte (including an activity) that is below the limits of detection, using know detection methods or assays (e.g., described herein). In certain embodiments, the genetic modification may partially reduce or decrease a composition or an activity. The term "reduces" or "decreases" with reference to a composition or an activity as used herein refers to a level of the composition or activity in an engineered cell or organism that is lower than the level of the composition or activity found in the host or starting cell or organism. A "lower" level can be a level that is detectable or undetectable. The term "partially reduces" or "partially decreases" with reference to a composition or an activity as used herein refers to a level of the composition or activity in an engineered cell or organism that is lower than the level of the composition or activity found in the host or starting cell or organism but that is still detectable. Thus, an activity or composition can be reduced to undetectable levels in some embodiments, or detectable levels in certain embodiments. An activity or composition can be decreased to any suitable level for production of a target molecule product (e.g., an organic acid), including but not limited to less than 2-fold (e.g., about 10% decrease to about 99% decrease; about 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% decrease), 2-fold, 3-fold, 4-fold, 5-fold, 6-fold, 7-fold, 8-fold, 9-fold, of 10-fold decrease, or greater than about 10-fold decrease.

The term "level", as used herein, often refers to an amount (e.g., a quantitative or relative amount) of a nucleic acid (e.g. an RNA (e.g. an mRNA) or DNA), polypeptide or activty.

In some embodiments, an activity or composition may be reduced or removed by decreasing the number of copies of a polynucleotide that encodes a composition polypeptide or polypeptide having a target activity. In some embodiments, an activity or composition can be reduced or removed by (i) inserting a polynucleotide within a polynucleotide that encodes a protein having

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1523 of 3077

the target activity or the target composition (disruptive insertion), and/or (ii) removing a portion of or all of a polynucleotide that encodes a polypeptide having the target activity or the target composition (deletion or knock out, respectively). In certain embodiments, an activity or composition can be reduced or removed by inserting into a host cell or microorganism a heterologous polynucleotide that is (i) operably linked to another polynucleotide that encodes a polypeptide having the target activity or target composition, and (ii) down regulates production of the polypeptide. Thus, an activity or composition can be reduced or removed by inserting or modifying a regulatory polynucleotide operably linked to another polynucleotide that encodes a polypeptide having the target activity or target composition.

An activity or composition also can be reduced or removed by (i) inhibiting a polynucleotide that encodes a polypeptide having the activity or the targeted composition or (ii) inhibiting a polynucleotide operably linked to another polynucleotide that encodes a polypeptide having the activity or targeted composition. A polynucleotide can be inhibited by a suitable technique known in the art, such as by contacting an RNA encoded by the polynucleotide with a specific inhibitory RNA (e.g., RNAi, siRNA, ribozyme). An activity also can be reduced or removed by contacting a polypeptide having the activity with a molecule that specifically inhibits the activity (e.g., enzyme inhibitor, antibody). In certain embodiments, an activity or composition can be reduced or removed by subjecting a host cell or organism to a selective environment and screening for cells or organisms that have a reduced level or removal of the activity or composition.

In some embodiments, an untranslated ribonucleic acid or a cDNA can be used to reduce the expression of a particular activity or enzyme. For example, a host cell or organism can be engineered by genetic modification to express a nucleic acid reagent that reduces the expression of an activity by producing an RNA molecule that is partially or substantially homologous to a nucleic acid sequence of interest which encodes the activity of interest. The RNA molecule can bind to the nucleic acid sequence of interest and inhibit the nucleic acid sequence from performing its natural function, in certain embodiments. In some embodiments, the RNA may alter the nucleic acid sequence of interest which encodes the activity of interest in a manner that the nucleic acid sequence of interest is no longer capable of performing its natural function (e.g., the action of a ribozyme for example).

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1524 of 3077

In some embodiments, an activity and/or composition may be reduced in, or removed from, a host cell or organism by increasing or adding a separate activity or composition in the host cell or organism. For example, an activity and/or composition that inhibits a targeted activity or composition in a host cell or organism can be increased or added thereby reducing or eliminating the targeted activity or composition by adding or increasing an inhibiting activity or composition. Methods of increasing or adding an activity or composition in a cell or organism are described herein.

Increasing an amount of a composition and/or activity in a cell

An altered activity and/or composition in an engineered cell or organism is sometimes an added composition or activity that is not detectable in a host cell or organism. An altered activity or composition can also be an increased or elevated activity or amount of a composition in an engineered cell or organism. An increased or elevated activity or composition generally is an activity or an amount of the composition that is greater than the activity or composition amount detectable in a host cell or organism. However, an increased or elevated activity or amount of a composition in an engineered cell or organism can also be a detectable activity or detectable composition that is not detectable in a host cell or organism. An activity or amount of a composition can be increased to any suitable level for example, for production of a target molecule product (e.g., an organic acid), including but not limited to less than 2-fold (e.g., about 10% increase to about 99% increase; about 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% increase), 2-fold, 3-fold, 4-fold, 5-fold, 6-fold, 7-fold, 8-fold, 9-fold, of 10-fold increase, or greater than about 10-fold increase.

In some embodiments, an activity and/or composition may be added to or increased in a host cell or organism by increasing the number of copies of a polynucleotide that encodes a polypeptide composition or polypeptide having the activity. In some embodiments, the activity and/or amount of a native or endogenous polypeptide can be increased by introducing heterologous nucleic acid into a host cell or organism that includes copies of a polynucleotide that encodes the polypeptide, for example, introducing 1 to about 100 additional heterologous copies of the polynucleotide (e.g., introducing 1 or more, 2 or more, 3 or more, 4 or more, 5 or more, 6 or more, 7 or more, 8 or more, 9 or more, 10 or more, 11 or more, 12 or more, 13 or more, 14 or more, 15 or more, 16 or more, 17 or more, 18 or more, 19 or more, 20 or more, 22 or more, 24 or more, 25 or more, 26 or more, 28 or more, 30 or more additional copies of the

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1525 of 3077

polynucleotide). In certain embodiments, an activity and/or composition can be added or increased by inserting into a host cell or organism a polynucleotide that encodes a heterologous polypeptide from a different species having the added activity or composition, or encodes a heterologous polypeptide that is a modified version of an endogenous polypeptide. In such embodiments, 1 to about 100 copies of the polynucleotide can be introduced (e.g., introducing 1 or more, 2 or more, 3 or more, 4 or more, 5 or more, 6 or more, 7 or more, 8 or more, 9 or more, 10 or more, 11 or more, 12 or more, 13 or more, 14 or more, 15 or more, 16 or more, 17 or more, 18 or more, 19 or more, 20 or more, 22 or more, 24 or more, 25 or more, 26 or more, 28 or more, 30 or more copies). A heterologous polypeptide that is a "modified endogenous polypeptide" often has an activity different than an activity of a native polypeptide counterpart (e.g., different catalytic activity and/or different substrate specificity), and often is active (e.g., an activity (e.g., substrate turnover) is detectable). A heterologous polypeptide that is a "modified endogenous polypeptide" also often includes or lacks a cell location-targeting amino acid sequence that a native polypeptide counterpart has or doesn't have (e.g., in order to modify the cellular location of the expressed polypeptide). In certain embodiments, an activity or composition can be added or increased by inserting into a host cell or organism a heterologous polynucleotide that is (i) operably linked to another polynucleotide that encodes a polypeptide composition or a polypeptide having the added activity, and (ii) up regulates production of the polypeptide. Thus, a composition or an activity can be added or increased by inserting or modifying a regulatory polynucleotide operably linked to another polynucleotide that encodes a composition polypeptide or polypeptide having the targeted activity. In certain embodiments, an activity or composition can be added or increased by subjecting a host cell or organism to a selective environment and screening for cells or organisms that have a detectable level of the activity or composition. Examples of a selective environment include, without limitation, a medium containing a substrate that a host cell or organism can process and a medium lacking a substrate that a host cell or organism can process.

In some embodiments, an activity and/or composition may be added to or increased in a host cell or organism by decreasing or removing a separate activity or composition in a host cell or organism. For example, an activity and/or composition in a host cell or organism that inhibits a desired target activity or composition can be decreased or removed thereby reducing or eliminating the inhibition of the desired activity or composition and adding or increasing the desired activity. Methods of decreasing or removing an activity or composition in a cell or organism are described herein.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1526 of 3077

Nucleic acid manipulation

In certain embodiments, nucleotide sequences sometimes are added to, modified or removed from one or more of the nucleic acid reagent elements, such as the promoter, 5' UTR, target sequence, or 3' UTR elements, to enhance, potentially enhance, reduce, or potentially reduce transcription and/or translation before or after such elements are incorporated in a nucleic acid reagent. In some embodiments, one or more of the following sequences may be modified or removed if they are present in a 5' UTR: a sequence that forms a stable secondary structure (e.g., quadruplex structure or stem loop stem structure (e.g., EMBL sequences X12949, AF274954, AF139980, AF152961, S95936, U194144, AF116649 or substantially identical sequences that form such stem loop stem structures)); a translation initiation codon upstream of the nucleotide sequence start codon; a stop codon upstream of the nucleotide sequence translation initiation codon; an ORF upstream of the nucleotide sequence translation initiation codon; an iron responsive element (IRE) or like sequence; and a 5' terminal oligopyrimidine tract (TOP, e.g., consisting of 5-15 pyrimidines adjacent to the cap). Computer-assisted software programs are available for nucleic acid sequence evaluation to optimize untranslated region sequences (see, e.g., World Wide Web Uniform Resource Locator (URL) bioinformatics.ua.pt/software/mrna-optimiser/). A translational enhancer sequence and/or an internal ribosome entry site (IRES) sometimes is inserted into a 5'UTR (e.g., EMBL nucleotide sequences J04513, X87949, M95825, M12783, AF025841, AF013263, AF006822, M17169, M13440, M22427, D14838 and M17446 and substantially identical nucleotide sequences).

An AU-rich element (ARE, e.g., AUUUA repeats) and/or splicing junction that follows a non-sense codon sometimes is removed from or modified in a 3' UTR. A polyadenosine tail sometimes is inserted into a 3' UTR if none is present, sometimes is removed if it is present, and adenosine moieties sometimes are added to or removed from a polyadenosine tail present in a 3' UTR. Thus, some embodiments are directed to a process that includes: determining whether any nucleotide sequences that increase, potentially increase, reduce or potentially reduce translation efficiency are present in the elements, and adding, removing or modifying one or more of such sequences if they are identified. Certain embodiments are directed to a process that includes: determining whether any nucleotide sequences that increase or potentially increase translation efficiency are not present in the elements, and incorporating such sequences into the nucleic acid reagent.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1527 of 3077

In some embodiments, an activity and/or composition can be altered by modifying the nucleotide sequence of an ORF. An ORF sometimes is mutated or modified (for example, by point mutation, deletion mutation, insertion mutation, PCR based mutagenesis and the like) to alter, enhance or increase, reduce, substantially reduce or eliminate the activity of the encoded protein or peptide. The protein or peptide encoded by a modified ORF sometimes is produced in a lower amount or may not be produced at detectable levels, and in other embodiments, the product or protein encoded by the modified ORF is produced at a higher level (e.g., codons sometimes are modified so they are compatible with tRNA's preferentially used in the host or engineered cell or organism). To determine the relative activity, the activity from the product of the mutated ORF (or cell containing it) can be compared to the activity of the product or protein encoded by the unmodified ORF (or cell containing it).

In some embodiments, an ORF nucleotide sequence sometimes is mutated or modified to alter the triplet nucleotide sequences used to encode amino acids (e.g., amino acid codon triplets, for example). Modification of the nucleotide sequence of an ORF to alter codon triplets sometimes is used to change the codon found in the original sequence to better match the preferred codon usage of the organism in which the ORF or nucleic acid reagent will be expressed. The codon usage, and therefore the codon triplets encoded by a nucleic acid sequence, in bacteria may be different from the preferred codon usage in eukaryotes, like yeast or plants for example. Preferred codon usage also may be different between bacterial species. In certain embodiments, an ORF nucleotide sequence sometimes is modified to eliminate codon pairs and/or eliminate mRNA secondary structures that can cause pauses during translation of the mRNA encoded by the ORF nucleotide sequence. Translational pausing sometimes occurs when nucleic acid secondary structures exist in an mRNA, and sometimes occurs due to the presence of codon pairs that slow the rate of translation by causing ribosomes to pause. In some embodiments, the use of lower abundance codon triplets can reduce translational pausing due to a decrease in the pause time needed to load a charged tRNA into the ribosome translation machinery. Therefore, to increase transcriptional and translational efficiency in bacteria (e.g., where transcription and translation are concurrent, for example) or to increase translational efficiency in eukaryotes (e.g., where transcription and translation are functionally separated), the nucleotide sequence of a nucleotide sequence of interest can be altered to better suit the transcription and/or translational machinery of the host and/or genetically modified cell or organism. In certain embodiments, slowing the rate of translation by the use of

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1528 of 3077

lower abundance codons, which slow or pause the ribosome, can lead to higher yields of the desired product due to an increase in correctly folded proteins and a reduction in the formation of inclusion bodies.

Codons can be altered and optimized according to the preferred usage by a given organism by determining the codon distribution of the nucleotide sequence donor organism and comparing the distribution of codons to the distribution of codons in the recipient or host organism. Techniques described herein (e.g., site directed mutagenesis and the like) can then be used to alter the codons accordingly. Comparisons of codon usage can be done by hand, or using nucleic acid analysis software commercially available to the artisan (see, e.g., World Wide Web Uniform Resource Locator (URL) kazusa.or.jp/codon/, World Wide Web Uniform Resource Locator (URL) jcat.de, World Wide Web Uniform Resource Locator (URL) idtdna.com/CodonOpt).

Modification of the nucleotide sequence of an ORF also can be used to correct codon triplet sequences that have diverged in different organisms. For example, certain yeast (e.g., *Candida tropicalis*, *Candida viswanathii* and *Candida maltosa*) use the amino acid triplet CUG (e.g., CTG in the DNA sequence) to encode serine. CUG typically encodes leucine in most organisms. In order to maintain the correct amino acid in the resultant polypeptide or protein, the CUG codon must be altered to reflect the organism in which the nucleic acid reagent will be expressed. Thus, if an ORF from a bacterial donor is expressed in such a *Candida* yeast strain mentioned above, the heterologous nucleotide sequence must first be altered or modified to the appropriate leucine codon. Therefore, in some embodiments, the nucleotide sequence of an ORF sometimes is altered or modified to correct for differences that have occurred in the evolution of the amino acid codon triplets between different organisms. In some embodiments, the nucleotide sequence can be left unchanged at a particular amino acid codon, if the amino acid encoded is a conservative or neutral change in amino acid when compared to the originally encoded amino acid.

In some embodiments, an activity can be altered by modifying translational regulation signals, like a stop codon for example. A stop codon at the end of an ORF sometimes is modified to another stop codon, such as an amber stop codon. In some embodiments, a stop codon is introduced within an ORF, sometimes by insertion or mutation of an existing codon. An ORF comprising a modified terminal stop codon and/or internal stop codon often is translated in a

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1529 of 3077

system comprising a suppressor tRNA that recognizes the stop codon. An ORF comprising a stop codon sometimes is translated in a system comprising a suppressor tRNA that incorporates an unnatural amino acid during translation of the target protein or target peptide. Methods for incorporating unnatural amino acids into a target protein or peptide are known, which include, for example, processes utilizing a heterologous tRNA/synthetase pair, where the tRNA recognizes an amber stop codon and is loaded with an unnatural amino acid (e.g., World Wide Web Uniform Resource Locator (URL) iupac.org/news/prize/2003/wang.pdf).

Depending on the portion of a nucleic acid reagent (e.g., promoter, 5' or 3' UTR, ORI, ORF, and the like) chosen for alteration (e.g., by mutagenesis, introduction or deletion, for example), the modifications described above can alter a given activity by (i) increasing or decreasing feedback inhibition mechanisms, (ii) increasing or decreasing promoter initiation, (iii) increasing or decreasing translation initiation, (iv) increasing or decreasing translational efficiency, (v) modifying localization of peptides or products expressed from nucleic acid reagents described herein, (vi) increasing or decreasing the copy number of a nucleotide sequence of interest, or (vii) expression of an anti-sense RNA, RNAi, siRNA, ribozyme and the like. In some embodiments, alteration of a nucleic acid reagent or nucleotide sequence can alter a region involved in feedback inhibition (e.g., 5' UTR, promoter and the like). A modification sometimes is made that can add or enhance binding of a feedback regulator and sometimes a modification is made that can reduce, inhibit or eliminate binding of a feedback regulator.

In certain embodiments, alteration of a nucleic acid reagent or nucleotide sequence can alter sequences involved in transcription initiation (e.g., promoters, 5' UTR, and the like). A modification sometimes can be made that can enhance or increase initiation from an endogenous or heterologous promoter element. A modification sometimes can be made that removes or disrupts sequences that increase or enhance transcription initiation, resulting in a decrease or elimination of transcription from an endogenous or heterologous promoter element.

In some embodiments, alteration of a nucleic acid reagent or nucleotide sequence can alter sequences involved in translational initiation or translational efficiency (e.g., 5' UTR, 3' UTR, codon triplets of higher or lower abundance, translational terminator sequences and the like, for example). A modification sometimes can be made that can increase or decrease translational initiation, modifying a ribosome binding site for example. A modification sometimes can be made that can increase or decrease translational efficiency. Removing or adding sequences

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1530 of 3077

that form hairpins and changing codon triplets to a more or less preferred codon are non-limiting examples of genetic modifications that can be made to alter translation initiation and translation efficiency.

In certain embodiments, alteration of a nucleic acid reagent or nucleotide sequence can alter sequences involved in localization of peptides, proteins or other desired products (e.g., an organic acid, for example). A modification sometimes can be made that can alter, add or remove sequences responsible for targeting a polypeptide, protein or product to an intracellular organelle, the periplasm, cellular membranes, or extracellularly. Transport of a heterologous product to a different intracellular space or extracellularly sometimes can reduce or eliminate the formation of inclusion bodies (e.g., insoluble aggregates of the desired product).

In some embodiments, alteration of a nucleic acid reagent or nucleotide sequence can alter sequences involved in increasing or decreasing the copy number of a nucleotide sequence of interest. A modification sometimes can be made that increases or decreases the number of copies of an ORF stably integrated into the genome of an organism or on an epigenetic nucleic acid reagent. Non-limiting examples of alterations that can increase the number of copies of a sequence of interest include, adding copies of the sequence of interest by duplication of regions in the genome (e.g., adding additional copies by recombination or by causing gene amplification of the host genome, for example), cloning additional copies of a sequence onto a nucleic acid reagent, or altering an ORI to increase the number of copies of an epigenetic nucleic acid reagent. Non-limiting examples of alterations that can decrease the number of copies of a sequence of interest include, removing copies of the sequence of interest by deletion or disruption of regions in the genome, removing additional copies of the sequence from epigenetic nucleic acid reagents, or altering an ORI to decrease the number of copies of an epigenetic nucleic acid reagent.

In certain embodiments, increasing or decreasing the expression of a nucleotide sequence of interest can also be accomplished by altering, adding or removing sequences involved in the expression of an anti-sense RNA, RNAi, siRNA, ribozyme and the like. The methods described herein can be used to modify expression of anti-sense RNA, RNAi, siRNA, ribozyme and the like.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1531 of 3077

Nucleic acid sequences of interest can be genetically modified using methods known in the art. Mutagenesis techniques are particularly useful for small scale (e.g., 1, 2, 5, 10 or more nucleotides) or large scale (e.g., 50, 100, 150, 200, 500, or more nucleotides) genetic modification. Mutagenesis allows the artisan to alter the genetic information of a cell or organism in a stable manner, either naturally (e.g., isolation using selection and screening) or experimentally by the use of chemicals, radiation or inaccurate DNA replication (e.g., PCR mutagenesis). In some embodiments, genetic modification can be performed by whole scale synthetic synthesis of nucleic acids, using a native nucleotide sequence as the reference sequence, and modifying nucleotides that can result in the desired alteration of activity. Mutagenesis methods sometimes are specific or targeted to specific regions or nucleotides (e.g., site-directed mutagenesis, PCR-based site-directed mutagenesis, and in vitro mutagenesis techniques such as transplacement and in vivo oligonucleotide site-directed mutagenesis, for example). Mutagenesis methods sometimes are non-specific or random with respect to the placement of genetic modifications (e.g., chemical mutagenesis, insertion element (e.g., insertion or transposon elements) and inaccurate PCR based methods, for example).

Site directed mutagenesis is a procedure in which a specific nucleotide or specific nucleotides in a nucleic acid molecule are mutated or altered. Site directed mutagenesis typically is performed using a nucleic acid sequence of interest cloned into a circular plasmid vector. Site-directed mutagenesis requires that the wild type sequence be known and used a platform for the genetic alteration. Site-directed mutagenesis sometimes is referred to as oligonucleotide-directed mutagenesis because the technique can be performed using oligonucleotides which have the desired genetic modification incorporated into the complement of a nucleotide sequence of interest. The wild type sequence and the altered nucleotide are allowed to hybridize and the hybridized nucleic acids are extended and replicated using a DNA polymerase. The double stranded nucleic acids are introduced into a host (e.g., *E. coli*, for example) and further rounds of replication are carried out in vivo. The transformed cells carrying the mutated nucleic acid sequence are then selected and/or screened for those cells carrying the correctly mutagenized sequence. Cassette mutagenesis and PCR-based site-directed mutagenesis are further modifications of the site-directed mutagenesis technique. Site-directed mutagenesis can also be performed in vivo (e.g., transplacement “pop-in pop-out”, in vivo site-directed mutagenesis with synthetic oligonucleotides and the like, for example).

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1532 of 3077

PCR-based mutagenesis can be performed using PCR with oligonucleotide primers that contain the desired mutation or mutations. The technique functions in a manner similar to standard site-directed mutagenesis, with the exception that a thermocycler and PCR conditions are used to replace replication and selection of the clones in a microorganism host. As PCR-based mutagenesis also uses a circular plasmid vector, the amplified fragment (e.g., linear nucleic acid molecule) containing the incorporated genetic modifications can be separated from the plasmid containing the template sequence after a sufficient number of rounds of thermocycler amplification, using standard electrophoretic procedures. A modification of this method uses linear amplification methods and a pair of mutagenic primers that amplify the entire plasmid. The procedure can take advantage of the *E. coli* Dam methylase system which causes DNA replicated in vivo to be sensitive to the restriction endonucleases DpnI. PCR synthesized DNA is not methylated and is therefore resistant to DpnI. This approach allows digestion of the template plasmid, leaving the genetically modified, PCR synthesized plasmids for isolating and transforming into a host bacteria for DNA repair and replication, thereby facilitating subsequent cloning and identification steps. A certain amount of randomness can be added to PCR-based sited directed mutagenesis by using partially degenerate primers.

DNA shuffling is a method which uses DNA fragments from members of a mutant library and reshuffles the fragments randomly to generate new mutant sequence combinations. The fragments are typically generated using DNaseI, followed by random annealing and re-joining using self priming PCR. The DNA overhanging ends, from annealing of random fragments, provide “primer” sequences for the PCR process. Shuffling can be applied to libraries generated by any of the above mutagenesis methods.

Error prone PCR and its derivative rolling circle error prone PCR uses increased magnesium and manganese concentrations in conjunction with limiting amounts of one or two nucleotides to reduce the fidelity of the Taq polymerase. The error rate can be as high as 2% under appropriate conditions, when the resultant mutant sequence is compared to the wild type starting sequence. After amplification, the library of mutant coding sequences must be cloned into a suitable plasmid. Although point mutations are the most common types of mutation in error prone PCR, deletions and frameshift mutations are also possible. There are a number of commercial error-prone PCR kits available, including those from Agilent and Takara Bio, U.S.A. (e.g., World Wide Web Uniform Resource Locator (URL) agilent.com and World Wide Web Uniform Resource Locator (URL) clontech.com, respectively, for example). Rolling circle error-

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1533 of 3077

prone PCR is a variant of error-prone PCR in which wild-type sequence is first cloned into a plasmid, then the whole plasmid is amplified under error-prone conditions.

In contrast to site-directed or specific mutagenesis, random mutagenesis does not require any sequence information and can be accomplished by a number of widely different methods. Random mutagenesis often is used to generate mutant libraries that can be used to screen for the desired genotype or phenotype. Non-limiting examples of random mutagenesis include; chemical mutagenesis, UV-induced mutagenesis, insertion element or transposon-mediated mutagenesis, DNA shuffling, error-prone PCR mutagenesis, and the like.

Chemical mutagenesis often involves chemicals like ethyl methanesulfonate (EMS), nitrous acid, mitomycin C, N-methyl-N-nitrosourea (MNU), diepoxybutane (DEB), 1, 2, 7, 8-diepoxyoctane (DEO), methyl methane sulfonate (MMS), N-methyl- N'-nitro-N-nitrosoguanidine (MNNG), 4-nitroquinoline 1-oxide (4-NQO), 2-methyloxy-6-chloro-9(3-[ethyl-2-chloroethyl]-aminopropylamino)-acridinedihydrochloride (ICR-170), 2-amino purine (2AP), and hydroxylamine (HA), provided herein as non-limiting examples. These chemicals can cause base-pair substitutions, frameshift mutations, deletions, transversion mutations, transition mutations, incorrect replication, and the like. In some embodiments, the mutagenesis can be carried out in vivo. Sometimes the mutagenic process involves the use of the host organism's DNA replication and repair mechanisms to incorporate and replicate the mutagenized base or bases.

Another type of chemical mutagenesis involves the use of base-analogs. The use of base-analogs causes incorrect base pairing which in the following round of replication is corrected to a mismatched nucleotide when compared to the starting sequence. Base analog mutagenesis introduces a small amount of non-randomness to random mutagenesis, because specific base analogs can be chosen which can be incorporated at certain nucleotides in the starting sequence. Correction of the mispairing typically yields a known substitution. For example, bromo-deoxyuridine (BrdU) can be incorporated into DNA and replaces T in the sequence. The host DNA repair and replication machinery can sometime correct the defect, but sometimes will mispair the BrdU with a G. The next round of replication then causes a G-C transversion from the original A-T in the native sequence.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1534 of 3077

Ultra violet (UV) induced mutagenesis is caused by the formation of thymidine dimers when UV light irradiates chemical bonds between two adjacent thymine residues. Excision repair mechanism of the host organism correct the lesion in the DNA, but occasionally the lesion is incorrectly repaired typically resulting in a C to T transition.

In some embodiments, an altered activity can be found by screening cells or an organism under conditions that select for the desired change in activity. For example, certain microorganisms can be adapted to increase or decrease an activity by selecting or screening the organism in question on a media containing substances that are poorly metabolized or even toxic. An increase in the ability of an organism to grow a substance that is normally poorly metabolized may result in an increase in the growth rate on that substance, for example. A decrease in the sensitivity to a toxic substance might be manifested by growth on higher concentrations of the toxic substance, for example. Modifications obtained in this manner are not limited to alterations in promoter sequences. That is, screening microorganisms by selective pressure, as described above, can yield genetic alterations that can occur in non-promoter sequences, and sometimes also can occur in sequences that are not in the nucleotide sequence of interest, but in a related nucleotide sequences (e.g., a gene involved in a different step of the same pathway, a transport gene, and the like). Such mutants sometimes can be found by isolating variants from unique environments.

Cells or organisms with altered activities can also be isolated using genetic selection and screening of cells or organisms challenged on selective media or by identifying naturally occurring variants from unique environments. For example, 2-deoxy-D-glucose is a toxic glucose analog. Growth of yeast on this substance yields mutants that are glucose-deregulated. A number of mutants have been isolated using 2-deoxy-D-glucose including transport mutants, and mutants that ferment glucose and galactose simultaneously instead of glucose first then galactose when glucose is depleted. Similar techniques have been used to isolate mutant microorganisms that can metabolize plastics (e.g., from landfills), petrochemicals (e.g., from oil spills), and the like, either in a laboratory setting or from unique environments.

Similar methods can be used to isolate cells or organisms having existing mutations in a desired activity when the activity exists at a relatively low or nearly undetectable level in the cell or organism of choice, in some embodiments. The method generally consists of growing the cell or organism to a specific density in liquid culture, concentrating the cells, and plating the cells on

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1535 of 3077

various concentrations of the substance to which an increase in metabolic activity is desired. The cells are incubated at a moderate growth temperature, for 5 to 10 days. To enhance the selection process, the plates can be stored for another 5 to 10 days at a low temperature. The low temperature sometimes can allow strains that have gained or increased an activity to continue growing while other strains are inhibited for growth at the low temperature. Following the initial selection and secondary growth at low temperature, the plates can be replica plated on higher or lower concentrations of the selection substance to further select for the desired activity.

Insertion element or transposon-mediated mutagenesis makes use of naturally occurring or modified naturally occurring mobile genetic elements. Transposons often encode accessory activities in addition to the activities necessary for transposition (e.g., movement using a transposase activity, for example). In many examples, transposon accessory activities are antibiotic resistance markers (e.g., Tn903 kan^r). Insertion elements typically only encode the activities necessary for movement of the nucleic acid sequence. Insertion element and transposon mediated mutagenesis often can occur randomly, however specific target sequences are known for some transposons. Mobile genetic elements like IS elements or Transposons (Tn) often have inverted repeats, direct repeats or both inverted and direct repeats flanking the region coding for the transposition genes. Recombination events catalyzed by the transposase cause the element to remove itself from the genome and move to a new location, leaving behind a portion of an inverted or direct repeat. Classic examples of transposons are the “mobile genetic elements” discovered in maize. Transposon mutagenesis kits are commercially available which are designed to leave behind a 5 codon insert (e.g., Mutation Generation System kit, ThermoFisher Scientific, World Wide Web Uniform Resource Locator (URL) thermofisher.com/, for example). This allows the artisan to identify the insertion site, without fully disrupting the function of most genes.

Introduction of nucleic acids into cells

Engineered cells and organisms can be prepared by altering, introducing and/or removing nucleotide sequences in the host genome or in stably maintained epigenetic nucleic acid reagents, as described herein. The nucleic acid reagents used to alter, introduce or remove nucleotide sequences in the host genome or epigenetic nucleic acids can be prepared using the methods described herein and/or available to the artisan.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1536 of 3077

Nucleic acid sequences having a desired activity can be isolated from cells of a suitable organism using lysis and nucleic acid purification procedures described in a known reference manual (e.g., Maniatis, T., E. F. Fritsch and J. Sambrook (1982) *Molecular Cloning: a Laboratory Manual*; Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.) or using commercially available cell lysis and DNA purification reagents and kits. In some embodiments, nucleic acids used to engineer cells or microorganisms can be provided for conducting methods described herein after processing of the organism containing the nucleic acid. For example, the nucleic acid of interest may be extracted, isolated, purified or amplified from a sample (e.g., from a cell(s) or organism of interest or culture containing a plurality of cells or organisms of interest, like yeast or bacteria for example). The term “isolated” as used herein refers to nucleic acid removed from its original environment (e.g., the natural environment if it is naturally occurring, or a host cell if expressed exogenously), and thus is altered "by the hand of man" from its original environment. An isolated nucleic acid generally is provided with fewer non-nucleic acid components (e.g., protein, lipid) than the amount of components present in a source sample. A composition containing isolated sample nucleic acid can be substantially isolated (e.g., about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or greater than 99% free of non-nucleic acid components). The term “purified” as used herein refers to sample nucleic acid provided that contains fewer nucleic acid species than in the sample source from which the sample nucleic acid is derived. A composition containing sample nucleic acid may be substantially purified (e.g., about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or greater than 99% free of other nucleic acid species). The term “amplified” as used herein refers to subjecting nucleic acid of a cell, organism or sample to a process that linearly or exponentially generates amplicon nucleic acids having the same or substantially the same nucleotide sequence as the nucleotide sequence of the nucleic acid in the sample, or portion thereof. As noted herein, the nucleic acids used to prepare nucleic acid reagents as described herein can be subjected to fragmentation or cleavage.

Amplification of nucleic acids is sometimes necessary when dealing with cells or organisms that are difficult to culture. Where amplification may be desired, any suitable amplification technique can be utilized. Non-limiting examples of methods for amplification of polynucleotides include, polymerase chain reaction (PCR); ligation amplification (or ligase chain reaction (LCR)); amplification methods based on the use of Q-beta replicase or template-dependent polymerase (see US Patent Publication Number US20050287592); helicase-dependent isothermal

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1537 of 3077

amplification (Vincent et al., "Helicase-dependent isothermal DNA amplification". EMBO reports 5 (8): 795–800 (2004)); strand displacement amplification (SDA); thermophilic SDA nucleic acid sequence based amplification (3SR or NASBA) and transcription-associated amplification (TAA). Non-limiting examples of PCR amplification methods include standard PCR, AFLP-PCR, Allele-specific PCR, Alu-PCR, Asymmetric PCR, Colony PCR, Hot start PCR, Inverse PCR (IPCR), In situ PCR (ISH), Intersequence-specific PCR (ISSR-PCR), Long PCR, Multiplex PCR, Nested PCR, Quantitative PCR, Reverse Transcriptase PCR (RT-PCR), Real Time PCR, Single cell PCR, Solid phase PCR, combinations thereof, and the like. Reagents and hardware for conducting PCR are commercially available.

Protocols for conducting the various type of PCR listed above are readily available to the artisan. PCR conditions can be dependent upon primer sequences, target abundance, and the desired amount of amplification, and therefore, one of skill in the art may choose from a number of PCR protocols available (see, e.g., U.S. Pat. Nos. 4,683,195 and 4,683,202; and PCR Protocols: A Guide to Methods and Applications, Innis et al., eds, 1990). PCR often is carried out as an automated process with a thermostable enzyme. In this process, the temperature of the reaction mixture is cycled through a denaturing region, a primer-annealing region, and an extension reaction region automatically. Machines specifically adapted for this purpose are commercially available. A non-limiting example of a PCR protocol that may be suitable for embodiments described herein is, treating the sample at 95ºC for 5 minutes; repeating forty-five cycles of 95ºC for 1 minute, 59ºC for 1 minute, 10 seconds, and 72ºC for 1 minute 30 seconds; and then treating the sample at 72ºC for 5 minutes. Multiple cycles frequently are performed using a commercially available thermal cycler. Suitable isothermal amplification processes known and selected by the person of ordinary skill in the art also may be applied, in certain embodiments. In some embodiments, nucleic acids encoding polypeptides with a desired activity can be isolated by amplifying the desired sequence from a cell or organism having the desired activity using oligonucleotides or primers designed based on sequences described herein.

Synthetic nucleic acids, e.g., codon-optimized sequences, can be generated using a variety of methods. For example, whole-scale synthetic chemistry can be used to generate an entire sequence. Other methods include use of chemically-generated oligonucleotides in amplification methods, e.g., recursive PCR, that build an entire nucleotide sequence (see, e.g., Prodromou and Pearl (1992) *Protein Engineering* 5(8):827-829; Yehezkel et al. (2013) *Gene Synthesis:*

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1538 of 3077

Methods and Protocols in Methods in Molecular Biology 852:35-47, Jean Piccoud (ed.) Springer Science and Business Media LLC).

Amplified, isolated and/or purified nucleic acids can be cloned into the recombinant DNA vectors described herein or into suitable commercially available recombinant DNA vectors. Cloning of nucleic acid sequences of interest into recombinant DNA vectors can facilitate further manipulations of the nucleic acids for preparation of nucleic acid reagents, (e.g., alteration of nucleotide sequences by mutagenesis, homologous recombination, amplification and the like, for example). Standard cloning procedures (e.g., enzymatic digestion, ligation, and the like) are known (e.g., described in Maniatis, T., E. F. Fritsch and J. Sambrook (1982) *Molecular Cloning: a Laboratory Manual*; Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.).

In some embodiments, nucleic acid sequences prepared by isolation or amplification can be used, without any further modification, to add an activity to a cell or microorganism and thereby create a genetically modified or engineered cell or microorganism. In certain embodiments, nucleic acid sequences prepared by isolation or amplification can be genetically modified to alter (e.g., increase or decrease, for example) a desired activity. In some embodiments, nucleic acids, used to add an activity or composition to a cell or organism, sometimes are genetically modified to optimize the heterologous polynucleotide sequence encoding the desired activity (e.g., polypeptide or protein, for example). The term “optimize” as used herein can refer to alteration to increase or enhance expression by preferred codon usage. The term optimize can also refer to modifications to the amino acid sequence to increase the activity of a polypeptide or protein, such that the activity exhibits a higher catalytic activity as compared to the “natural” version of the polypeptide or protein.

A heterologous, recombinant, or mutagenized polynucleotide can be introduced into a nucleic acid reagent for introduction into a host cell or organism, thereby generating an engineered cell or microorganism. Standard recombinant DNA techniques (restriction enzyme digests, ligation, and the like) can be used by the artisan to combine a nucleic acid of interest into a suitable nucleic acid reagent capable of (i) being stably maintained by selection in the host cell or organism, or (ii) being integrated into the genome of the host cell or organism. Sometimes nucleic acid reagents include two replication origins to allow manipulation of the same nucleic acid reagent in bacteria before final introduction of the final product into the host cell or organism (e.g., yeast or fungus for example). Standard molecular biology and recombinant

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1539 of 3077

DNA methods are known (e.g., described in Maniatis, T., E. F. Fritsch and J. Sambrook (1982) *Molecular Cloning: a Laboratory Manual*; Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.).

Nucleic acid reagents can be introduced into cells or microorganisms using various techniques. Non-limiting examples of methods used to introduce heterologous nucleic acids into various cells or organisms include; transformation, transfection, transduction, electroporation, ultrasound-mediated transformation, particle bombardment and the like. In some instances the addition of carrier molecules (e.g., bis-benzimdazolyl compounds, for example, see US Patent 5595899) can increase the uptake of DNA in cells that may be difficult to transform by conventional methods. Conventional methods of transformation are known (e.g., described in Maniatis, T., E. F. Fritsch and J. Sambrook (1982) *Molecular Cloning: a Laboratory Manual*; Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.).

Linear DNA transformed into a host cell can be integrated into the genome by homologous recombination. The localization of genomic integration is determined by the homologous sequence at the ends of the transformed linear DNA. Fig. 1 is a diagrammatic representation of a cassette for the addition of a gene of interest (GOI) into a host non-functional ura3 locus using the single crossover integration method. The core of the cassette contains the GOI gene with a promoter (P_{XXX}) and terminator (T_{XXX}) for controlling transcription of the GOI gene. These DNA cassettes are typically generated by overlap extension PCR assembly of the cassette elements or by PCR amplification from circular plasmids containing the entire cassette. Additionally, circular plasmids containing cassette elements may be cut within (e.g., the middle of) the *URA3* ORF to generate a linear DNA fragment used in transforming cells. A circular DNA vector containing the cassette core and an intact *URA3* gene can be linearized by endonuclease-mediated cutting the vector such that it splits the *URA3* selectable marker within (e.g., the middle of) the ORF. The resulting linear DNA contains the expression cassette of the gene of interest (GOI) positioned between the *URA3* promoter (P_{URA3}) and terminator (T_{URA3}). Parts A and B of Fig. 1 show the results of integration of one copy (A) and two copies (B) of the cassette into a Ura^- auxotrophic mutant strain. Integration of one cassette generates an added, functional GOI expression unit and may or may not provide for expression of a functional Ura3p, depending on the nature of ura3 locus and the location of the split in the *URA3* selectable marker. Integration of two or more copies of the cassette in tandem arrays generates a complete, functional *URA3* sequence by combining the 5' end of *URA3* from one copy of the

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1540 of 3077

cassette and the 3' end of *URA3* from the second copy of the cassette. Additional copies may also be integrated. Transformants can be selected by growth on uracil-free media. This integration method thus favors selection of transformants containing multiple copies of the GOI.

In some embodiments, other auxotrophic or dominant selection markers can be used in place of URA3 (e.g., an auxotrophic selectable marker), with the appropriate change in selection media and selection agents. Auxotrophic selectable markers are used in strains deficient for synthesis of a required biological molecule (e.g., amino acid or nucleoside, for example). Non-limiting examples of additional auxotrophic markers include; HIS3, TRP1, LEU2, LEU2-d, and LYS2. Certain auxotrophic markers (e.g., URA3 and LYS2) allow counter selection to select for the second recombination event that pops out all but one of the direct repeats of the recombination construct. HIS3 encodes an activity involved in histidine synthesis. TRP1 encodes an activity involved in tryptophan synthesis. LEU2 encodes an activity involved in leucine synthesis. LEU2-d is a low expression version of LEU2 that selects for increased copy number (e.g., gene or plasmid copy number, for example) to allow survival on minimal media without leucine. LYS2 encodes an activity involved in lysine synthesis, and allows counter selection for recombination out of the LYS2 gene using alpha-aminoadipate (α-aminoadipate).

Dominant selectable markers are useful because they also allow use of industrial and/or prototrophic strains for genetic manipulations. Additionally, dominant selectable markers provide the advantage that rich medium can be used for plating and culture growth, and thus growth rates are markedly increased. Non-limiting examples of dominant selectable markers include; Tn903 kan^r, Cm^r, Hyg^r, CUP1, and DHFR. Tn903 kan^r encodes an activity involved in kanamycin antibiotic resistance (e.g., typically neomycin phosphotransferase II or NPTII, for example). Cm^r encodes an activity involved in chloramphenicol antibiotic resistance (e.g., typically chloramphenicol acetyl transferase, for example). Hyg^r encodes an activity involved in hygromycin resistance by phosphorylation of hygromycin B (e.g., hygromycin phosphotransferase, or HPT). CUP1 encodes an activity involved in resistance to heavy metal (e.g., copper, for example) toxicity. DHFR encodes a dihydrofolate reductase activity which confers resistance to methotrexate and sulfanilamde compounds.

Homologous recombination can also be used as a tool for mutagenesis. Homologous recombination can be used to specifically target regions of known sequence for insertion of heterologous nucleotide sequences using the host cell's natural DNA replication and repair

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1541 of 3077

enzymes. Homologous recombination methods sometimes are referred to as mutagenesis, transplacement, knock-out mutagenesis or knock-in mutagenesis. Integration of a nucleic acid sequence into a host genome by a double crossover homologous recombination event inserts the entire nucleic acid reagent at the targeted location. A second homologous recombination event driven by direct repeat DNA sequences contained in the integrated nucleic acid cassette excises (e.g., "pop out" or "loop out") all but a portion of the nucleic acid reagent, leaving behind a heterologous sequence, often referred to as a "footprint" or "scar". Mutagenesis by insertion (e.g., knock in) or by leaving behind a disrupting heterologous nucleic acid (e.g., knock out) serves to disrupt or "knock out" the function of the gene or nucleic acid sequence in which insertion occurs. By combining selectable markers and/or auxotrophic markers with nucleic acid reagents designed to provide the appropriate nucleic acid target sequences, the artisan can target a selectable nucleic acid reagent to a specific genomic region, and then select for recombination events that "pop out" a portion of the inserted nucleic acid reagent.

Such methods take advantage of nucleic acid reagents that have been specifically designed with known target nucleic acid sequences at or near a nucleic acid or genomic region of interest. Popping out typically leaves a "foot print" of left over sequences that remain after the recombination event. The left over sequence can disrupt a gene and thereby reduce or eliminate expression of that gene. In some embodiments, the method can be used to insert sequences, upstream or downstream of genes that can result in an enhancement or reduction in expression of the gene. In certain embodiments, new genes can be introduced into the genome of a host cell or organism using similar homologous recombination methods. An example of a yeast recombination system using the *URA3* gene and 5-FOA is described herein.

One method for genetic modification is described by Alani et al. ("A method for gene disruption that allows repeated use of URA3 selection in the construction of multiply disrupted yeast strains", *Genetics* 116(4):541-545 August 1987). The original method uses a *URA3* gene cassette with 1000 base pairs (bp) of the same nucleotide sequence cloned in the same orientation on either side of the *URA3* cassette. Targeting sequences of about 50 bp are added to each side of the construct. The double-stranded targeting sequences are complementary to sequences in the genome of the host cell or organism. The targeting sequences allow site-specific recombination in a region of interest. A modification of the original technique replaces the two 1000 bp sequence direct repeats with two 200 bp direct repeats. The modified method also uses 50 bp targeting sequences. The modification reduces or eliminates recombination of

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1542 of 3077

a second knock out into the 1000 bp repeat left behind in a first mutagenesis, therefore allowing multiply knocked out yeast. Additionally, the 200 bp sequences used in the method are uniquely designed, self-assembling sequences that leave behind identifiable footprints. The technique used to design the sequences incorporate design features such as low identity to the yeast genome, and low identity to each other. Therefore, a library of the self-assembling sequences can be generated to allow multiple knockouts in the same organism, while reducing or eliminating the potential for integration into a previous knockout.

Fig. 2 is a diagrammatic illustration of an exemplary gene knock out cassette. In this example, two slightly different cassettes are shown for use in separately disrupting each of the two *FAT1* alleles in a diploid yeast such as *Candida viswanathii*. The two cassette-containing nucleic acid segments are referred to as "Deletion 1" and "Deletion 2," respectively, in the figure. Each cassette contains a *URA3* gene including a URA3 promoter (P_{URA3}) and terminator (T_{URA3}). The complete *URA3* expression cassette provides for expression of orotidine-5'-monophosphate (OMP) dicarboxylase in a Ura^- host cell into which the cassette has integrated, and yields a prototrophic transformant that can be selected for by growth in uracil-free media. Integration into the *FAT1* locus by a first crossover event is provided for by the presence of sequences located on either side of the cassette that are homologous to sequences in the target locus (e.g., *FAT1*). Additionally, each cassette contains a repeat of the P_{URA3} sequence immediately downstream of the terminator sequence. This repeat sequence can be used in a second recombination event that results in the looping out of the *URA3* gene sequence which is facilitated by growth of the prototrophic transformants in the presence of 5-FOA yielding a ura^- auxotroph. All or a portion of the P_{URA3} sequence repeat remains in the genome and disrupts the *FAT1* gene allele such that it no longer yields a functional gene product. The heterozygous transformant can then be transformed with the second cassette (e.g., Deletion 2) and undergo the same two crossover events to yield a homozygous Ura^- cell. The two separate deletion cassette-containing fragments differ in the sequences of the target gene that they contain on each side of the *URA3* cassette which results in integration into different positions in the target gene.

The URA3 cassette makes use of the toxicity of 5-FOA in yeast carrying a functional URA3 gene. Uracil synthesis-deficient host yeast are transformed with the modified URA3 cassette, using standard yeast transformation protocols, and the transformed cells are plated on minimal media minus uracil. In some embodiments, PCR can be used to verify correct insertion into the

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1543 of 3077

region of interest in the host genome, and in certain embodiments the PCR step can be omitted. Inclusion of the PCR step can reduce the number of transformants that are counter selected to "pop out" the URA3 cassette. The transformants (e.g., all or the ones determined as correct by PCR, for example) can then be counter-selected on media containing 5-FOA, which will select for recombination events looping out the URA3 cassette, thus rendering the yeast Ura^- again, and resistant to 5-FOA toxicity. Targeting sequences used to direct recombination events to specific regions are presented herein. A modification of the method described above can be used to integrate genes into the chromosome in which, following recombination, a functional gene is left in the chromosome next to a, e.g., 200-bp, footprint. Such methods provide for addition of a desired nucleic acid into the host genome in combination with disruption of an endogenous nucleic acid.

Fig. 3A, Fig. 3B, and Fig 3C show diagrammatic illustrations of a knock-in gene disruption method which disrupts one target gene ("GOI1") and also adds a desired gene of interest ("GOI2") at the disrupted locus. As shown in Fig. 3A, the basic URA3 disruption cassette can be the same as that described in Fig. 2; however, there is an additional expression cassette immediately downstream of the second P_{URA3} repeat sequence. This expression cassette contains the gene of interest, GOI2, for adding to the endogenous GOI1 locus and includes a promoter (P_{XXX}) and terminator (T_{XXX}) for controlling transcription of GOI2. Immediately upstream of the first P_{URA3} sequence is a sequence of nucleotides of the GOI1 gene, and immediately downstream of the terminator (T_{XXX}) for GOI2 is another sequence of the GOI1 gene. These sequences are for use in integration of the cassettes into the GOI1 locus. Fig. 3B shows the locus after the first crossover event. These transformants are selected for growth in uracil-free media. In order to remove the *URA3* gene and regenerate an auxotrophic cell that can be further modified using the *URA3* marker method, the transformants are grown in the presence of 5-FOA to facilitate the second crossover event. The result of that event is shown in Fig. 3C which depicts the P_{URA3} sequence that remains followed by a functional GOI2 cassette.

Protein engineering methods

As described herein, one method of altering carbon flux in cells and organisms is to modify one or more activities involved in carbon processing in cells. These activities can be modified by altering one or more elements directly and/or indirectly involved in the activities. Such elements include, but are not limited to, nucleic acids (e.g., transcription regulatory elements, addition

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1544 of 3077

and/or deletion of nucleic acids), peptides (e.g., signal peptides regulating protein localization in cells) and polypeptides (e.g., enzymes regulating reactions in metabolic pathways). Peptides and polypeptides can be modified in multiple ways, including, for example, alteration of the primary structure (i.e., amino acid sequence), secondary structure, post-translational chemical modification (e.g., phosphorylation, acylation, glycosylation) and processing (e.g., proteolytic cleavage). Many protein modifications can be achieved through alteration of the nucleic acid encoding the protein in a cell. Alteration of the nucleic acid coding sequence can result in alteration of the amino acid sequence which in turn can modify the intra- and inter-polypeptide interactions of the encoded protein. Such alterations can thus result in modification of the activity of the polypeptide and the activity of any metabolic processes in which it may participate.

In some embodiments of the cells, organisms, compositions and methods provided herein, a modified polypeptide can be expressed in a cell or organism by introducing a modification into nucleic acid encoding the polypeptide in the cell or organism. Modified polypeptides often have an activity different than the activity of an unmodified counterpart. A modified activity sometimes is a different transport activity, a different catalytic activity, a different substrate specificity, or a different catalytic activity and a different substrate specificity. A different activity sometimes is an activity that is higher (e.g., increased activity) or lower (e.g., decreased activity) than the activity of an unmodified counterpart polypeptide. In some embodiments, the catalytic activity of a modified polypeptide is higher or lower than the catalytic activity of the unmodified counterpart for a particular substrate. In certain embodiments, the specificity of a modified polypeptide for a particular substrate is higher or lower than the specificity of the unmodified counterpart for a particular substrate. A modified polypeptide often is active and an activity of a modified polypeptide often can be detected (e.g., substrate turnover can be detected). An activity for a particular polypeptide that is modified sometimes is referred to as a "target activity." As described herein, target activities include, but are not limited to, activities of ω-oxidation, β-oxidation, acetyl-CoA processing, carnitine/acetylcarnitine shuttle, membrane transport, fatty acid biosynthesis, acyl-CoA formation/degradation. Non-limiting examples of particular target activities include carnitine acetyltransferase, carnitine translocase, acetyl-CoA carboxylase, ATP citrate lyase, acetyl-CoA hydrolase, acetyl-CoA synthetase, thioesterase, acyl-CoA synthetase, monooxygenase, cytochrome P450 reductase, alcohol dehydrogenase, alcohol oxidase, aldehyde dehydrogenase, acyl-CoA oxidase, 3-ketoacyl-CoA thiolase, peroxisomal transporter, peroxisome biogenesis factor and multifunctional enzyme (e.g., enoyl-CoA hydratase and/or 3-

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1545 of 3077

hydroxyacyl-CoA dehydrogenase) activities. In some of the embodiments provided herein, these and other activities can be modified in a cell or organism.

One or more particular modifications can be selected to generate a modified polypeptide having a target activity. Modifications often are amino acid modifications (e.g., deletion, insertion of one or more amino acids). Amino acid modifications sometimes are amino acid substitutions. Amino acid substitutions sometimes are conservative, non-limiting examples of which include substitution of an amino acid containing an acidic moiety for another amino acid containing an acidic moiety (e.g., D, E), substitution of an amino acid containing a basic moiety for another amino acid containing a basic moiety (e.g., H, K, R), substitution of an amino acid containing an aliphatic chain moiety for another amino acid containing an aliphatic chain moiety (e.g., V, L, I, A), substitution of an amino acid containing a cyclic moiety for another amino acid containing a cyclic moiety (e.g., W, F, Y), and substitution of an amino acid containing a polar moiety for another amino acid containing a polar moiety (e.g., S, T). Amino acid substitutions sometimes are non-conservative, non-limiting examples of which include substitution of an amino acid containing an acidic moiety for an amino acid containing a basic moiety, substitution of an amino acid containing a basic moiety for an amino acid containing an acidic moiety, substitution of an amino acid containing relatively small moiety (e.g., G, A) for another amino acid containing a relatively large moiety (e.g., Y, W, F, I, L), and substitution of an amino acid containing a relatively large moiety for another amino acid containing an relatively small moiety.

Particular modifications can be selected using any suitable method known in the art. In certain embodiments, a reference structure is known for a related polypeptide with a known activity, and modifications to a target polypeptide can be guided by alignment of the target polypeptide structure to the reference structure. A reference structure sometimes is a primary structure (e.g., polynucleotide or polypeptide sequence) and the primary structure of a target can be aligned to the reference structure using an alignment method known in the art. Particular amino acids in the target that align with (e.g., are identical to or homologous to) or do not align with (e.g., are not identical to or not homologous to) particular amino acids in the reference can be selected for modification. Selections can be made by inspection of an alignment or by software known in the art that identifies, scores and/or ranks amino acids for modification based on an alignment. A reference structure sometimes is a secondary structure, tertiary structure or quaternary structure, each of which are three dimensional structures pertaining to a polypeptide. A primary structure of a target polypeptide can be modeled to a secondary, tertiary or

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1546 of 3077

quaternary reference structure using three-dimensional modeling software known in the art. A secondary, tertiary or quaternary structure of a target polypeptide can be compared to a secondary, tertiary or quaternary reference structure using three-dimensional comparative software known in the art. Particular structures (e.g., a particular individual amino acid; a particular group of contiguous or non-contiguous amino acids) in the target that align with or map to, or do not align with or map to, particular structures in the reference can be selected for modification. Also, particular structures in the target that are in proximity to a substrate or co-factor can be selected for modification. Selections can be made by inspection of an alignment or map or by software known in the art that identifies, scores and/or ranks amino acids and/or structures for modification based on an alignment and map. After particular amino acids and/or structures are selected for modification in a first polypeptide, amino acids and structures in a second polypeptide that align with the selected amino acids and structures in the first polypeptide may be identified.

For example, a structural model of a protein can be created based on the crystal structure of the protein using SWISS-MODEL, which has been described by Arnold *et al.* ((2006) *Bioinformatics* 22: 195-201), Guex *et al.* ((2009) *Electrophoresis* 30 Supplement 1: S162-S173) and Kiefer *et al*. ((2009) *Nucleic Acids Res.* 37 (Database issue): D387-D392). As described herein, the resulting structural model can be analyzed to identify sites in the protein that potentially participate in determining an activity of the protein. HotSpot Wizard is an example of a tool for identifying sites for engineering of substrate specificity and/or activity of enzymes using a combination of structural, functional and sequence analysis and has been described by Pavelka *et al*. ((2009) *Nucleic Acids Res*. 37 (Web Server issue): W376-W383) (see also: HotSpot Wizard 1.7 World Wide Web Uniform Resource Locator (URL) loschmidt.chemi.muni.cz/hotspotwizard/index.jsp). Identification of such sites facilitates a determination of possible amino acids to target for mutagenesis in modifying the activity of a protein (e.g., enzyme). Part of the HotSpot Wizard analysis is the identification of homologs by a BLAST search (see, e.g., Johnson *et al*. (2008) *Nucleic Acids Res*. 36 (Web Server issue): W5-W9) and their alignment using MUSCLE as described, for example, by Edgar ((2004) *BMC Bioinformatics* 5: 113 and *Nucleic Acids Res*. 32: 1792-1797). The multiple sequence alignment reveals the variety of amino acids found at each position and their relative frequency amongst all the sequences. This information can be useful in determining possible amino acid substitutions that may be made at identified sites in the protein.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1547 of 3077

In a non-limiting example, particular amino acid substitutions for a *Candida* spp. Pox5 acyl-CoA oxidase polypeptide are provided herein. For example, some substitutions were designed to modify a substrate specificity of an acyl-CoA oxidase polypeptide. As described herein, in embodiments in which the target product molecule is a six-carbon fatty acid (e.g., adipic acid) produced by β-oxidation of a longer-chain fatty acid, it is optimal to modify the activity of acyl-CoA oxidases (which can catalyze the first step in β-oxidation) in host cells or organisms such that there is little to no activity on substrates with chain lengths less than 8 carbons. Deletion of nucleic acids encoding acyl-CoA oxidases (e.g., Pox4 in *Candida viswanathii*) with relatively broad carbon-chain length specificity that are active on short-chain length substrates prevents generation of fatty acid products with fewer than eight carbon atoms (i.e., chain length shorter than C8) by peroxisomal β-oxidation. This is because the remaining acyl-CoA oxidase activity (e.g., Pox5) is specific for longer chain substrates and has low activity on substrates with carbon chain lengths less than 10. In order to increase the activity of Pox5 on substrates with a chain length of 8 carbons and thereby increase the amount of 6-carbon fatty acid target molecule products relative to 8-carbon fatty acid molecules, the Pox5 protein was subjected to engineering as described herein. Modified Pox5 proteins obtained by amino acid substitutions of the wild-type *Candida viswanathii* Pox5 protein (made via corresponding nucleotide sequence changes in the nucleic acid encoding the protein) resulted in an increased ratio of 6-carbon to 8-carbon fatty acid products in *Candida* cells expressing the modified enzymes as compared to cells expressing wild-type Pox5p (as shown by experimental results presented in the Examples herein). Using the teachings described herein, a primary structure of another acyl-CoA oxidase polypeptide can be aligned with the amino acid sequence or modeled structure of a Pox5 polypeptide and some or all amino acids of the other polypeptide that align with those selected for modification in the Pox5 polypeptide also can be selected for modification.

Additional nonlimiting examples of protein modifications that can be made in altering the carbon flux in a cell or organism include modifications to alter a substrate specificity of an acyl-CoA dehydrogenase polypeptide produced in the cell and that is involved in β-oxidation. An acyl-CoA dehydrogenase enzyme can require an NAD cofactor in carrying out a catalytic function. Sometimes a co-factor specificity of an acyl-CoA dehydrogenase is modified, and in some embodiments the modified polypeptide prefers to utilize oxygen as a co-factor.

In another non-limiting example of a protein modification designed to modify an enzyme activity, amino acid substitutions can be made to enhance or reduce regulation of the enzyme. For

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1548 of 3077

example, enzymes can be regulated in a number of ways, including, for example, covalent modification of an enzyme such as phosphorylation/dephosphorylation and acetylation/deacetylation. The activity of an enzyme can be modified by altering its ability to be activated or inhibited within a cell. In one embodiment, regulation of an enzyme by phosphorylation can be decreased or eliminated by modifying a nucleic acid encoding the enzyme to substitute codons for phosphorylatable amino acid residues (e.g., serine) with codons for non-phosphorylatable residues (e.g, alanine). Computer-assisted software programs are available for identifying potential phosphorylatable amino acid residues (see, e.g., NetPhos (World Wide Web Uniform Resource Locator (URL) cbs.dtu.dk/services/NetPhos/), NetPhosYeast (World Wide Web Uniform Resource Locator (URL) cbs.dtu.dk/services/NetPhosYeast/)). In an example described herein, an acetyl-CoA carboxylase protein (e.g., Acc1 of *Candida viswanathii*) is modified to reduce regulation of the enzyme by phosphorylation. Because the dephosphorylated state is the active state of the enzyme, the protein (and nucleic acid encoding it) was modified to eliminate one or more phosphorylatable serine residues by substituting them with alanine residues, thereby relieving the regulation by phosphorylation. For example, as described herein, a *Candida viswanathii* acetyl-CoA carboxylase endogenous, wild-type enzyme was modified to substitute alanine residues for one or more of the following serine amino acid residues: S652, S1131, S1138, S1153, S1158.

One or more activities of a modified polypeptide can be characterized using any suitable assay known in the art. A modified polypeptide can be expressed in a cell or organism other than a target organism in which a target product will be produced, for assaying activity. For example, a modified polypeptide can be expressed in a bacterium (e.g., *E. coli*), assayed and then introduced into a yeast (e.g., *Candida* spp. yeast) for production of a target molecule product.

Engineered carbon flux pathways for efficient production of target molecules

Provided herein are multiple compositions for, and methods of, modifying cells and organisms to alter carbon flux. Also provided are the modified cells and organisms generated by the methods. The modification methods can be combined in a number of ways as described herein to engineer cell- or organism-based systems for enhanced, efficient production of target molecules. Also provided herein are methods of producing target molecules, including, for

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1549 of 3077

example, organic acids, polyketides and terpenes, using the modified cells or organisms provided herein.

Included in the cells, organisms, compositions and methods provided herein are modified cells and organisms in which carbon processing activities have been engineered to enhance carbon flow through cellular oxidative metabolism pathways. One advantage of such modified bioproduction systems is that they are well suited for use with lower cost, alternative carbon sources, including, for example, non-carbohydrate and non-fermentable carbon sources such as aliphatic compounds and hydrocarbons (e.g, alkanes, fatty acids and fatty alcohols). Use of such carbon sources is not only more cost-effective but can also have the added advantage of reducing the environmental impact of harmful wastes (e.g., agro-industrial by-products, waste cooking oil and waste motor oil) that can be used as feedstocks in target molecule production instead of being discarded. As also described herein, embodiments of the cell- and microbial-based systems in which carbon processing activities have been engineered to direct carbon flow through oxidative metabolism can be controlled to provide for maximal, coordinated and highly efficient target molecule production based on, for example, use of carbon source-dependent transcription regulation.

Figs. 5-11 schematically illustrate non-limiting embodiments of engineered carbon flux pathways of modified cells and organisms that can be used to produce a target molecule (e.g., adipic acid, malonyl-CoA, 3-hydroxypropionic acid, polyketide, triacetic acid lactone, terpene) from various starting carbon sources or feedstocks.

Engineered carbon recycling loop pathways for a platform target molecule production system

In order to minimize carbon loss and increase process efficiency of biological cell-based production systems, provided herein are cells and organisms (e.g., microorganisms) that have been modified to reduce, inhibit, slow and/or delay carbon flow into one or more growth and/or energy production metabolic pathways so that it is available for use in other inherent and/or engineered production processes. In doing so, carbon that would be lost to metabolic pathways uninvolved in target molecule production in an unmodified cell or organism are rescued or captured for use in target molecule production processes. As such, modified cells and organisms provided herein in some embodiments are useful as platform systems (as well as

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1550 of 3077

production systems) that can be used as the basis for further engineering for enhanced production of many different desired target molecules either singly or multiply in co-production cell- and microbial-based systems.

Engineered pathways for capturing carbon atoms expelled from the β-oxidation pathway

Figs. 5 and 6 depict possible cellular modifications in exemplary embodiments of a eukaryotic (i.e., yeast in this example) platform system designed to capture carbon atoms in the cytosol by enhancing carbon flow through cellular oxidative metabolism pathways (ω-oxidation and peroxisomal β-oxidation) and reducing flow of carbon into mitochondria, and the endoplasmic reticulum and lipid particles (in the form of acyl-CoA). Although multiple, possible, cellular modifications are illustrated in Figs. 5 and 6, as described herein, some of the modifications depicted in the figures are optional enhancements of exemplary engineered systems and may or may not be included in a modified cell or organism depending on, for example, the intended use of the system (e.g., development of a particular single, or multiple, target molecule(s) production system) and the selection of variable features (e.g., host cell or organism, carbon source, regulatory controls (such as transcription control elements), culture conditions and the like) of the system. Thus, it is understood that any optional modifications set forth in the exemplary systems shown in Figs. 5 and 6 are non-limiting and may or may not be included in a particular engineered system and, if included, may be in utilized in different combinations than illustrated in the figures.

Acetyl-CoA is a central molecule in the biochemical network of a cell that can be utilized for the biosynthesis of many useful chemicals. The β-oxidation pathway produces acetyl-CoA through the oxidation of fatty acids. In yeast, β-oxidation can be localized to the peroxisomal compartment which thus can be a primary location of fatty acid-derived acetyl-CoA. Generally, the peroxisomal acetyl-CoA would be converted to acetyl-carnitine by carnitine acetyltransferase (Cat2p). The acetyl-carnitine, being smaller, can diffuse out of the peroxisome and be transported across the mitochondrial inner membrane by Crc1p (a translocase protein). Once inside the mitochondria, the acetyl-carnitine is converted back into acetyl-CoA by mitochondrial Cat2p and can be used in the TCA cycle for energy generation or the synthesis of other biomolecules (Fig. 4). To take advantage of the acetyl-CoA generated by peroxisomal β-oxidation, genetic engineering strategies may be employed to reroute the carbon

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1551 of 3077

in acetyl-CoA destined for the mitochondria to the cytosol instead, thereby making it available for use in biosynthetic pathways making desired chemical products. Thus, the amount of carbon that is lost to the TCA and lipid-generating (e.g., glycerol-3-phosphate (G3P) and/or dihydroxyacetone phosphate (DHAP)) pathways at the expense of target molecule production is reduced in this platform system. In embodiments of a platform system such as, for example, systems shown in Figs. 5 and 6, carbon processing activities can be engineered to enhance carbon flow through cellular oxidative metabolism pathways, e.g., ω-oxidation and peroxisomal β-oxidation, and decrease carbon flow to mitochondria and other organelles. Such platform systems can include an acetyl group carbon recycle loop that diverts acetyl moieties generated in the breakdown of fatty acids in peroxisomal β-oxidation into cytosolic fatty acid synthesis to regenerate a fatty acid that can be subjected to another cycle of peroxisomal β-oxidation. The recycle loop is depicted in Figs. 5 and 6 by the dark, solid reaction arrows beginning with extracellular fatty acid internalization in the upper left corner of the figure, extending through ω-oxidation and into the peroxisome for β-oxidation which yields acetyl-CoA that is transported out of the peroxisome and into the cytosol (initially in the form of either acetyl-carnitine or acetate), utilized in fatty acid synthesis to generate acyl-CoA which is then hydrolyzed to free fatty acid for re-entry into the loop at the starting point of ω-oxidation.

As shown in Figs. 5 and 6, carbon flux through the ω-oxidation and β-oxidation (peroxisomal) pathways can be enhanced through one or more of multiple modifications introduced via genetic manipulation of the cell. The enhancements can begin with the cellular internalization of external carbon. A non-fermentable or alternative carbon source (e.g., fatty acids, alkanes) enters the modified cell through the plasma membrane from the extracellular medium (shown in the upper left corner of Figs. 5 and 6). In an unmodified cell, a long-chain fatty acid (either as the carbon source or generated from processing of a carbon source, e.g., alkane) would be activated (through thioesterifcation with CoA) to acyl-CoA upon cell entry by acyl-CoA synthetase (encoded, e.g., by *FAA1* and/or *FAT1*). However, in some embodiments, a gene(s) encoding cytosolic and/or membrane-bound acyl-CoA synthetase can be disrupted or deleted resulting in a decrease or elimination of cytosolic and/or membrane-bound acyl-CoA synthetase and, thus, cytosolic and/or membrane-bound acyl-CoA synthetase activity in the cytosol. Cytosolic activation of fatty acids can thus also be decreased or eliminated in such modified cells (indicated in Figs. 5 and 6 as a lightly shaded dotted line reaction arrow labeled as “faa1Δ” and “fat1Δ” in blackened ovals). Most fatty acid metabolic pathways, including lipid (e.g, triacylglycerides (TAG) and phospholipids (PL)) biosynthesis and protein acylation, require that

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1552 of 3077

a free fatty acid be activated to acyl-CoA (or to acyl-ACP) prior to being metabolized. Therefore, in embodiments that include an enhancement such as a reduced or abolished acyl-CoA synthetase activity, the decreased cytosolic fatty acid activation can result in fewer internalized fatty acid carbons being lost to such pathways at the expense of target molecule-producing processes.

Free fatty acids that have entered the cell (shown as "FA" in Figs. 5 and 6), or generated from metabolism of an alkane carbon source, can then undergo oxidation to dicarboxylic acids (DCA) through ω-oxidation ("ω-ox" in Figs. 5 and 6). The availability of this oxidative process in the cell presents multiple advantages in these platform systems for target molecule production. For example, long-chain fatty acids that have not been activated to acyl-CoA do not readily cross the peroxisomal membrane; however, long-chain dicarboxylic acids are able to enter peroxisomes. Therefore, conversion of free fatty acids to DCA through ω-oxidation can be a further enhancement of carbon flow toward peroxisomal β-oxidation, particularly because there are no or limited other pathways in the cell for processing of free dicarboxylic acids. The availability of ω-oxidation-processing of free fatty acids in the modified cell is also beneficial to engineered production systems in which the target molecule (or an intermediate in target molecule production) is a dicarboxylic acid (e.g., adipic acid, suberic acid, sebacic acid, dodecanedioic acid, tetradecanedioic acid). For example, as shown in Figs. 5 and 6, under certain conditions, a dicarboxylic acid processed in β-oxidation can be converted into a shorter chain diacid which can be secreted from the cell as a target molecule upon removal of the coenzyme A carrier via hydrolysis catalyzed by peroxisomal thioesterase. Thus, the ω-oxidation pathway in a modified cell can serve as a cellular gateway for funneling internalized fatty acids into oxidative metabolism and target molecule production and away from cytosolic activation. Modification of ω-oxidation activity is another potential enhancement of these embodiments. For example, if a host cell or organism (e.g., *Candida* spp, *Yarrowia* spp, *Bacillus* spp, *Blastobotrys* spp) expresses an endogenous ω-oxidation pathway, one or more enzymes (e.g., monooxygenase, cytochrome P450 reductase, such as CPRB, and others) of the pathway can be modified (e.g., as described herein) to increase catalytic activity and/or alter substrate specificity in order to increase fatty acid processing in the pathway and/or target specific fatty acids for processing into dicarboxylic acids. If a host cell or organism does not express an endogenous ω-oxidation pathway, it can be genetically modified to express heterologous enzymes to engineer an ω-oxidation pathway in the cell or organism.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1553 of 3077

In an oxidative metabolism-enhanced platform system, dicarboxylic acids, such as those generated by ω-oxidation, can traverse the peroxisomal membrane and move into peroxisomes where they can be activated via thioesterification to a dicarboxylic acid ester (shown as DCA-CoA in Figs. 5 and 6) and enter β-oxidation ("β -ox" in Figs. 5 and 6). In each cycle of β-oxidation, fatty acids are degraded through removal of two carbons from the carbon chain which are released as acetyl-CoA. The remaining fatty acid carbon chain can reenter another cycle of oxidation as an acyl-CoA shortened by two carbons atoms. Through successive cycles, a monocarboxylic fatty acid can be completely degraded such that only acetyl-CoA (for fatty acids with an even number of carbon atoms in the chain) or propionyl-CoA (for fatty acids with an odd number of carbon atoms in the chain) remains. Through successive cycles of β-oxidation of a dicarboxylic acid, the molecule can be completely degraded such that only succinyl-CoA (for fatty diacids with an even number of carbon atoms in the chain) or malonyl-CoA (for fatty diacids with an odd number of carbon atoms in the chain) remains. Thus, in these platform systems, the enhanced movement of fatty acids toward degradation via β-oxidation can yield acetyl-CoA (which can be used in target molecule production), and, in certain instances as described herein, shorter chain diacids at the completion of the oxidative process. A short-chain diacid thus produced can be a final target molecule (or a precursor or intermediate in the production of a target molecule).

The oxidative metabolism aspect of some platform systems, such as those shown in Figs. 5 and 6, can be further enhanced through modification of β-oxidation activity. For example, one or more enzymes (e.g., acyl-CoA oxidase, ketoacyl-CoA thiolase, multifunctional enzyme hydratase and/or dehydrogenase, and others) of the pathway can be modified (e.g., as described herein) to increase catalytic activity and/or alter substrate specificity in order to increase fatty acid processing in the pathway and/or select for specific fatty diacids for processing into target dicarboxylic acids. One example of a modification of β-oxidation activity, as described herein, is alteration of the substrate specificity of one or more acyl-CoA oxidase enzymes in the pathway, such as Pox4 and/or Pox5 of *Candida* yeast strains. In so doing, the process can be optimized for the production of fatty diacids of particular carbon chain lengths. For example, by genetically modifying a host cell or microorganism to decrease or eliminate Pox4 expression and/or activity in the host (e.g., *Candida*), the amount of shorter-chain (e.g., having less than about 8-10 carbons) fatty acids or diacids resulting from β-oxidation of longer chain fatty acids can be increased. Production of fatty acids or diacids of particular lengths can also be enhanced by genetically modifying (e.g., mutagenesis of the gene coding sequence to

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1554 of 3077

alter the encoded amino acid sequence) the activity of another acyl-CoA oxidase, such as Pox5, to alter the substrate specificity. For example, as described herein, some alterations of a *Candida* Pox5 amino acid sequence increase activity of the enzyme on C8 substrates and provide for a relative increase in the amount of C6 diacid (adipic acid) produced and decrease in the amount of C8 and longer diacids resulting from β-oxidation of a longer chain fatty acid. Thus, the platform system shown in Figs. 5 and 6 can also serve as a production system for diacids of particular carbon chain lengths. Additional optional modifications that can provide for enhanced carbon flux through β-oxidation (and enhanced target molecule production) in these systems include, but are not limited to, modification of β-oxidation-associated activities, such as peroxisome biogenesis and proliferation activities. For example, as described herein, the abundance and/or volume of peroxisomes in which β-oxidation occurs can be increased in host cells through genetic modification. An example of such a modification is increasing the transcription of, and/or number of copies of, one or more peroxin-encoding nucleic acids (e.g., *PEX11*) in a host cell. Amplification of such peroxin-encoding nucleic acids and/or activities can lead to an overall increased β-oxidation capacity.

One feature of the carbon recycle loops of the platform systems shown in Figs. 5 and 6 is the management and capture of acetyl-CoA generated during β-oxidation. Peroxisomal acetyl-CoA generally has two main fates: (i) conversion to acetyl-carnitine for transfer to mitochondria for use in the TCA cycle and (ii) the generation of malate in the glyoxylate cycle ("GlyOx" in Figs. 5 and 6) which is then used in gluconeogenesis or moves into mitochondria. In unmodified cells, these uses of acetyl-CoA generated in β-oxidation represent loss of carbon atoms that could be used in target molecule production. Through modifications that are a part of the platform systems shown in Figs. 5 and 6, acetyl groups can be captured either (1) as they move through the cytosol toward the mitochondria in the form of acetyl-carnitine or (2) in the form of acetate generated in peroxisomes. In capturing these acetyl group carbons, they thus can be diverted from the TCA cycle.

Capture of carbon from acetyl-carnitine

In the example platform system depicted in Fig. 5, modifications in the host cell that enhance the capture and diversion of acetyl group carbon include, but are not limited to, modification of acetyl-carnitine entry into mitochondria, and modification of conversion of cytosolic acetyl-carnitine to acetyl-CoA. As described herein, unmodified cells may contain a cytosolic carnitine

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1555 of 3077

acetyltransferase activity for conversion of cytosolic acetyl-carnitine to acetyl-CoA. However, in some instances, it may not be as catalytically active and/or abundant as it is in organelles, e.g., peroxisomes and mitochondria. In a platform system provided herein and depicted in Fig. 5, one modification that can be made to enhance capture of acetyl groups in the cytosol is to increase the amount and/or activity of cytosolic carnitine acetyltransferase. As described herein, methods of achieving this include increasing the copy number of nucleic acids encoding cytosolic carnitine acetyltransferase in the cell, increasing the transcription of such nucleic acids and/or introducing nucleic acid encoding a more active cytosolic carnitine acetyltransferase enzyme into the cell (e.g., modifying an endogenous cytosolic enzyme activity by replacing it with, or adding to it, a heterologous enzyme activity). For example, in one embodiment described herein, a *Candida* mitochondrial/peroxisomal carnitine acetyltransferase (e.g, Cat2) with greater catalytic activity than an endogenous *Candida* cytoplasmic carnitine acetyltransferase (e.g., Yat1) can be recombinantly expressed cytosolically in a host cell by engineering a nucleic acid encoding the more active enzyme such that the encoded enzyme lacks a mitochondrial (and a peroxisomal) targeting sequence of amino acids (shown, as $CAT2^{cyt}$ in Fig. 5; see also, e.g., amino acid SEQ ID NO: 4 and a nucleotide sequence (SEQ ID NO: 61) encoding the amino acid sequence). Once acetyl-carnitine in transit from the peroxisomes to the mitochondria has been converted to acetyl-CoA in the cytosol by carnitine acetyltransferase activity in the cytosol (e.g., $CAT2^{cyt}$), it cannot cross the mitochondrial inner membrane and is diverted from the TCA cycle. This acetyl-CoA is now available for use in target molecule production. The amount of carnitine acetyltransferase activity in the cytosol of such a modified cell or organism can be further increased by using a strong and/or fatty acid-inducible heterologous promoter (e.g., a yeast *HDE* gene promoter) to regulate transcription of the engineered nucleic acid encoding a carnitine acetyltransferase activity.

Another modification that can enhance cytosolic capture and diversion of acetyl moieties in cells is an alteration of acetyl-carnitine uptake into mitochondria from the cytosol (shown as faded, dotted reaction arrow lines into and in the mitochondrial compartment in Fig. 5). One method of modifying mitochondrial acetyl-carnitine uptake is by altering the processing of acetyl-carnitine that occurs in the mitochondria to convert it to acetyl-CoA for use in the TCA cycle. For example, by decreasing the amount and/or activity level of the enzyme that catalyzes this processing, i.e., mitochondrial carnitine acetyltransferase, there can be a corresponding decrease in conversion of acetyl-carnitine to acetyl-CoA in the mitochondria. Without being limited or bound by theory, this can introduce a bottleneck in acetyl-carnitine processing in the

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1556 of 3077

mitochondria which slows acetyl-CoA entry into the TCA cycle. If the mitochondrial carnitine acetyltransferase activity is not sufficient to handle the acetyl carbon flux coming from the peroxisome, then the cytoplasmic acetyl-carnitine concentration should build up and, in effect, acetyl-carnitine is diverted from the TCA cycle. The increased concentration of cytoplasmic acetyl-carnitine can thus be a source of substrate for carnitine acetyltransferase activity in the cytosol which converts the substrate to cytosolic acetyl-CoA for use in target molecule production. The amount and/or activity of mitochondrial carnitine acetyltransferase can be decreased in a number of ways, as described herein. For example, the number of copies of nucleic acid encoding the enzyme in a host cell can be reduced (e.g, an endogenous gene encoding the enzyme can be disrupted or deleted), the transcription of such nucleic acid can be decreased and/or nucleic acid encoding a less active mitochondrial carnitine acetyltransferase enzyme can be introduced into the cell (i.e., replacing the endogenous mitochondrial enzyme with a heterologous enzyme). For example, in one embodiment described herein, a *Candida* cytoplasmic carnitine acetyltransferase (e.g, Yat1) which is less active than an endogenous *Candida* mitochondrial carnitine acetyltransferase (e.g., Cat2) can be recombinantly expressed in a host cell mitochondria by engineering a nucleic acid encoding the less active enzyme such that the encoded enzyme includes a mitochondrial targeting sequence of amino acids (shown as "*CAT2*" in a diagonal line-hatched background in the mitochondria in Fig. 5; see also, e.g., amino acid SEQ ID NOS: 10, 11 and 12 and nucleotide SEQ ID NOS: 67, 68 and 69 encoding such amino acid sequences). The modified nucleic acid can be introduced into a host cell in which the endogenous mitochondrial carnitine acetyltransferase gene has been disrupted or deleted. Although not specifically indicated in Fig. 5, in some cells and organisms (e.g., *Candida* spp), an endogenous gene encoding a mitochondrial carnitine acetyltransferase may also encode the cell's peroxisomal carnitine acetyltransferase. For example, such a gene can encode an enzyme that includes mitochondrial and peroxisomal targeting sequences for localization to each of these areas of the cell. If the gene encoding an endogenous mitochondrial carnitine acetyltransferase in such a cell is disrupted or deleted, it may be optimal (e.g., for cell viability and/or efficient processing of peroxisomal acetyl-CoA) to introduce a heterologous nucleic acid encoding carnitine acetyltransferase that includes a peroxisomal targeting sequence into the cell.

Another method of modifying mitochondrial acetyl-carnitine uptake is by altering a transport mechanism that moves acetyl-carnitine into the mitochondrial matrix, e.g., an acetyl-carnitine translocase. A mitochondrial inner-membrane transport protein (e.g., Crc1p) may function as an

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1557 of 3077

acetyl-carnitine transporter providing for transport of acetyl-carnitine into the mitochondrial matrix. By decreasing the amount and/or activity level of the transport protein, movement of acetyl-carnitine from the cytosol into mitochondria can be reduced thereby increasing the concentration of acetyl-carnitine in the cytosol that can be converted to acetyl-CoA by cytoplasmic carnitine acetyltransferase. The amount and/or activity of mitochondrial acetyl-carnitine transport protein can be decreased in a number of ways, as described herein. For example, the number of copies of nucleic acid encoding a mitochondrial acetyl-carnitine transport protein in a host cell can be reduced (e.g, an endogenous gene encoding the protein can be disrupted or deleted), the transcription of such nucleic acid can be decreased and/or nucleic acid encoding a less active transport protein can be introduced into the cell (e.g., replacing the endogenous mitochondrial transport protein with a heterologous protein). For example, in one embodiment described herein, the transcription of nucleic acid encoding a *Candida* acetyl-carnitine translocase (shown as "CRC1" in a diagonal line-hatched background in Fig. 5) can be reduced in a host cell by introducing such nucleic acid, which is operably linked to a heterologous promoter (e.g., a yeast glucose-6-phosphate isomerase gene promoter such as, for example, SEQ ID NO: 118) that provides for less transcription and/or a reduced transcription rate, and/or that can be regulated to provide for alternately weak and stronger transcription, into host cells in which the endogenous gene has been disrupted. The resulting reduction in transcription of the nucleic acid results in decreased amounts of the transporter protein in the mitochondrial membrane of modified cells. Thus, modification of mitochondrial acetyl-carnitine transporter expression and/or mitochondrial carnitine acetyltransferase activity, as shown in Fig. 5, can serve to divert acetyl-carnitine from use in the TCA cycle and increase the concentration of acetyl-carnitine in the cytosol of cells including one or more of these modifications. This, combined with increased amounts and/or activity of cytosolic carnitine acetyltransferase can result in increased amounts of acetyl-CoA available in the cytosol for use in target molecule production. Although all three of these modifications, i.e., alteration of mitochondrial carnitine acetyltransferase activity (e.g., Cat2), alteration of mitochondrial acetyl-carnitine transporter protein activity (e.g., Crc1) and modification of cytosolic carnitine acetyltransferase activity (e.g., Yat1), are shown in the depiction of an example platform system in Fig. 5, each can be used singly or in any combination in modifying carbon processing in cells or microorganisms for the production of target molecules. For example, modifications of a host cell in generating a platform and/or production system with respect to these three modifications include, but are not limited to: (1) a decreased mitochondrial carnitine acetyltransferase expression and/or activity and an increased cytosolic carnitine acetyltransferase expression

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1558 of 3077

and/or activity, (2) a decreased mitochondrial acetyl-carnitine transporter protein expression and/or activity and an increased cytosolic carnitine acetyltransferase expression and/or activity, (3) a decreased mitochondrial carnitine acetyltransferase expression and/or activity, a decreased mitochondrial acetyl-carnitine transporter protein expression and/or activity, and an increased cytosolic carnitine acetyltransferase expression and/or activity, (4) a decreased mitochondrial carnitine acetyltransferase expression and/or activity, and a decreased mitochondrial acetyl-carnitine transporter protein expression and/or activity, (5) a decreased mitochondrial carnitine acetyltransferase expression and/or activity, (6) a decreased mitochondrial acetyl-carnitine transporter protein expression and/or activity and (7) an increased cytosolic carnitine acetyltransferase expression and/or activity.

Capture of carbon from acetate

In the platform system depicted in Fig. 6, another example of a modification of a host cell that can enhance the capture and diversion of acetyl group carbons that have been generated in β-oxidation is modification of acetyl-CoA processing in peroxisomes. In a cell that does not include such a modification, acetyl-CoA generated during the degradation of fatty acids in β-oxidation typically is converted by peroxisomal carnitine acetyltransferase into acetyl-carnitine for transport into the cytosol and eventually to the mitochondria. The engineered carbon recycling loop of the embodiment of the platform system depicted in Fig. 6 can capture the acetyl group carbons through conversion of peroxisomal acetyl-CoA into acetate. The acetate readily traverses the peroxisomal membrane and can move into the cytosol where it can be reconverted to acetyl-CoA and thus diverted from entry into, and loss to, mitochondrial metabolism.

Modification of peroxisomal acetyl-CoA processing in a host cell can be accomplished, for example, as described herein. Peroxisomal acetyl-CoA can be converted to acetate through hydrolysis catalyzed by acetyl-CoA hydrolase which also liberates coenzyme A for reuse in β-oxidation. A cell or microorganism may be modified to increase (or introduce) acetyl-CoA hydrolase and/or acetyl-CoA hydrolase activity in the peroxisomes. For example, the copy number of nucleic acids encoding a peroxisomal acetyl-CoA hydrolase in the cell can be increased, transcription of such nucleic acids can be increased and/or, if a cell expresses an endogenous peroxisomal acetyl-CoA hydrolase, a nucleic acid encoding a more active hydrolase enzyme can be introduced into the cell (e.g., modifying an endogenous peroxisomal

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1559 of 3077

hydrolase enzyme activity by replacing it with, or adding to it, a heterologous enzyme activity). For example, in one embodiment described herein, a *Candida* acetyl-CoA hydrolase (e.g, Ach) that is expressed in mitochondria of unmodified cells can be recombinantly expressed in host cell peroxisomes by engineering a nucleic acid encoding the enzyme such that the encoded protein lacks a mitochondrial targeting sequence of amino acids and includes a peroxisomal targeting sequence (shown as *ACH* in the peroxisomal compartment in Fig. 6; see e.g, amino acid SEQ ID NO: 16 and a nucleotide sequence (SEQ ID NO: 73) encoding the amino acid sequence). The amount of acetyl-CoA hydrolase activity in peroxisomes of such a modified cell or organism can be further increased by using a strong and/or fatty acid-inducible heterologous promoter (e.g., a yeast *HDE* gene promoter) to regulate transcription of the engineered nucleic acid encoding an acetyl-CoA hydrolase activity. In order to reduce or eliminate peroxisomal conversion of acetyl-CoA to acetyl-carnitine (shown as "Ac-Carn" in the peroxisome compartment of Fig. 6) so that a maximal amount of the acetyl-CoA is converted to acetate, the host cell can also be modified to decrease or eliminate (e.g., by disrupting or deleting a gene encoding peroxisomal carnitine acetyltransferase) carnitine acetyltransferase activity in the peroxisomes (shown in Fig. 6 as "*cat2Δ*" in a black oval).

Once in the cytoplasm, the acetate can be converted to acetyl-CoA by acetyl-CoA synthetase (shown as "*ACS*" in Fig. 6) which catalyzes the ligation of acetate and coenzyme A to produce acetyl-CoA. To provide optimal processing of the increased cytosolic acetate generated in this embodiment into cytosolic acetyl-CoA, the amount and/or activity of cytosolic acetyl-CoA synthetase can be increased in modified host cells or organisms. For example, the copy number of nucleic acids encoding a cytosolic acetyl-CoA synthetase in the cell can be increased, transcription of such nucleic acids can be increased (e.g., using a heterologous strong and/or fatty acid-inducible promotor, for example, a yeast *HDE* gene promoter) and/or a nucleic acid encoding a more active acetyl-CoA synthetase enzyme can be introduced into the cell (e.g., modifying an endogenous cytosolic acetyl-CoA synthetase enzyme activity by replacing it with, or adding to it, a heterologous enzyme activity).

As described herein, in some host cells or organisms, e.g., some yeast species, such as *Candida*, peroxisomal and mitochondrial carnitine acetyltransferase may be encoded by the same gene which can contain two in-frame start codons. Carbon source-dependent alternate transcription initiation can result in expression of a carnitine acetyltransferase initiated from the first start codon or a shorter carnitine acetyltransferase initiated from the second start codon.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1560 of 3077

The longer version of carnitine acetyltransferase encodes an N-terminal mitochondrial targeting signal wherease the shorter version does not. Therefore, in an embodiment in which a peroxisomal carnitine acetyltransferase activity is decreased by disruption or deletion of the corresponding gene in such cells, the mitochondrial carnitine acetyltransferase activity may also be decreased or eliminated. Therefore, because cell survival may require a minimal amount of energy generated through mitochondrial metabolism which in turn requires a supply of acetyl-CoA, a host cell or microorganism for the platform system shown in Fig. 6 can also be modified to express a mitochondrial carnitine acetyl transferase. This can be accomplished by introducing a recombinant nucleic acid encoding a carnitine acetyltransferase that includes a mitochondrial targeting sequence of amino acids into the host cell or microorganism. Because a system such as that shown in Fig. 6 is designed to direct most of the acetyl group carbons from acetyl-CoA generated in β-oxidation to the next segment of the carbon recycling loop (i.e., cytosolic fatty acid synthesis), the recombinant mitochondrial carnitine acetyltransferase can be one with reduced catalytic activity relative to the endogenous mitochondrial carnitine acetyltransferase (e.g., a Yat1 enzyme instead of a Cat2 enzyme). This can serve to minimize the rate at which any acetyl-carnitine that does enter the mitochondria is converted to acetyl-CoA which, in turn, could minimize the amount of carbon loss to mitochondrial metabolism at the expense of the recycling mechanism in this system. Loss of carbon to mitochondrial metabolism could also be minimized (in addition to, or as an alternative to, introducing nucleic acid encoding a less active mitochondrial carnitine acetyltransferase) by decreasing the amount and/or activity level of a mitochondrial acetyl-carnitine translocase (e.g., Crc1) into the system shown in Fig. 6. This can result in slowing of acetyl-carnitine transport into mitochondria and serve to divert some of the cytoplasmic acetyl-carnitine from use in the TCA cycle. Although both of these modifications, i.e., alteration of mitochondrial carnitine acetyltransferase activity (e.g., Cat2) and alteration of mitochondrial acetyl-carnitine transporter protein activity (e.g., Crc1), are shown in the depiction of an example platform system in Fig. 6, each can be used singly or in combination in modifying carbon processing in cells or microorganisms for the production of target molecules.

Additionally, because the only cytoplasmic acetyl-carnitine being generated in the system shown in Fig. 6 is through the carnitine acetyltransferase activity present in the cytosol (e.g., endogenous Yat1p in a *Candida* cell), the amount of acetyl-carnitine available to mitochondria is limited due to competition between the cytosolic carnitine acetyltransferase and acetyl-CoA carboxylase (shown as “ACC1” in Fig. 6) for the acetyl-CoA substrate. In some instances, it

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1561 of 3077

may be optimal to increase the amount and/or activity of carnitine acetyltransferase in the cytosol in a system such as that depicted in Fig. 6 in order to insure sufficient generation of acetylcarnitine for any minimal amount of acetylcarnitine that may be needed for mitochondrial metabolism. This can be accomplished, for example, by increasing the copy number of nucleic acids encoding cytosolic carnitine acetyltransferase in the cell, increasing the transcription of such nucleic acids and/or introducing nucleic acid encoding a more active cytosolic carnitine acetyltransferase enzyme into the cell (e.g., modifying an endogenous cytosolic enzyme activity by replacing it with, or adding to it, a heterologous enzyme activity (e.g., Cat2p)).

Redirecting carbon flow toward fatty acid biosynthesis

At this point in an acetyl group carbon recycling loop, such as that illustrated in the systems depicted in Figs. 5 and 6, when acetyl moieties are accumulated in the cytosol as acetyl-CoA, they are primarily directed into the cytosolic fatty acid synthesis pathway to regenerate fatty acids. Cytoplasmic acetyl-CoA can be converted into malonyl-CoA, which can be a carbon donor in the synthesis of a fatty acid chain in repeated cycles of the addition of 2 carbon atoms per cycle to extend the chain and generate a fatty acid. The reactions of each cycle are typically catalyzed by fatty acid synthase (FAS) and generally continue until a 16- or 18-carbon fatty acid (palmitic acid or stearic acid) is completed in the form of palmitoyl-CoA or stearoyl-CoA. To enhance flow of the accumulated cytosolic acetyl-CoA into the fatty acid biosynthesis pathway, the process of converting acetyl-CoA into malonyl-CoA can optionally be modified in host cells or microorganisms. For example, the amount and/or activity of an enzyme that can catalyze the reaction, acetyl-CoA carboxylase, can be modified, e.g., increased, in host cells or microorganisms using genetic engineering methods as described herein (e.g., increasing the copy number and/or transcription of nucleic acid encoding acetyl-CoA carboxylase, increasing the activity of the enzyme by introducing nucleic acid encoding a modified amino acid sequence of the enzyme into host cells). In one embodiment described herein, the activity of a yeast (e.g., *Candida*) cytosolic acetyl-CoA carboxylase (shown as "*ACC1*" in a diagonal line-hatched background oval in Figs. 5 and 6) can be increased through substitution of phosphorylatable serine residues with alternate (e.g., alanine) residues to reduce inhibition of the enzyme by phosphorylation (see, e.g., SEQ ID NO: 19). Heterologous nucleic acid encoding the modified protein can be introduced into a host cell for expression of the enzyme therein. Increasing the amount and/or activity of acetyl-CoA carboxylase can reduce or prevent any limitation on flow of accumulated acetyl-CoA into fatty acid biosynthesis due to insufficient enzyme activity. The

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1562 of 3077

amount of cytosolic acetyl-CoA carboxylase activity in such a modified cell or organism can also be increased by using a strong and/or fatty acid-inducible heterologous promoter (e.g., a yeast *HDE* gene promoter) to regulate transcription of the engineered nucleic acid encoding an acetyl-CoA carboxylase activity.

Additional optional modifications of cells or organisms to enhance the platform systems shown in Figs. 5 and 6 include increasing the amount, activity, and/or altering the specificity, of enzymes in the fatty acid synthesis (FAS) enzyme complex, shown as "FAS" in Figs. 5 and 6 (e.g., the enzyme activities of the FAS1 and FAS2 subunits of yeast). A fatty acid synthase (e.g., FAS) activity can catalyze a series of decarboxylative Claisen condensation reactions from acetyl-CoA and malonyl-CoA. Without being limited or bound by any theory, it is believed that following each round of elongation the beta keto group is reduced to the fully saturated carbon chain by the sequential action of a ketoreductase activity, a dehydratase activity, and an enol reductase activity. In the case of Type I FAS enzymes, the growing fatty acid chain typically is carried between these active sites while attached covalently to the phosphopantetheine prosthetic group of an acyl carrier protein (ACP), and can be released by the action of a thioesterase (TE) upon reaching a carbon chain length of, for example, 16 (e.g., palmitic acid). In some instances, the collection of activities is found in a multifunctional, multi-subunit protein complex (e.g., Type I FAS activity). A fatty acid synthase enzyme (FAS) can be coded by fatty acid synthase subunit alpha (FAS2) and fatty acid synthase subunit beta (FAS1) genes. Thus, a fatty acid synthase activity usually includes a collection of activities (e.g., an enzymatic system) that perform functions associated with the synthesis of fatty acids. Therefore, the terms "fatty acid synthase activity", "fatty acid synthase", "FAS", and "FAS activity", as used herein refer to a collection of activities, or an enzymatic system, that perform functions associated with the synthesis of fatty acids. Fatty acid synthase activity may be amplified by over-expression of the FAS2 and/or FAS1 genes by any suitable method. Non-limiting examples of methods suitable to amplify or over express FAS2 and FAS1 genes include amplifying the number of FAS2 and/or FAS1 genes in a host cell following transformation with a high-copy number plasmid (e.g., such as one containing a 2u origin of replication), integration of multiple copies of FAS2 and/or FAS1 genes into the host genome, over-expression of the FAS2 and/or FAS1 genes directed by a strong and/or fatty acid-inducible promoter, the like or combinations thereof. Examples of polynucleotides from *Candida* strain ATCC 20336 that encode fatty acid synthase molecules (FAS1, FAS2) are provided herein (SEQ ID NOS: 102

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1563 of 3077

and 103) and are also described in International patent application no. PCT/US2012/056562 (publication no. WO 2012/056562).

Redirecting synthesized fatty acids toward ω- and β-oxidation

A typical product of cytosolic fatty acid synthesis is an acyl-CoA (e.g., palmitoyl- or stearoyl-CoA), which is shown as "FA-CoA" in Figs. 5 and 6. Because this is an activated (i.e., thioester) form of a fatty acid, it can be used in cellular metabolic pathways (e.g., synthesis of triacylglycerides and phospholipids) other than desired engineered target molecule production processes. This represents a loss of the carbon atoms in the acyl-CoA which could have been incorporated into target products. To reduce loss of the cytosolic acetyl group carbons (now in the form of an acyl-CoA) captured in the recycling loop of the exemplary platform systems shown in Figs. 5 and 6, the final segment of the loop often includes an engineered cytosolic thioesterase enzyme to, in effect, "deactivate" the fatty acid-CoA through hydrolysis and removal of coenzyme A. This can divert the carbons in the acyl-CoA from use in cellular processes not involved in target molecule production and/or generation of cytosolic acetyl-CoA and can complete the recycling loop by generating a cytosolic free fatty acid that can then begin the loop pathway at the initial point of ω-oxidation. Many cells (e.g., some eukaryotic cells, such as yeast) do not have a cytosolic thioesterase activity that is active on long-chain acyl-CoA substrates. Host cells that express an endogenous cytosolic thioesterase activity may also benefit from increasing the activity to enhance the flow of the acyl-CoA carbons through the final segment of the recycling loop. In the embodiment of the platform system shown in Fig.5, a host cell can be modified to increase (in this case by introducing) a thioesterase activity in the cytosol in order to direct acyl-CoA carbon flux toward oxidative metabolism pathways (ω- and β-oxidation). This can be achieved using genetic engineering methods as described herein (e.g., increasing the copy number and/or transcription of nucleic acid encoding a thioesterase, increasing the activity of the enzyme by introducing nucleic acid encoding a modified amino acid sequence of the enzyme into host cells). For example, in one embodiment described herein, a *Candida* peroxisomal thioesterase (e.g, Tes) can be recombinantly expressed cytosolically in a host cell by engineering a nucleic acid encoding the enzyme such that the encoded enzyme lacks a peroxisomal targeting sequence of amino acids (shown, as $TES3^{cyt}$ in a diagonal line-hatched background oval in Figs. 5 and 6; see also, e.g, nucleotide SEQ ID NO: 88 and encoded amino acid SEQ ID NO: 32). Some organisms, such as some yeast (e.g., *Candida*), for example, may express several distinct thioesterases (e.g., *Candida viswanathii* has 8

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1564 of 3077

peroxisomal thioesterase genes) having varied activities. The activities of thioesterases encoded by different genes can be evaluated, using methods known in the art and/or described herein, to compare the enzymes and select the type and level of activity that is optimal for achieving conversion of cytosolic acyl-CoA to free fatty acid in a recycle loop such as the one depicted in Figs. 5 and 6.

Another modification that can optionally be included in platform systems, such as those depicted in Figs. 5 and 6, is a decrease in the amount and/or activity of, or elimination of, mechanisms for the transport of acyl-CoA across the peroxisomal membrane and into peroxisomes. This modification can be beneficial in embodiments in which a target molecule (or precursor or intermediate in the production of a target molecule) is a dicarboxylic acid. For example, the platform systems shown in Figs. 5 and 6 can be used as a modified cell or microorganism for the enhanced production of dicarboxylic acids via ω- and β-oxidation of a fatty acid or alkane carbon source. Feedstock fatty acid (or alkane) carbon atoms that would have been lost as acetyl-CoA formed during β-oxidation degradation of a long-chain diacid (that had been generated by ω-oxidation) in an unmodified cell are captured and used in the generation of additional target diacid molecules (e.g., adipic acid) through the engineered recycling loops in these diacid production systems. Thus, a recycling loop such as the one depicted in Fig. 5 or Fig. 6 can provide for enhanced, highly efficient fatty acid production by significantly reducing "waste" of feedstock carbons in other cellular processes not involved in target molecule production. When the target molecule is a diacid, the carbon atoms recycled through the cytosolic fatty acid synthesis segment of the loop can optimally be directed through the final loop segment of conversion of the synthesized acyl-CoA into free fatty acid so that the carbon atoms can be used again in generating more diacid target through ω- and β-oxidation. Therefore, it can be beneficial to reduce or eliminate any transport of the synthesized acyl-CoA across the peroxisomal membrane and into peroxisomes where it would directly enter into β-oxidation without first being converted to a diacid through ω-oxidation. This can be achieved through disruption or deletion of genes encoding peroxisomal acyl-CoA transporters. In one embodiment described herein, a yeast (e.g., *Candida*) peroxisomal transport protein (e.g, Pxa1) activity is decreased or eliminated by disrupting the gene encoding the protein in the host cell (e.g., shown as "pxa1Δ" in a black background oval in Figs. 5 and 6).

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1565 of 3077

Engineered pathways for capturing carbon atoms from the TCA cycle

Fig. 7 depicts possible cellular modifications in an exemplary embodiment of a eukaryotic (i.e., yeast in this example) platform system designed to capture carbon atoms in the cytosol by enhancing carbon flow through cellular oxidative metabolism pathways (ω-oxidation and peroxisomal β-oxidation) and reducing flow of carbon into the endoplasmic reticulum and lipid particles (in the form of acyl-CoA). Thus, the amount of carbon that is lost to lipid-generating (e.g., glycerol-3-phosphate (G3P) and/or dihydroxyacetone phosphate (DHAP)) pathways at the expense of target molecule production can be reduced in this platform system. Although multiple, possible, cellular modifications are illustrated in Fig. 7, as described herein, some of the modifications depicted in the figure are optional enhancements of an exemplary engineered system and may or may not be included in a modified cell or organism depending on, for example, the intended use of the system (e.g., development of a particular single, or multiple, target molecule(s) production system) and the selection of variable features (e.g., host cell or organism, carbon source, regulatory controls (such as transcription control elements), culture conditions and the like) of the system. Thus, it is understood that any optional modifications set forth in the exemplary system shown in Fig. 7 are non-limiting and may or may not be included in a particular engineered system and, if included, may be in utilized in different combinations than illustrated in the figure.

In the system shown in Fig. 7, as in the embodiments of the platform systems depicted in Figs. 5 and 6, carbon processing activities can be engineered to enhance carbon flow through cellular oxidative metabolism pathways, e.g., ω-oxidation and peroxisomal β-oxidation. However, unlike the embodiments depicted in Figs. 5 and 6, carbon flow from the peroxisomes to the mitochondria is usually not decreased in the platform system shown in Fig. 7. Instead, a portion of the acetyl carbon that is allowed to be utilized in mitochondrial metabolism in this embodiment can be captured from citrate that moves out of the mitochondria and accumulates in the cytosol, particularly in conditions of low nitrogen and high carbon. In the cytosol, the citrate can be converted to oxaloacetate in a reaction which releases acetyl-CoA that, as in the embodiments depicted in Figs. 5 and 6, can be directed into cytosolic fatty acid synthesis. Thus, this platform system can also include a form of an acetyl group carbon recycling loop that can capture acetyl moieties in the cytosol after they have participated in the TCA cycle in mitochondria instead of before they enter mitochondria as acetyl-carnitine. The carbon recycle loop is depicted in Fig. 7 by the dark, solid reaction arrows beginning with extracellular fatty acid

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1566 of 3077

internalization in the upper left corner of the figure. Starting with carbon internalization, the initial segments of the loop (i.e., carbon processing in ω-oxidation and β-oxidation yielding acetyl-CoA that moves out of the peroxisome and into the cytosol as acetyl-carnitine) are the same as those of the recycling loop shown in Fig. 5 until the point of the processing of cytosolic acetyl-carnitine. In the platform system shown in Fig. 7, acetyl group carbon can continue to flow unimpeded into mitochondria, typically in the form of acetyl-carnitine, and can be converted to acetyl-CoA, utilized in the TCA cycle, and then can be regenerated as acetyl-CoA in the cytosol from a TCA cycle intermediate, citrate ("Cit" in Fig. 7). From that point on, the processing of carbon in the recycling loop is typically similar to that of cytosolic acetyl-CoA in the systems shown in Figs. 5 and 6. The acetyl-CoA can enter the final segments of the loop with the acetyl carbons being utilized in fatty acid synthesis to generate acyl-CoA which is then hydrolyzed to free fatty acid for re-entry into another cycle of the loop at the starting point of ω-oxidation.

Many of the exemplary host cell or organism modifications in the platform system shown in Fig. 7 are the same as those in the platform system shown in Fig. 5, e.g., decreased or eliminated cytosolic acyl-CoA synthetase activity, increased amount and/or activity of cytosolic acetyl-CoA carboxylase and cytosolic thioesterase. Additionally, modifications to enhance ω- and β-oxidation pathway activities and peroxisome proliferation, as well as to reduce transport of acyl-CoA into peroxisomes, can optionally be included in the engineering of a platform system shown in Fig. 7 in a host cell or organism. If a host cell or organism being modified to create a system shown in Fig. 7 does not express an endogenous cytosolic ATP citrate lyase activity (shown as "ACL1/2" in Fig. 7) to catalyze the conversion of citrate to oxaloacetate thereby releasing acetyl-CoA, it can be geneticially modified to generate the enzyme activity. For example, one or more copies of a heterologous nucleic acid encoding an ATP citrate lyase operably linked to a suitable promoter element (e.g., a strong and/or fatty acid-inducible promoter such as a yeast *HDE* gene promoter) can be introduced into the cell or organism for recombinant expression of the activity. If a host cell or organism (for example, an oleaginous yeast) does express an endogenous cytosolic ATP citrate lyase activity, the activity can be increased using genetic engineering methods as described herein (e.g., increasing the copy number and/or transcription of nucleic acid encoding an ATP citrate lyase, increasing the activity of the enzyme by introducing nucleic acid encoding a modified amino acid sequence of the enzyme into host cells).

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1567 of 3077

Capturing a portion of the acetyl-carbon that is utilized in mitochondrial metabolism in the form of citrate that moves out of the mitochondria and accumulates in the cytosol is a modification that can also be incorporated into the platform systems shown in Figs. 5 and 6. For example, one or more copies of a heterologous nucleic acid encoding an ATP citrate lyase operably linked to a suitable promoter element can be introduced into the cell or organism modified as shown in Fig. 5 or Fig. 6 for expression of ATP citrate lyase activity in the cytosol and conversion of cytosolic citrate to oxaloacetate with concommitant release of acetyl-CoA. The released acetyl-CoA can then be utilized in the cytosolic fatty acid synthesis segment of the carbon recycle loop along with acetyl-CoA generated from acetate (as shown in Fig. 6) or from acetyl-carnitine (as shown in Fig. 5).

Engineered pathways for producing fatty acids

An example of a target molecule that can be produced using the modified cells or organisms and methods provided herein is a fatty acid. Examples of enhanced fatty acid-producing cell- or microbial-based systems provided herein include the systems depicted in Figs. 5, 6 and 7. The fatty acid target molecule generation pathways in these examples center on a modified oxidative metabolism (ω- and β-oxidation) pathway through which a longer-chain hydrocarbon feedstock carbon source (e.g., fatty acids and/or alkanes) can be degraded to yield a shorter-chain target fatty acid molecule. In a typical unmodified cell or organism, the β-oxidation cycle releases two carbon atoms in the form of acetyl-CoA generated during chain shortening which, in eukaryotic cells, can then be used in other metabolic processes, including, for example, the TCA cycle after moving from the peroxisome to the mitochondria. In engineered cell- or organism-based production systems such as shown in Figs. 5, 6 and 7, acetyl group carbons released during β-oxidation can be captured in the cytoplasm (e.g., as acetyl-carnitine, acetate and/or citrate) and converted back to acetyl-CoA so that they can be used for the synthesis of longer-chain fatty acids in the cytoplasm and eventually be incorporated into shorter-chain fatty acid target molecule product through, for example, modified oxidative metabolism. These high-efficiency systems can enhance fatty acid target molecule production by including an engineered carbon recycling loop that can increase the efficiency of fatty acid production and can provide greater target molecule yields by decreasing loss of feedstock carbons to cellular energy-generating and/or growth processes. This can be accomplished through rerouting carbon that would have been lost as acetyl-CoA during β-oxidation degradation back into the target molecule fatty acid product.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1568 of 3077

For example, one target fatty acid molecule that can be produced using cells or organisms modified to include a carbon-capturing recycling loop such as those shown in Figs. 5, 6 and 7, is adipic acid, which is a 6-carbon dicarboxylic acid. A cell or microorganism that has been modified as shown in any of Figs. 5, 6 or 7, when provided with a fatty acid carbon source (e.g., a typical 18-carbon vegetable-oil fatty acid), can metabolize the fatty acid first through ω-oxidation, in which it can be converted to a diacid containing the same number of carbons as the fatty acid carbon source, e.g., 18 carbons, and can then metabolize the diacid through β-oxidation. The products of the first cycle of β-oxidation typically are a 16-carbon diacid and a 2-carbon acetyl-CoA molecule. The products of a second cycle of β-oxidation starting with the 16-carbon diacid are typically a 14-carbon diacid and another 2-carbon acetyl-CoA molecule. After four more cycles of β-oxidation, the 6-carbon adipic acid target molecule can be produced, in which case 6 molecules of acetyl-CoA (a total of 12 carbon atoms) will have been released into the peroxisome. In a cell- or microbial-based production system that does not include an engineered carbon recycling pathway such as those shown in Figs. 5, 6 and 7, the 12 carbon atoms released during β-oxidation are usually not used in generating 2 more 6-carbon adipic acid target molecule products but would be "lost" to cellular metabolism pathways involved in energy generation and growth. Thus, only one-third of the source carbons would be used in product generation whereas two-thirds of the source carbons would be lost to processes that do not require all of the lost carbon in order for the cell or microbe to survive. However, cells and microbes modified as described herein to alter cellular carbon flux can capture more of the source carbon atoms and use them in generating more product. Accordingly, the modified cells and microorganisms provided herein as target molecule production systems can be significantly more efficient and can provide a greater product yield from a given amount of feedstock than cell- or microbial-based systems that have not been modified for enhanced production.

Engineered malonyl-CoA-producing pathways in a platform target molecule production system

Figure 8 depicts possible cellular modifications in some embodiments of a eukaryotic (e.g., yeast in this example) platform system for the enhanced production of malonyl-CoA. Malonyl-CoA is a versatile precursor molecule in the synthesis of many industrially valuable molecules. Accordingly, a system such as that depicted in Fig. 8 can be used as a platform for incorporating pathways for target molecule production. Although multiple, possible, cellular

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1569 of 3077

modifications are illustrated in Fig. 8, as described herein, some of the modifications depicted in the figure are optional enhancements of an exemplary engineered system and may or may not be included in a modified cell or organism depending on, for example, the intended use of the system (e.g., development of a particular single, or multiple, target molecule(s) production system) and the selection of variable features (e.g., host cell or organism, carbon source, regulatory controls (such as transcription control elements), culture conditions and the like) of the system. Thus, it is understood that any optional modifications set forth in the exemplary system shown in Fig. 8 are non-limiting and may or may not be included in a particular engineered system and, if included, may be in utilized in different combinations than illustrated in the figure.

The cell-based platform system shown in Fig. 8 incorporates carbon flux modifications designed to capture carbon atoms as acetyl-CoA, the precursor to malonyl-CoA, in the cytosol by enhancing carbon flow through cellular oxidative metabolism pathways (ω-oxidation and peroxisomal β-oxidation) and reducing flow of carbon into mitochondria. Thus, the amount of carbon that is lost to the TCA cycle at the expense of malonyl-CoA production can be reduced in this platform system. At the core of this platform system are acetyl group diversion elements that impede the flow of acetyl group carbons into mitochondria and enhance flow of the carbons into generation of malonyl-CoA. These elements can optionally be combined with a carbon recycle loop that can capture any cytosolic acetyl moieties that are not funneled into the target molecule production pathway. Such cytosolic acetyl moieties can be captured in the fatty acid biosynthesis segment of the carbon recycle loop and can be used to regenerate an acyl-CoA that can be diverted from lipid synthesis and converted into a fatty acid that can be subjected to another cycle of peroxisomal β-oxidation (depicted as a dashed line extending from malonyl-CoA to FA-CoA in Fig. 8).

As set forth in connection with the description of the platform system shown in Fig. 5, carbon flux alteration in the system shown in Fig. 8 can begin with modifications, e.g., decreased acyl-CoA synthetase (Faa1 and/or Fat1) activity, that enhance carbon flow usually in the form of a fatty acid (from, for example, a fatty acid carbon source or derived from an alkane source) into ω- and β-oxidation pathways in the cell or organism. The oxidative metabolism aspect of the platform system can be further enhanced through optional modification of ω- and/or β-oxidation activity as described herein. For example, with respect to the ω-oxidation pathway, one or more enzymes (e.g., monooxygenase, cytochrome P450 reductase, such as CPRB, and others) of

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1570 of 3077

the pathway can be modified (e.g., as described herein) to increase catalytic activity and/or alter substrate specificity in order to increase fatty acid processing in the pathway and/or to target specific fatty acids for processing into dicarboxylic acids. If a host cell or organism does not express an endogenous ω-oxidation pathway, it can be genetically modified to express heterologous enzymes to engineer an ω-oxidation pathway in the cell or organism. Optional enhancements of the β-oxidation activity can include, for example, modification of one or more enzymes (e.g., acyl-CoA oxidase, ketoacyl-CoA thiolase, multifunctional enzyme hydratase and/or dehydrogenase, and others) of the pathway (e.g., as described herein) to increase catalytic activity and/or alter substrate specificity in order to increase fatty acid processing in the pathway and/or to target specific fatty diacids for processing into target dicarboxylic acids (in an instance in which the system shown in Fig. 8 could be used as a multiple target molecule production system, i.e., a "co-production" system). One example of a modification of β-oxidation activity, as described herein, is alteration of the substrate specificity of one or more acyl-CoA oxidase enzymes in the pathway, such as Pox4 and/or Pox5 of *Candida* yeast strains. Additional modifications that can provide for enhanced carbon flux through β-oxidation (and enhanced target molecule production) in these systems include, but are not limited to, modification of β-oxidation-associated activities, such as peroxisome biogenesis and proliferation activities. For example, as described herein, the abundance and/or volume of peroxisomes in which β-oxidation occurs can be increased in host cells through genetic modification. An example of such a modification is increasing the transcription of, and/or number of copies of, one or more peroxin-encoding nucleic acids (e.g., *PEX11*) in a host cell. Amplification of such peroxin-encoding nucleic acids and/or activities leads to an overall increased β-oxidation capacity.

Carbon capture modifications of this platform system, as in the system depicted in Fig. 5, occurring in the processing of the acetyl-CoA generated during β-oxidation can include diversion elements that can impede the flow of acetyl group carbons into mitochondria (e.g., decreasing mitochondrial carnitine transporters and carnitine acetyltransferase activities) and enhance flow of the carbons into generation of malonyl-CoA (e.g., increasing cytosolic carnitine acetyltransferase and acetyl-CoA carboxylase activities). Modifications in the host cell or organism that can enhance the capture and diversion of acetyl-CoA in the cytosol include, but are not limited to, modification of acetyl-carnitine entry into mitochondria and modification of conversion of cytosolic acetyl-carnitine to acetyl-CoA. As described herein, one modification that can be made to enhance capture of acetyl groups in the cytosol is to increase the amount

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1571 of 3077

and/or activity of cytosolic carnitine acetyltransferase which converts acetyl-carnitine into acetyl-CoA, a thioester form that cannot move into the mitochondrial matrix. Methods of achieving this include increasing the copy number of nucleic acids encoding cytosolic carnitine acetyltransferase in the cell, increasing the transcription of such nucleic acids and/or introducing nucleic acid encoding a more active cytosolic carnitine acetyltransferase enzyme into the cell (e.g., modifying an endogenous cytosolic enzyme activity by replacing it with, or adding to it, a heterologous enzyme activity; see, e.g., "*CAT2*^{cyt}" in Fig. 8). Another modification that can enhance cytosolic capture and diversion of acetyl moieties in cells is an alteration of acetyl-carnitine uptake into mitochondria from the cytosol (depicted as faded, dotted reaction arrow lines into and in the mitochondrial compartment in Fig. 8). One method of modifying mitochondrial acetyl-carnitine uptake can be altering the processing of acetyl-carnitine that occurs in the mitochondria to convert it to acetyl-CoA for use in the TCA cycle. For example, by decreasing the activity level of an enzyme that catalyzes this processing, e.g., mitochondrial carnitine acetyltransferase, there can be a corresponding decrease in conversion of acetyl-carnitine to acetyl-CoA in the mitochondria. Without being limited or bound by theory, if the mitochondrial carnitine acetyltransferase activity is not sufficient to efficiently process the acetyl carbon flux coming from the peroxisome, then the cytoplasmic acetyl-carnitine concentration should build up and, in effect, acetyl-carnitine can be diverted from the TCA cycle. The increased concentration of cytoplasmic acetyl-carnitine thus can be a source of substrate for carnitine acetyltransferase activity in the cytosol which can convert the substrate to cytosolic acetyl-CoA for use in target molecule production. The amount and/or activity of mitochondrial carnitine acetyltransferase can be decreased in a number of ways, as described herein. For example, the number of copies of nucleic acid encoding the enzyme in a host cell can be reduced, the transcription of such nucleic acid can be decreased and/or nucleic acid encoding a less active mitochondrial carnitine acetyltransferase enzyme can be introduced into the cell (e.g., replacing the endogenous mitochondrial enzyme with a heterologous enzyme; see, e.g., "*CAT2*" in a diagonal line-hatched background in the mitochondria in Fig. 8).

As also described with reference to Fig. 5, another method of modifying mitochondrial acetyl-carnitine uptake can be altering the transport mechanism that moves acetyl-carnitine into the mitochondrial matrix, e.g., an acetyl-carnitine translocase. By decreasing the amount and/or activity level of the transport protein, movement of acetyl-carnitine from the cytosol into mitochondria can be slowed and/or reduced thereby increasing the concentration of acetyl-carnitine in the cytosol that can be converted to acetyl-CoA by cytoplasmic carnitine

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1572 of 3077

acetyltransferase. The amount and/or activity of mitochondrial acetyl-carnitine transport protein can be decreased, for example, by reducing the number of copies of nucleic acid encoding a mitochondrial acetyl-carnitine transport protein in a host cell, reducing the transcription of such nucleic acid and/or introducing nucleic acid encoding a less active transport protein into the cell (e.g., replacing the endogenous mitochondrial transport protein with a heterologous protein). For example, in one embodiment described herein, the transcription of nucleic acid encoding a *Candida* acetyl-carnitine translocase (shown as "CRC1" in a diagonal line-hatched background in Fig. 8) can be reduced in a host cell by introducing such nucleic acid, which is operably linked to a heterologous promoter that provides for less transcription and/or a reduced transcription rate, and/or that can be regulated to provide for alternately weak and stronger transcription, into host cells in which the endogenous gene has been disrupted. Thus, modification of mitochondrial acetyl-carnitine transporter expression and/or mitochondrial carnitine acetyltransferase activity, as shown in Fig. 8, can serve to divert acetyl-carnitine from use in the TCA cycle and increase the concentration of acetyl-carnitine in the cytosol of cells that include one or more these modifications. This, combined with increased amounts and/or activity of cytosolic carnitine acetyltransferase, can result in increased amounts of acetyl-CoA available in the cytosol for use in target molecule production. As also described with reference to Fig. 5, although all three of these modifications, i.e., alteration of mitochondrial carnitine acetyltransferase activity (e.g., Cat2), alteration of mitochondrial acetyl-carnitine transporter protein activity (e.g., Crc1) and modification of cytosolic carnitine acetyltransferase activity (e.g., Yat1), are shown in the depiction of a platform system in Fig. 8, each can be used singly or in any combination in modifying carbon processing in cells or microorganisms for the production of malonyl-CoA and/or other target molecules.

At this point in the platform pathway system depicted in Fig. 8, when acetyl moieties are accumulated in the cytosol as acetyl-CoA, they can be directed into generation of malonyl-CoA. To enhance flow of the accumulated cytosolic acetyl-CoA toward malonyl-CoA generation, the process of converting acetyl-CoA into malonyl-CoA can be modified in host cells or microorganisms. For example, the amount and/or activity of an enzyme that catalyzes the reaction, acetyl-CoA carboxylase, can be modified, e.g., increased, in host cells or organisms using genetic engineering methods as described herein (e.g., increasing the copy number and/or transcription of nucleic acid encoding acetyl-CoA carboxylase, increasing the activity of the enzyme by introducing nucleic acid encoding a modified amino acid sequence of the enzyme into host cells). In one embodiment described herein, the activity of a yeast (e.g.,

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1573 of 3077

Candida) cytosolic acetyl-CoA carboxylase (shown as "*ACC1*" in a diagonal line-hatched background oval in Fig. 8) can be increased through substitution of phosphorylatable serine residues with alternate (e.g., alanine) residues to reduce inhibition of the enzyme by phosphorylation. Heterologous nucleic acid encoding the modified protein can be introduced into a host cell for expression of the enzyme therein. Increasing the amount and/or activity of acetyl-CoA carboxylase can reduce or prevent any limitation on flow of accumulated acetyl-CoA into malonyl-CoA generation.

The system depicted in Fig. 8 serves as a platform that can be used in multiple ways. For example, as shown, the pathway can serve as an enhanced, high-efficiency malonyl-CoA production system that can be further modified for use in generating a target molecule in an engineered pathway that initiates with a malonyl-CoA precursor. Examples of this are provided herein in which cells or organisms modified to contain the platform system are genetically modified to express pathways for the production of 3-hydroxypropionic acid or triacetic acid lactone from malonyl-CoA. Additionally, the system depicted in Fig. 8 can be used in the co-production of a fatty dicarboxlic acid, e.g., adipic acid, and a target molecule generated using malonyl-CoA as a precursor in the synthesis pathway.

Cells or microorganisms that have been modified to incorporate the biological platform system depicted in Fig. 8 can be further modified depending on the purpose(s) for which the system is being used. For example, if the system is being used solely for the production of a target molecule synthesized from a malonyl-CoA precursor, then it may be beneficial to modify the cell or microorganism to decrease cytosolic fatty acid synthesis. A committed step in fatty acid biosynthesis is the conversion of cytoplasmic acetyl-CoA into malonyl-CoA. Malonyl-CoA can serve as a carbon donor in the synthesis of a fatty acid chain in repeated cycles of the addition of 2 carbon atoms per cycle to extend the chain and generate a fatty acid. The fatty acid synthesis (FAS) pathway in host cells or organisms can compete with any pathways for target molecule synthesis for the malonyl-CoA substrate produced in this platform system. Therefore, to enhance malonyl-CoA carbon atom flux toward target molecule synthesis and direct carbon flux away from fatty acid biosynthesis, a host cell or organism can be modified to alter FAS enzyme activities. Non-limiting examples of methods suitable to decrease the amount and/or activity of enzymes involved in fatty acid synthesis, e.g., FAS2 and FAS1 enzymes, include decreasing the number of nucleic acids encoding fatty acid synthase enzymes, such as FAS2 and/or FAS1, in a host cell (e.g., disruption or deletion of one or more genes) and reducing the

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1574 of 3077

transcription of nucleic acids encoding fatty acid synthase enzymes, e.g., FAS2 and/or FAS1 genes, by replacing an endogenous promoter of one or more genes with a weak and/or constitutive promoter, the like or combinations thereof. Examples of polynucleotides from *Candida* strain ATCC 20336 that encode fatty acid synthase molecules (FAS1, FAS2) are provided herein and are also described in International patent application no. PCT/US2012/056562 (publication no. WO 2012/056562).

If the platform system depicted in Fig. 8 is being used for co-production of a fatty dicarboxylic acid and a target molecule derived from malonyl-CoA, then it may be beneficial to either make no modifications of the FAS pathway activities or modify the host cell or organism to increase the amount, activity, and/or alter the specificity, of enzymes in the fatty acid synthesis (FAS) enzyme complex, shown as "FAS" in Fig. 8 (e.g., the enzyme activities of the FAS1 and FAS2 subunits of yeast). Examples of methods of altering fatty acid synthesis through modification of the amount and/or activity of fatty acid synthase enzymes (e.g., increasing cellular copy number of nucleic acids encoding one or more enzymes, increasing transcription of nucleic acids encoding one or more enzymes) are described herein. A system such as that shown in Fig. 8 being used for co-production of a fatty dicarboxylic acid and a target molecule derived from malonyl-CoA, could include modifications, such as those described herein, of one or more activities of the ω- and/or β-oxidation pathways to enhance production of a desired dicarboxylic acid as described herein (e.g., modification of acyl-CoA oxidases, such as Pox4 and Pox5, monooxygenase, cytochrome P450 reductase). Such a co-production system would be depicted in Fig. 8 as including another arrow extending from a dicarboxylic acid (DCA-CoA) in the peroxisome to an acyl-CoA and then to a chain-shortened diacid (as a result of β-oxidation, "β-Ox") as is shown in the systems depicted in Figs. 5 and 6.

Additionally, the efficiency of the system for co-production of a dicarboxylic acid may optionally be enhanced by modifying the host cell or organism to introduce the final segment of the carbon recycling loop as described with reference to Fig. 5 (and shown at the top of Fig. 8 as a solid reaction arrow extending from "FA-CoA" to "FA" and including the "TES^{cyt}" thioesterase enzyme as catalyst). In platform systems depicted in Fig. 8 that include a functional FAS pathway, the final product of the cytosolic fatty acid synthesis is typically an acyl-CoA (e.g., palmitoyl-CoA), which is shown as "FA-CoA" in Fig. 8. To avoid loss of the carbon atoms in the acyl-CoA lipid synthesis pathways, and complete the carbon recycling loop, the host cell or organism can be modified to include an engineered cytosolic thioesterase enzyme to, in effect, "deactivate" the

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1575 of 3077

fatty acid-CoA through hydrolysis and removal of coenzyme A. This can divert the carbons in the acyl-CoA from use in cellular processes not involved in target molecule production and/or generation of cytosolic acetyl-CoA and can complete the recycling loop by generating a cytosolic free fatty acid that can then begin a new cycle of the loop pathway at the initial point of ω-oxidation. A host cell can be modified to increase (or introduce) a thioesterase activity in the cytosol in order to direct acyl-CoA carbon flux toward oxidative metabolism pathways (ω- and β-oxidation). This can be achieved using genetic engineering methods as described herein (e.g., increasing the copy number and/or transcription of nucleic acid encoding a thioesterase, increasing the activity of the enzyme by introducing nucleic acid encoding a modified amino acid sequence of the enzyme into host cells). For example, in one embodiment described herein, a *Candida* peroxisomal thioesterase (e.g, Tes) can be recombinantly expressed cytosolically in a host cell (shown as $TES3^{cyt}$ in a diagonal line-hatched background oval in Fig. 8).

Another modification that can optionally be included in, but that is not required for, platform systems, such as those depicted in Fig. 8, is a decrease in the amount and/or activity of, or elimination of, mechanisms for the transport of acyl-CoA across the peroxisomal membrane and into peroxisomes. This modification can be beneficial in embodiments in which one of the target molecules (or precursor or intermediate in the production of a target molecule) is a dicarboxylic acid. This is because if an acyl-CoA generated from fatty acid synthesis can move across the peroxisomal membrane and into peroxisomes, it could directly enter into β-oxidation without first being converted to a diacid through ω-oxidation. A reduction in the transfer of acyl-CoA into peroxisomes can be achieved through disruption or deletion of genes encoding peroxisomal acyl-CoA transporters. In one embodiment described herein, a yeast (e.g., *Candida*) peroxisomal transport protein (e.g, Pxa1) activity can be decreased or eliminated by disrupting the gene encoding the protein in the host cell (e.g., shown as “pxa1Δ” in a black background oval in Fig. 8).

Engineered pathways for producing 3-hydroxypropionic acid

Modified cells, organisms, compositions and methods provided herein can also be used for enhanced production of other organic acids, such as, for example, 3-hydroxypropionic acid (3HP). An example of an engineered production pathway for cell- or microbial-based synthesis of 3HP is shown in Fig. 9. This biochemical pathway is based in a synthetic method using malonyl-CoA as a precursor molecule and incorporates elements as described for the platform

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1576 of 3077

system depicted in Fig. 8. As shown in Fig. 9, 3HP can be produced from malonyl-CoA in a 2-step reduction process. In the first step, malonyl-CoA ("Mal-CoA" in Fig. 9) can be reduced to malonate semialdehyde (MSA) in a reaction catalyzed by malonyl-CoA reductase (e.g., EC 1.2.1.75 ; "MCR" in Fig. 9). Malonate semialdehyde can be further reduced to 3HP in a reaction catalyzed by 3-hydroxy-propionate-dehydrogenase (e.g., EC 1.1.1.59; "HPD1" in Fig. 9). If a host cell or organism being modified to create a system shown in Fig. 9 does not express an endogenous cytosolic malonyl-CoA reductase or 3-hydroxy-propionate-dehydrogenase activity, it can be geneticially modified to generate one or both of the enzyme activities. For example, one or more copies of a heterologous nucleic acid encoding an Mcr (e.g., nucleotide SEQ ID NO: 323 or any nucleotide sequence encoding amino acid SEQ ID NO: 322) or Hpd1 (e.g., nucleotide SEQ ID NO: 104 or any nucleotide sequence encoding amino acid SEQ ID NO: 48) operably linked to a suitable promoter element (e.g., a strong and/or fatty acid-inducible promoter such as a yeast *HDE* gene promoter) can be introduced into the cell or organism for recombinant expression of the activity. If a host cell or organism does express an endogenous cytosolic malonyl-CoA reductase and/or 3-hydroxy-propionate-dehydrogenase activity, the activity may be increased using genetic engineering methods as described herein (e.g., increasing the copy number and/or transcription of nucleic acid encoding an Mcr or Hpd1, increasing the activity of the enzyme by introducing nucleic acid encoding a modified amino acid sequence of the enzyme into host cells). Nucleic acid encoding an MCR activity can be obtained, for example, from bacteria, including, but not limited to, *Sulfolobus islandicus* (see, e.g, nucleic acid SEQ ID NO: 323 and encoded amino acid SEQ ID NO:322) *Sulfolobus tokodaii* (nucleotide sequence: EMBL-EBI accession no. BAB67276.2) strain 7 (DSMZ 16693; available from the Leibniz Institute DSMZ-German Collection of Microorganisms and cell lines). An example of a *Candida viswanathii* nucleotide sequence (SEQ ID NO:104) encoding an Hpd1 protein (amino acid SEQ ID NO:48) is provided herein. Nucleic acid encoding Hpd1 can be obtained from additional sources, for example, yeast strains such as *Candida albicans* (nucleotide sequence: Genbank accession no. XM_714034), e.g., strain SC5314 (ATCC No. MYA-2876). Methods for detecting the presence and/or activity of a malonyl Co-A reductase include spectrophotometric assays such as described by Alber et al. ((2006) *J. Bacteriol.* 188:8551-8559). Methods for detecting the presence and/or activity of a malonyl 3-hydroxy-propionate-dehydrogenase include assays such as described by Otzen et al. ((2014) *J. Biol. Chem.* 289(12):8151-8169).

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1577 of 3077

Some cells and microbial hosts, e.g., yeast spp., may express an endogenous semialdehyde dehydrogenase (e.g., EC 1.2.1.18; Ald6p) enzyme that catalyzes the oxidation of malonate semialdehyde to acetyl-CoA and CO_2 which would result in loss of the MSA intermediate in the synthesis of 3HP. Therefore, a gene encoding a semialdehyde dehydrogenase in any such host cell or organism can be disrupted or deleted, for example, using gene disruption techniques known in the art and/or described herein, to decrease or eliminate the enzyme activity. An example of a *Candida viswanathii* nucleotide sequence (SEQ ID NO:105) encoding an Ald6 protein (amino acid SEQ ID NO:49) is provided herein. Additional nucleotide sequences encoding Ald6 include, but are not limited to, *Candida albicans* strain SC5314 *ALD6* (Genbank accession no. XM-705897), *Saccharomyces cerevisiae ALD6* (Genbank accession no. NM_001183875). Methods for detecting the presence and/or activity of a semialdehyde dehydrogenase include assays such as described by Otzen et al. ((2014) *J. Biol. Chem.* 289(12):8151-8169) and Banerjee et al. ((1970) *J. Biol. Chem.* 245:1828-1835).

As shown in Fig. 9, a cellular platform pathway enhanced for production of malonyl-CoA, such as that depicted in Fig. 8, can provide a highly compatible system for use in the efficient production of 3HP. The platform is designed to enable maximal use of a lower cost carbon source (e.g., fatty acid or alkane) through modifications of cellular carbon flux that can enhance flow of source carbons to the generation of the malonyl-CoA precursor in the reaction scheme for production of 3HP. The presence and/or amount of 3HP in a sample can be determined, for example, using analytical methods such as HPLC (see, e.g., Raj et al. (2008) *Process Biochem.* 43:1440-1446 and International patent application no. PCT/US2016/023243 (publication no. WO 2016/154046)).

Cells or organisms that have been modified to incorporate a 3HP production system such as that depicted in Fig. 9 can be further modified depending on the purpose(s) for which the system is being used. For example, if the system is being used solely for the production of 3HP, then it may be beneficial to modify the cell or organism to decrease cytosolic fatty acid synthesis (e.g., as described with respect to the sytem shown in Fig. 8). If the system depicted in Fig. 9 is being used for co-production of a fatty dicarboxylic acid and 3HP, then it may be beneficial to either make no modifications of the FAS pathway activities or modify the host cell or organism to increase the amount, activity, and/or alter the specificity, of enzymes in the fatty acid synthesis (FAS) enzyme complex, shown as “FAS” in Fig. 9 (and as described, for example, with respect to the system shown in Fig. 8). A system such as that shown in Fig. 9 being used for co-

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1578 of 3077

production of a fatty dicarboxylic acid and 3HP could also optionally include modifications, such as those described herein, of one or more activities of the ω- and/or β-oxidation pathways to enhance production of a desired dicarboxylic acid as described herein (e.g., modification of acyl-CoA oxidases, such as Pox4 and Pox5, monooxygenase, cytochrome P450 reductase). Such a co-production system would be depicted in Fig. 9 as including another arrow extending from a dicarboxylic acid (DCA-CoA) in the peroxisome to an acyl-CoA and then to a chain-shortened diacid (as a result of β-oxidation, "β-Ox") as is shown in the systems depicted in Figs. 5 and 6.

Although multiple, possible, cellular modifications are illustrated in Fig. 9, as described herein, some of the modifications depicted in the figure are optional enhancements of an exemplary engineered system and may or may not be included in a modified cell or organism depending on, for example, the intended use of the system (e.g., development of a particular single, or multiple, target molecule(s) production system) and the selection of variable features (e.g., host cell or organism, carbon source, regulatory controls (such as transcription control elements), culture conditions and the like) of the system. Thus, it is understood that any optional modifications set forth in the exemplary system shown in Fig. 9 are non-limiting and may or may not be included in a particular engineered system and, if included, may be in utilized in different combinations than illustrated in the figure.

Engineered pathways for producing polyketides

Modified cells, organisms, compositions and methods provided herein can also be used for enhanced production of polyketides, such as, for example, triacetic acid lactone (TAL). An example of an engineered production pathway for cell- or microbial-based synthesis of TAL is shown in Fig. 10. This biochemical pathway is based in a synthetic method using malonyl-CoA as a precursor molecule. As shown in Fig. 10, TAL can be produced from malonyl-CoA in two condensation reactions with acetyl-CoA catalyzed by 2-pyrone synthase (EC 2.3.1; "2PS" in Fig. 10). A malonyl-CoA-producing cell or organism can be genetically modified to express a 2-pyrone synthase enzymatic activity by introducing heterologous nucleic acid encoding the enzyme into a host cell or organism. For example, one or more copies of a heterologous nucleic acid encoding 2PS operably linked to a suitable promoter element (e.g., a strong and/or fatty acid-inducible promoter such as a yeast *HDE* gene promoter) can be introduced into the cell or organism for recombinant expression of the activity. Nucleic acid encoding a 2-pyrone

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1579 of 3077

synthase can be obtained, for example, from plant species such as *Gerbera hybrida* (e.g., nucleotide SEQ ID NO: 325 encoding amino acid SEQ ID NO: 324, and Genbank nucleotide sequence accession no. Z38097). The enzymatic activity of 2-pyrone synthase can be determined using a TLC-based radiometric assay as described by Jez et al. ((2000) *Chemistry and Biology* 7(12):919-930).

As shown in Fig. 10, a cellular platform pathway enhanced for production of malonyl-CoA, such as that depicted in Fig. 8, can provide a highly compatible system for use in the efficient production of TAL. The platform is designed to enable maximal use of a lower cost carbon source (e.g., fatty acid or alkane) through modifications of cellular carbon flux that can enhance flow of source carbons to the generation of the malonyl-CoA precursor in the reaction scheme for production of TAL. TAL concentrations can be determined using reversed-phase HPLC as described by Xie et al. ((2006) *Biotechnol Bioengineering* 93(4):727-736). Cells or organisms that have been modified to incorporate the TAL production system depicted in Fig. 10 can be further modified depending on the purpose(s) for which the system is being used. For example, if the system is being used solely for the production of TAL, then it may be beneficial to modify the cell or organism to decrease cytosolic fatty acid synthesis (e.g., as described with respect to the sytem shown in Fig. 8). If the system depicted in Fig. 10 is being used for co-production of a fatty dicarboxylic acid and TAL, then it may be beneficial to either make no modifications of the FAS pathway activities or modify the host cell or organism to increase the amount, activity, and/or alter the specificity, of enzymes in the fatty acid synthesis (FAS) enzyme complex, shown as “FAS” in Fig. 10 (and as described, for example, with respect to the system shown in Fig. 8). A system such as that shown in Fig. 10 being used for co-production of a fatty dicarboxylic acid and TAL could also optionally include modifications, such as those described herein, of one or more activities of the ω- and/or β-oxidation pathways to enhance production of a desired dicarboxylic acid as described herein (e.g., modification of acyl-CoA oxidases, such as Pox4 and Pox5, monooxygenase, cytochrome P450 reductase). Such a co-production system would be depicted in Fig. 10 as including another arrow extending from a dicarboxylic acid (DCA-CoA) in the peroxisome to an acyl-CoA and then to a chain-shortened diacid (as a result of β-oxidation, “β-Ox”) as is shown in the systems depicted in Figs. 5 and 6.

Although multiple, possible, cellular modifications are illustrated in Fig. 10, as described herein, some of the modifications depicted in the figure are optional enhancements of an exemplary engineered system and may or may not be included in a modified cell or organism depending

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1580 of 3077

on, for example, the intended use of the system (e.g., development of a particular single, or multiple, target molecule(s) production system) and the selection of variable features (e.g., host cell or organism, carbon source, regulatory controls (such as transcription control elements), culture conditions and the like) of the system. Thus, it is understood that any optional modifications set forth in the exemplary system shown in Fig. 10 are non-limiting and may or may not be included in a particular engineered system and, if included, may be in utilized in different combinations than illustrated in the figure.

Engineered acetyl-CoA-generating pathways in a platform target molecule production system

Figure 11 depicts possible cellular modifications in some embodiments of a eukaryotic (i.e., yeast in this example) platform system for the enhanced generation of acetyl-CoA. Acetyl-CoA can be a versatile precursor molecule in the synthesis of many industrially valuable molecules. Accordingly, a system such as that depicted in Fig. 11 can be used as a platform for incorporating particular pathways for the production of a diverse array of target molecules. The cell-based platform system shown in Fig. 11 can incorporate carbon-flux modifications shown in Fig. 8 that are designed to capture carbon atoms as acetyl-CoA in the cytosol by enhancing carbon flow through cellular oxidative metabolism pathways (ω-oxidation and peroxisomal β-oxidation) and reducing flow of carbon into mitochondria. Fig. 11 differs from Fig. 8 in that it shows an embodiment of the platform system in which target molecule production pathways extend from acetyl-CoA, instead of malonyl-CoA, as a precursor molecule. The core of this platform system, as it is for the system shown in Fig. 8, centers on acetyl group diversion elements that can impede the flow of acetyl group carbons into mitochondria and can enhance flow of the carbons into generation of acetyl-CoA. In systems designed for production of target molecules using acetyl-CoA as a precursor, these elements can optionally be combined with a carbon recycle loop that captures any cytosolic acetyl moieties that are not funneled into the target molecule production pathway into cytosolic fatty acid synthesis to regenerate an acyl-CoA that can be diverted from lipid synthesis and converted into a fatty acid that can be subjected to another cycle of peroxisomal β-oxidation.

Figure 11 shows an embodiment of a platform system for the enhanced generation of acetyl-CoA in connection with terpene target molecule production as an example of the use of acetyl-CoA as a precursor molecule. As shown in the figure, terpenes can be generated, for example,

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1581 of 3077

from isopentenyl diphosphate produced through mevalonate pathways in cells. Through a series of reactions in the mevalonate pathway present in eukaryotes and some bacteria, acetyl-CoA can be converted to isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) which are precursors in the production of polyisoprenoids (e.g., dolichol) and sterols (e.g., ergosterol, cholesterol). Terpenes, e.g., valencene, lycopene, carotenes and the like, can be produced from IPP through further reactions involving enzymes such as, for example, valencene synthase, and enzymes encoded by carotenoid biosynthesis *CRT* genes (e.g., geranylgeranyl pyrophosphate synthase (CrtE), phytoene synthase (CrtB), phytoene desaturase (CrtI), beta-carotene ketolase (CrtW), beta-carotene hydroxylase (CrtZ)). Cells and organisms engineered for production of terpenes as shown in Fig. 11 can be modified to enhance flow of carbon through the terpene production pathways. For example, such engineered cells can be genetically modified to express, or increase expression of, any of the mevalonate pathway enzymes and/or enzymes involved in terpene synthesis pathways.

As set forth in connection with the description of the platform systems shown in Figs. 5 and 8, carbon flux alteration in the system shown in Fig. 11 can begin with modifications (e.g., decreased acyl-CoA synthetase (Faa1 and/or Fat1) activity) that can enhance carbon flow in the form of a fatty acid (from, for example, a fatty acid carbon source or derived from an alkane source) into ω- and β-oxidation pathways in the cell or organism. The oxidative metabolism aspect of the platform system can be further enhanced through modification of ω- and/or β-oxidation activity as described herein. Carbon capture modifications of this platform system, as in the system depicted in Figs. 5 and 8, can occur in the processing of the acetyl-CoA generated during β-oxidation and can include diversion elements that can impede the flow of acetyl group carbons into mitochondria (e.g., decreasing mitochondrial carnitine transporters and carnitine acetyltransferase activities) and enhance flow of the carbons into generation of acetyl-CoA (e.g., increasing cytosolic carnitine acetyltransferase activity). At this point in the platform pathway system depicted in Fig. 11, when acetyl moieties are accumulated in the cytosol as acetyl-CoA, there primarily are two pathways in which they can be used: (1) fatty acid synthesis (FAS) through initial conversion into malonyl-CoA and (2) target molecule synthesis pathways that can be engineered into the host cell or organism. Thus, target molecule synthesis can be limited by competition for the acetyl-CoA substrate by the FAS pathway. To minimize this competition, in this platform system, unlike the systems shown in Figs. 5 and 8, the host cell or organism is not modified to increase acetyl-CoA carboxylase activity, which is an enzyme that catalyzes conversion of acetyl-CoA into malonyl-CoA and can commit it to fatty acid synthesis.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1582 of 3077

However, if an acetyl-CoA carboxylase activity is present in the cytosol of this cellular system, and the system includes a functional FAS pathway, any acetyl-CoA molecules that are lost to the fatty acid synthesis pathway can be recaptured for possible use in target molecule synthesis by including an optional carbon recycling loop (such as that described in reference to Fig. 5) in the system. The final product of cytosolic fatty acid synthesis is typically an acyl-CoA (e.g., palmitoyl-CoA), which is shown as "FA-CoA" in Fig. 11. To avoid loss of the carbon atoms in the acyl-CoA lipid synthesis pathways, and complete the carbon recycling loop, the host cell or organism can optionally be modified to include a cytosolic thioesterase enzyme to, in effect, "deactivate" the fatty acid-CoA through hydrolysis and removal of coenzyme A. This can divert the carbons in the acyl-CoA from use in cellular processes not involved in target molecule production and/or generation of cytosolic acetyl-CoA and can complete the recycling loop by generating a cytosolic free fatty acid that can then begin a new cycle of the loop pathway at the initial point of ω-oxidation. As described herein, a host cell can be modified to increase (or introduce) a thioesterase activity (shown as *TES3*^{cyt} in a diagonal line-hatched background oval in Fig. 11) in the cytosol in order to direct acyl-CoA carbon flux toward oxidative metabolism pathways (ω- and β-oxidation).

The system depicted in Fig. 11 can serve as a platform that can be used in multiple ways. For example, as shown, the pathway can serve as an enhanced, high-efficiency cytosolic acetyl-CoA-generating system that can be further modified for use in generating a target molecule (e.g., terpenes) in an engineered pathway that initiates with an acetyl-CoA precursor. Additionally, the system depicted in Fig. 11 can be used in the co-production of a fatty dicarboxylic acid, e.g., adipic acid, and a target molecule generated using acetyl-CoA as a precursor in the synthesis pathway. In embodiments in which the system is used for co-production of a fatty diacid and another target molecule, the host cell or organism can optionally also be modified as described herein to decrease the amount and/or activity of, or eliminate, mechanisms for the transport of acyl-CoA across the peroxisomal membrane and into peroxisomes (e.g., shown as "pxa1Δ" in a black background oval in Fig. 11). This modification avoids the by-passing of ω-oxidation by acyl-CoA generated from fatty acid synthesis which would allow the formation of monocarboxlic acids.

Cells or organisms that have been modified to incorporate a biological platform system such as that depicted in Fig. 11 can be further modified depending on the purpose(s) for which the

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1583 of 3077

system is being used. For example, if the system is being used solely for the production of one or more target molecules synthesized from an acetyl-CoA precursor (i.e., not being used as a fatty acid co-production system), then it may be beneficial to modify the cell or organism to decrease or slow cytosolic fatty acid synthesis. The fatty acid synthesis (FAS) pathway in host cells or organisms can compete with any pathways for target molecule synthesis for the cytosolic acetyl-CoA produced in this platform system. Therefore, to enhance acetyl-CoA carbon atom flux toward target molecule synthesis and direct carbon flux away from fatty acid biosynthesis, a host cell or organism can optionally be modified to decrease or attenuate activity of one or more enzymes involved in fatty acid synthesis. Flux of acetyl-CoA carbon into the mevalonate pathway may also optionally be enhanced by increasing one or more activities of the mevalonate pathway. For example, a mevalonate pathway activity may be enhanced by genetically modifying an enzyme that catalyzes a reaction in the pathway (e.g., increasing the copy number of nucleic acids encoding a mevalonate pathway enzyme in the cell, increasing the transcription of such nucleic acids and/or introducing nucleic acid encoding a more active enzyme into the cell).

If a platform system such as that depicted in Fig. 11 is being used for co-production of a fatty dicarboxylic acid and a target product derived from another pathway using acetyl-CoA precursors (e.g., the mevalonate pathway), then it may be beneficial to either make no modifications of the FAS pathway activities or modify the host cell or organism to increase the amount, activity, and/or alter the specificity, of enzymes in the fatty acid synthesis (FAS) enzyme complex, shown as “FAS” in Fig. 11 (e.g., the enzyme activities of the FAS1 and FAS2 subunits of yeast). A system such as that shown in Fig. 11 being used for co-production of a fatty dicarboxylic acid and other acetyl-CoA-dervied product could also optionally include modifications, such as those described herein, of one or more activities of the ω- and/or β-oxidation pathways to enhance production of a desired dicarboxylic acid as described herein (e.g., modification of acyl-CoA oxidases, such as Pox4 and Pox5, monooxygenase, cytochrome P450 reductase). Such a co-production system would be depicted in Fig. 11 as including another arrow extending from a dicarboxylic acid (DCA-CoA) in the peroxisome to an acyl-CoA and then to a chain-shortened diacid (as a result of β-oxidation, “β-Ox”) as is shown in the systems depicted in Figs 5 and 6.

Although multiple, possible, cellular modifications are illustrated in Fig. 11, as described herein, some of the modifications depicted in the figure are optional enhancements of an exemplary

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1584 of 3077

engineered system and may or may not be included in a modified cell or organism depending on, for example, the intended use of the system (e.g., development of a particular single, or multiple, target molecule(s) production system) and the selection of variable features (e.g., host cell or organism, carbon source, regulatory controls (such as transcription control elements), culture conditions and the like) of the system. Thus, it is understood that any optional modifications set forth in the exemplary system shown in Fig. 11 are non-limiting and may or may not be included in a particular engineered system and, if included, may be in utilized in different combinations than illustrated in the figure.

Methods for producing target molecules

Provided herein are methods for contacting an engineered cell, microorganism or organism with a carbon source (e.g., a feedstock containing fatty acids and/or alkanes) under conditions whereby one or more target molecules is produced. Biological methods provided herein for producing a target molecule can incorporate cells or organisms, such as those provided herein, that have been modified to enhance production efficiency by maximizing use and minimizing the costs of raw starting materials input into the process. The methods are designed to provide flexibility in culture conditions, particularly carbon source utilization, and coordinated regulatory mechanisms to enable efficient production of a variety of carbon-containing target molecules. In some embodiments of the methods, the cell or organism used for target molecule production is one that is able to assimilate a variety of carbon sources, including one or more non-fermentable (e.g., alkanes, fatty acids, alcohols) as well as fermentable carbon sources. In particular embodiments, the cell or organism is one that is able to survive under conditions in which the sole carbon source is a non-fermentable carbon source. The use of such cells and organisms in some embodiments of the production methods can contribute to the cost-effectiveness and environmental soundness of the methods.

In certain embodiments, a cell or organism used in the production methods has been modified to alter cellular carbon flux, including, for example, cells and organisms provided herein. Such carbon flux alterations include those that can increase the amount of the source carbon that is available for use in target molecule generation pathways. For example, one or more activities in one or more metabolic pathways of the cell or organism can be engineered to increase carbon flux through the pathways to produce a desired product. The engineered activities can be chosen such that there is an increased production of metabolic intermediates that can be

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1585 of 3077

utilized in one or more pathways to achieve increased production of a desired product relative to the unmodified host cell or organism. The engineered activities also can be chosen such that there is a decreased activity of enzymes that reduce production of a desired intermediate or end product (e.g., reverse activities).

This cellular carbon flux management can be optimized, for any chosen feedstock used in culturing the cells or organisms, by engineering the appropriate activities in the appropriate pathways. Non-limiting examples are given herein using pure alkanes (e.g., single chain length alkanes, dodecane for example), mixed chain-length alkanes, long-chain alkanes, pure fatty acids (e.g., single chain length fatty acids, capric acid or oleic acid for example) and mixed chain length fatty acids as a carbon source in a feedstock. The process of carbon flux management through engineered pathways can be used to produce a target molecule (e.g., an organic acid, a fatty acid, dicarboxylic acid, polyketide, terpene) at a level and/or rate closer to the calculated maximum theoretical yield for any given feedstock, than does a production process that has not been enhanced or modified using methods described herein. The terms "theoretical yield" or "maximum theoretical yield" as used herein refer to the yield of product of a chemical or biological reaction that can be formed if the reaction went to completion. Theoretical yield is based on the stoichiometry of the reaction and ideal conditions in which starting material is completely consumed, undesired side reactions do not occur, the reverse reaction does not occur, and there are no losses in the work-up procedure.

Growth conditions and fermentation

Large-scale cell- or microbial-based target molecule production is generally conducted by culturing the cells or organisms in a fermentor. The culture conditions can vary depending on the cell or organism and the target molecule being produced. In some embodiments of the production processes provided herein, the modified cell or organism is eukaryotic, such as, for example, a yeast. In particular embodiments, the yeast is one that is able to assimilate fatty acids and/or alkanes. For example, in some emodiments, the yeast is a species of *Candida*, e.g., *C. tropicalis* or C. *viswanathii* (e.g., ATCC 20336, ATCC 20913, ATCC 20962) or *Yarrowia*, e.g., *Y. lipolytica* (e.g., ATCC 20228). In some embodiments, the yeast is a thermotolerant yeast, e.g., a species of *Blastobotrys*, such as *Blastobotrys adeninivorans*. In some embdodiments, the yeast is a non-oleaginous yeast. Engineered organisms often are cultured under conditions that optimize yield of a target molecule. For example, culture conditions can

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1586 of 3077

be selected to balance the levels of one or more of the following activities in order to optimize target molecule yield: carnitine acetyltransferase, acetyl-carnitine translocase, acetyl-CoA carboxylase, ATP citrate lyase, acetyl-CoA hydrolase, acetyl-CoA synthetase, thioesterase, acyl-CoA synthetase, monooxygenase, cytochrome P450 reductase, alcohol dehydrogenase, alcohol oxidase, aldehyde dehydrogenase, acyl-CoA oxidase, 3-ketoacyl-CoA thiolase, peroxisomal transporter, peroxisome biogenesis factor, fatty acid synthase activity and multifunctional enzyme (e.g., enoyl-CoA hydratase and 3-hydroxyacyl-CoA dehydrogenase) activities. In general, non-limiting examples of conditions that may be optimized include the type and amount of carbon source, the type and amount of nitrogen source, the carbon-to-nitrogen ratio, the oxygen level, growth temperature, pH, length of the biomass production phase, length of target product accumulation phase, and time of cell harvest.

A suitable pH range for fermentation often is between about pH 4.0 to about pH 8.0, where a pH in the range of about pH 5.5 to about pH 7.0 sometimes is utilized for initial culture conditions. Depending on the host cell or organism, culturing may be conducted under aerobic or anaerobic conditions, where microaerobic conditions sometimes are maintained. In embodiments in which a *Candida* yeast (e.g., *C. tropicalis* or *C. viswanathii*) is used as the host microorganism, aerobic conditions can be optimal. A two-stage process may be utilized, where one stage promotes organism proliferation and another stage promotes production of target molecule. Suitable temperatures for culturing microorganisms generally are in the range of 28°C to 35°C. However, some organisms are able to survive and grow in more extreme temperatures. Thermotolerant and/or osmotolerant organisms can be well suited for use in industrial production systems operating at elevated temperatures and/or osmotic pressures that would impair growth and/or metabolism and/or completely inactivate organisms that are not thermotolerant and/or osmotolerant. In some instances, production efficiency can be improved and production costs reduced in using such organisms due to decreases in losses and avoidance of implementation of cooling processes. Nitrogen may be supplied from an inorganic (e.g., $(NH_4)_2SO_4$) or organic source (e.g., urea or glutamate). In addition to appropriate carbon and nitrogen sources, culture media also can contain suitable minerals, salts, cofactors, buffers, vitamins, metal ions (e.g., Mn^{+2}, Co^{+2}, Zn^{+2}, Mg^{+2}) and other components suitable for culture of microorganisms.

Engineered microorganisms sometimes are cultured in complex media (e.g., yeast extract-peptone-dextrose broth (YPD)). In some embodiments, engineered microorganisms are

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1587 of 3077

cultured in a defined minimal media that lacks a component necessary for growth and thereby forces selection of a desired expression cassette (e.g., Yeast Nitrogen Base (DIFCO Laboratories, Detroit, Mich.)). Culture media in some embodiments are common commercially prepared media, such as Yeast Nitrogen Base (DIFCO Laboratories, Detroit, Mich.). Other defined or synthetic growth media may also be used. In specific embodiments, yeast are cultured in YPD media (10 g/L Bacto Yeast Extract, 20 g/L Bacto Peptone, and 20 g/L Dextrose). Filamentous fungi, in particular embodiments, can be grown in CM (Complete Medium) containing 10 g/L Dextrose, 2 g/L Bacto Peptone, 1g/L Bacto Yeast Extract, 1 g/L Casamino acids, 50 mL /L 20X Nitrate Salts (120 g/L $NaNO_3$, 10.4 g/L KCl, 10.4 g/L $MgSO_4 \cdot 7 H_2O$), 1 mL/L 1000X Trace Elements (22 g/L $ZnSO_4 \cdot 7 H_2O$, 11 g/L H_3BO_3, 5 g/L $MnCl_2 \cdot 7 H_2O$, 5 g/L $FeSO_4 \cdot 7 H_2O$, 1.7 g/L $CoCl_2 \cdot 6 H_2O$, 1.6 g/L $CuSO_4 \cdot 5 H_2O$, 1.5 g/L $Na_2MoO_4 \cdot 2 H_2O$, and 50 g/L Na_4EDTA), and 1 mL/L Vitamin Solution (100 mg each of Biotin, pyridoxine, thiamine, riboflavin, p-aminobenzoic acid, and nicotinic acid in 100 mL water).

A variety of fermentation processes may be applied for commercial biological production of a target product. In some embodiments, commercial production of a target product from a recombinant microbial host is conducted using a batch, fed-batch or continuous fermentation process, for example. A batch fermentation process often is a closed system where the media composition is fixed at the beginning of the process and not subject to further additions beyond those required for maintenance of pH and oxygen level during the process. At the beginning of the culturing process the media is inoculated with the desired organism and growth or metabolic activity is permitted to occur without adding additional sources (i.e., carbon and nitrogen sources) to the medium. In batch processes the metabolite and biomass compositions of the system change constantly up to the time the culture is terminated. In a typical batch process, cells proceed through a static lag phase to a high-growth log phase and finally to a stationary phase, where the growth rate is diminished or halted. Left untreated, cells in the stationary phase will eventually die.

A variation of the standard batch process is the fed-batch process, where the carbon source is continually added to the fermenter over the course of the fermentation process. Fed-batch processes are useful when catabolite repression is apt to inhibit the metabolism of the cells or where it is desirable to have limited amounts of carbon source in the media at any one time. Measurement of the carbon source concentration in fed-batch systems may be estimated on the basis of the changes of measurable factors such as pH, dissolved oxygen and the partial

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1588 of 3077

pressure of waste gases (e.g., CO_2). Batch and fed-batch culturing methods are known in the art. Examples of such methods may be found in Thomas D. Brock in Biotechnology: A Textbook of Industrial Microbiology, 2.sup.nd ed., (1989) Sinauer Associates Sunderland, Mass. and Deshpande, Mukund V., *Appl. Biochem. Biotechnol. 36*:227 (1992).

In a continuous fermentation process, a defined media often is continuously added to a bioreactor while an equal amount of culture volume is removed simultaneously for product recovery. Continuous cultures generally maintain cells in the log phase of growth at a constant cell density. Continuous or semi-continuous culture methods permit the modulation of one factor or any number of factors that affect cell growth or end product concentration. For example, an approach may limit the carbon source and allow all other parameters to moderate metabolism. In some systems, a number of factors affecting growth may be altered continuously while the cell concentration, measured by media turbidity, is kept constant. Continuous systems often maintain steady state growth and thus the cell growth rate often is balanced against cell loss due to media being drawn off the culture. Methods of modulating nutrients and growth factors for continuous culture processes, as well as techniques for maximizing the rate of product formation, are known and a variety of methods are detailed by Brock, supra.

A non-limiting exemplary fermentation protocol, which includes an initial batch growth phase followed by a fed-batch production, or conversion, phase, that can be used for production of target molecules using modified yeast strains, such as some of those provided herein, is as follows. Fermentation medium of composition 14.0 g/L ammonium sulfate, 10.2 g/L potassium phosphate monobasic, 1.0 g/L magnesium sulfate, 0.2 g/L calcium chloride, 120 mg/L citric acid, 46 mg/L ferric chloride, 0.4 mg/L biotin, 54 g/L glucose and 2X trace metals mix is filter sterilized and transferred to a sterile fermentation vessel. Growth of a yeast strain, e.g., *Candida viswanathii*, is initiated with a 5% inoculum (initial OD600 nm = 1.0) and growth conditions of 35°C, 1000 rpm, 1 vvm, pH 5.8 and initial volume of 1.0 L. Growth continues for approximately 15 hours before exhaustion of the initial carbon source. The temperature control is changed to 30°C and the conversion phase is initiated by the addition of oleic acid to 5 g/L. At the same time as the oleic acid bolus, a continuous feed of oleic acid is initiated at a rate of 1.5 g/L-h. Fermentation conditions are maintained at 30°C, 1000 rpm, 1 vvm, and pH 5.8 for 24 hours at which point the pH set-point is changed to 3.5. The fermentation is carried out for a total of about 135 hours. Samples are collected for GC analysis every 24 hours after initiating

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1589 of 3077

the conversion phase. In this protocol, cells are allowed to grow on glucose as a carbon source in the initial phase (i.e., growth phase) until the glucose is depleted. At this point, a different carbon source, e.g., a fatty acid such as oleic acid, is introduced into the fermentor (i.e., the conversion phase). This new carbon source is continuously fed into the fermentor to initiate and maintain target molecule production (i.e., production phase) that involves oxidative metabolic pathways (e.g., ω-oxidation and β-oxidation) and fatty acid-induced enzyme expression.

Feedstocks, media, supplements and additives

Culture media generally contain a suitable carbon source. Carbon sources useful for culturing cells, microorganisms and/or fermentation processes sometimes are referred to as feedstocks. The term "feedstock" as used herein refers to a composition containing a carbon source that is provided to a cell or organism, which is used by the cell or organism to produce energy and metabolic products useful for growth. A feedstock may be a natural substance, a "man-made substance," a purified or isolated substance, a mixture of purified substances, a mixture of unpurified substances or combinations thereof. A feedstock often is prepared by and/or provided to a cell or organism by a person, and a feedstock often is formulated prior to administration to the cell or organism. A carbon source may include, but is not limited to including, one or more of the following substances: monosaccharides (e.g., also referred to as "saccharides," which include 6-carbon sugars (e.g., glucose, fructose), 5-carbon sugars (e.g., xylose and other pentoses) and the like), disaccharides (e.g., lactose, sucrose), oligosaccharides (e.g., glycans, homopolymers of a monosaccharide), polysaccharides (e.g., starch, cellulose, heteropolymers of monosaccharides or mixtures thereof), sugar alcohols (e.g., glycerol), and renewable feedstocks (e.g., cheese whey permeate, cornsteep liquor, sugar beet molasses, barley malt).

A carbon source also may include a metabolic product that can be used directly as a metabolic substrate in an engineered pathway described herein, or indirectly via conversion to a different molecule using engineered or native biosynthetic pathways in an engineered cell or microorganism. In some embodiments, a carbon source may include glycerol backbones generated by the action of an engineered pathway including at least a lipase activity. In certain embodiments, metabolic pathways can be preferentially biased towards production of a desired product by increasing the levels of one or more activities in one or more metabolic pathways having and/or generating at least one common metabolic and/or synthetic substrate. In some

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1590 of 3077

embodiments, a metabolic byproduct (e.g., glycerol) of an engineered activity (e.g., lipase activity) can be used in one or more metabolic pathways such as gluconeogenesis, pentose phosphate pathway, glycolysis, fatty acid synthesis, beta oxidation, and omega oxidation, to generate a carbon source that can be converted to a target molecule, e.g., adipic acid.

In some embodiments, a feedstock includes a mixture of carbon sources, where each carbon source in the feedstock is selected based on the genotype of the cultured cell or organism. In certain embodiments, a mixed carbon source feedstock includes one or more carbon sources selected from sugars, cellulose, fatty acids, triacylglycerides, paraffins, the like and combinations thereof.

In some embodiments a feedstock is selected according to the genotype and/or phenotype of the cell or organism that is cultured. For example, as described herein, for the production of certain target molecules (e.g., dicarboxylic acids of a particular carbon chain length) the activities of oxidative processes, such as β-oxidation, can be altered through genetic modification of a host cell or organism. In some instances, the catalytic activities and/or substrate specificities of, for example, one or more acyl-CoA oxidases of the host cell β-oxidation pathway can be modified in order to ensure that carbon sources of a particular chain length are or are not subject to degradation. The feedstock used in target molecule production by such modified cells or organisms can be selected to enhance the production process. For example, a feedstock rich in 12-carbon fatty acids, 12-carbon dicarboxylic acids or 12-carbon paraffins, or a mixture of 10, 12 and 14-carbon compounds can be useful for culturing yeast strains harboring an alteration that partially blocks beta oxidation by disrupting POX4 activity, as described herein. Non-limiting examples of carbon sources having 10 to 14 carbons include fats (e.g., coconut oil, palm kernel oil), paraffins (e.g., alkanes, alkenes, or alkynes) having 10 to 14 carbons, (e.g., dodecane (also referred to as adakane12, bihexyl, dihexyl and duodecane); tetradecane), alkene and alkyne derivatives), fatty acids (decanoic, dodecanoic acid, tetradecanoic acid), fatty alcohols (decanol, dodecanol, tetradecanol), the like, non-toxic substituted derivatives or combinations thereof.

In certain embodiments involving genetically modified cells or organisms having partially blocked beta-oxidation pathways, feedstocks suitable for use include, but are not limited to, fatty acid distillates or soapstocks of renewable oils (palm oil fatty acid distillate, soybean oil soapstock, coconut oil soapstock), renewable oils (coconut oil, palm oil, palm kernel oil,

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1591 of 3077

soybean oil, corn oil, and the like), fatty acids of chain length equal to or greater than C10 (in substantially single form (e.g., in substantially pure form) or in mixture form, alkanes of chain length equal to or greater than C10 in substantially single form (e.g., substantially pure form) or in mixture form. Any suitable alkane, fatty acid, fatty alcohol, plant based oil, seed based oil, non-petroleum derived soap stock or the like can be used as the feedstock for the cell or organism (e.g., dodecane, methyl laurate, lauric acid, carbon sources having 10 or greater carbons (e.g. for sebacic acid production) or carbon sources having 12 or greater carbons (e.g. for dodecanedioic acid production)). In some embodiments, carbon sources with greater than 12 carbons can be metabolized using naturally occurring and/or engineered pathways to yield molecules that can be further metabolized using the beta oxidation pathway.

In some embodiments, one acyl-CoA oxidase activity of the beta-oxidation pathway of a host cell or organism is engineered such that it is enhanced, and in certain embodiments, another acyl-CoA oxidase activity in the beta-oxidation pathway is altered to reduce or eliminate the activity, thereby optimizing the production of a diacid of a desired chain-length or diacids with a distribution of desired chain lengths. In some embodiments, an acyl-CoA oxidase is selected and/or engineered to alter the substrate specificity of the enzyme. In certain embodiments, the substrate specificity of a heterologous and/or engineered acyl-CoA oxidase is for carbon chain lengths of between about 12 carbons and about 18 carbons, and in some embodiments a heterologous and/or engineered acyl-CoA oxidase exhibits no activity on substrates below 12 carbons in length. In certain embodiments, a heterologous acyl-CoA oxidase with a desired chain length specificity can be isolated from any suitable organism. In certain embodiments, a carbon source starting material (e.g., alkane, fatty acid, fatty alcohol, dicarboxylic acid) of intermediate or long chain length (e.g., between about 10 carbons and 22 carbons) is converted into an acyl-CoA derivative for entry into the beta-oxidation pathway. A fatty acid can be processed using omega oxidation to yield a dicarboxylic acid (e.g., dodecanedioic acid).

Examples of carbon sources

A carbon source may include, but is not limited to including, one or more of the following substances: alkanes, alkenes, mono-carboxylic acids, di-carboxylic acids, monosaccharides (e.g., also referred to as “saccharides,” which include 6-carbon sugars (e.g., glucose, fructose), 5-carbon sugars (e.g., xylose and other pentoses) and the like), disaccharides (e.g., lactose, sucrose), oligosaccharides (e.g., glycans, homopolymers of a monosaccharide),

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1592 of 3077

polysaccharides (e.g., starch, cellulose, heteropolymers of monosaccharides or mixtures thereof), sugar alcohols (e.g., glycerol), and renewable feedstocks (e.g., cheese whey permeate, cornsteep liquor, sugar beet molasses, barley malt).

Carbon sources also can be selected from one or more of the following non-limiting examples: paraffin (e.g., saturated paraffin, unsaturated paraffin, substituted paraffin, linear paraffin, branched paraffin, or combinations thereof); alkanes (e.g., hexane, dodecane), alkenes or alkynes, each of which may be linear, branched, saturated, unsaturated, substituted or combinations thereof (described in greater herein); linear or branched alcohols (e.g., hexanol, dodecanol); saturated or unsaturated fatty acids (e.g., each fatty acid is about 1 carbon to about 60 carbons with 0 to 10 unsaturations, including free fatty acids, mixed fatty acids, single fatty acid, purified fatty acids (e.g., single fatty acid or mixture of fatty acids), fatty acid distillates, soap stocks, the like and combinations thereof); esters of fatty acids; salts of fatty acids, monoglycerides; diglycerides; triglycerides, phospholipids. Non-limiting commercial sources of products for preparing feedstocks include plants, plant oils or plant products (e.g., vegetable oils (e.g., almond oil, canola oil, cocoa butter, coconut oil, corn oil, cottonseed oil, flaxseed oil, grape seed oil, illipe, jatropha oil, olive oil, palm oil, palm olein, palm kernel oil, rapeseed oil, safflower oil, peanut oil, soybean oil, sesame oil, shea nut oil, sunflower oil, walnut oil, the like and combinations thereof) and vegetable oil products), purified fatty acids (e.g., myristoleic acid, palmitoleic acid, sapienic acid, oleic acid, elaidic acid, vaccenic acid, linoleic acid, linoelaidic acid, α-linolenic acid, arachidonic acid, eicosapentaenoic acid, erucic acid and docosahexaenoic acid) and animal fats (e.g., beef tallow, butterfat, lard, cod liver oil). A carbon source may include a petroleum product and/or a petroleum distillate (e.g., diesel, fuel oils, gasoline, kerosene, paraffin wax, paraffin oil, petrochemicals). In some embodiments, a feedstock comprises petroleum distillate. A carbon source can be a fatty acid distillate (e.g., a palm oil distillate or corn oil distillate). Fatty acid distillates can be by-products from the refining of crude plant oils. In some embodiments, a feedstock comprises a fatty acid distillate.

In some embodiments, a feedstock comprises a soapstock (i.e. soap stock). A widely practiced method for purifying crude vegetable oils for edible use is the alkali or caustic refining method. This process employs a dilute aqueous solution of caustic soda to react with the free fatty acids present which results in the formation of soaps. The soaps together with hydrated phosphatides, gums and prooxidant metals are typically separated from the refined oil as the heavy phase discharge from the refining centrifuge and are typically known as soapstock.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1593 of 3077

A carbon source also may include a metabolic product that can be used directly as a metabolic substrate in an engineered pathway described herein, or indirectly via conversion to a different molecule using engineered or native biosynthetic pathways in an engineered cell or microorganism. In some embodiments, a carbon source may include glycerol backbones generated by the action of an engineered pathway including at least a lipase activity. In certain embodiments, metabolic pathways can be preferentially biased towards production of a desired product by increasing the levels of one or more activities in one or more metabolic pathways having and/or generating at least one common metabolic and/or synthetic substrate. In some embodiments, a metabolic byproduct (e.g., fatty acid, glycerol) of an engineered activity (e.g., ω- oxidation activity, lipase activity) can be used in one or more metabolic pathways, such as gluconeogenesis, pentose phosphate pathway, glycolysis, fatty acid synthesis, beta oxidation, and omega oxidation, to generate a carbon source that can be converted to a fatty dicarboxylic acid (e.g., adipic acid, octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid) or other target molecule.

A carbon source can be an organic acid, including, but not limited to, fatty acids, diacids and β-hydroxy acids (e.g., hydroxyalkanoate monomers). As used herein, “organic acid” and “fatty acid” encompass the free-acid forms thereof and salts or esters thereof. Fatty acids are aliphatic acids of varying carbon chain lengths. Fatty acids generally have a formula that includes: R1-COOR2. In some embodiments, R1 can be an aliphatic group, and can include 1 to 30 carbon atoms, or 6 to 24 carbon atoms, and R2 can be hydrogen, methyl, ethyl, propyl or butyl. For example, R1 can include about 1 carbon atom, about 2 carbon atoms, about 3 carbon atoms, about 4 carbon atoms, about 5 carbon atoms, about 6 carbon atoms, about 7 carbon atoms, about 8 carbon atoms, about 9 carbon atoms, about 10 carbon atoms, about 12 carbon atoms, about 14 carbon atoms, about 16 carbon atoms, about 18 carbon atoms, about 20 carbon atoms, about 22 carbon atoms, about 24 carbon atoms, about 26 carbon atoms, about 28 carbon atoms or about 30 carbon atoms. Naturally occurring fatty acids in biological systems generally contain an even number of carbon atoms, typically between about 12 to about 24, or about 14 to about 24, and most commonly, 16 or 18 carbon atoms. Based on the number of carbons in a fatty acid carbon chain, it can be categorized as a short-, medium- or long-chain fatty acid. Generally, short-chain fatty acids have a chain length of about 2 to about 6 carbon atoms, medium-chain fatty acids have a chain length of about 8 to about 10 carbon atoms, long-chain fatty acids have a chain length of about 12 to about 20 carbon atoms and

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1594 of 3077

very long-chain length fatty acids have a chain length of about 22 or about 24 or more carbon atoms. The carbon atom bonds in the alkyl chain may all be single bonds (i.e., a saturated fatty acid) or may contain one or more double bonds (i.e., an unsaturated fatty acid). Unsaturated fatty acids having one double bond are also referred to as monoenoic; unsaturated fatty acids having two or more double bonds in the carbon chain are also referred to as polyenoic and polyunsaturated (PUFA). The carbon chain in a fatty acid may also be substituted with hydroxyl, methyl, or other groups in place of a hydrogen. Carboxylic acids, such as fatty acids, can partially dissociate in aqueous media and exist as undissociated, uncharged molecules and as a dissociated, anionic form.

Fatty acids containing one carboxyl group can also be referred to as monocarboxylic fatty acids. A fatty acid containing two carboxyl groups (e.g., α,ω-dicarboxylic acids) is a fatty dicarboxylic acid, also referred to herein as a diacid. Fatty dicarboxylic acids generally have a formula that includes: R1OOC-R-COOR2. In some embodiments, R can be an aliphatic group, and can include 1 to 30 carbon atoms, or 4 to 24 carbon atoms, and R1 and R2 can be hydrogen, methyl, ethyl, propyl or butyl. For example, R can include about 1 carbon atom, about 2 carbon atoms, about 3 carbon atoms, about 4 carbon atoms, about 5 carbon atoms, about 6 carbon atoms, about 7 carbon atoms, about 8 carbon atoms, about 9 carbon atoms, about 10 carbon atoms, about 12 carbon atoms, about 14 carbon atoms, about 16 carbon atoms, about 18 carbon atoms, about 20 carbon atoms, about 22 carbon atoms, about 24 carbon atoms, about 26 carbon atoms, about 28 carbon atoms or about 30 carbon atoms. An example of a diacid is adipic acid (hexanedioic acid) which contains six carbon atoms. A diacid sometimes is a C4 to a C24 diacid (i.e., a diacid containing 4 carbons to 24 carbons) and sometimes is a C8, C10, C12, C14, C16, C18, or C20 diacid. A hydrocarbon portion of a diacid sometimes is fully saturated and sometimes a diacid includes one or more unsaturations (e.g., double bonds). In some embodiments, genetically modified cells and organisms and processes provided herein are capable of producing a diacid.

Non-limiting examples of diacids include octadecanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid and other organic intermediates. Non-limiting examples of fatty dicarboxylic acids include adipic acid (hexanedioic acid, 1,4-butanedicarboxylic acid), suberic acid (i.e., octanedioic acid, 1,8-octanedioic acid, octanedioic acid, octane-1,8-dioic acid, 1,6-hexanedicarboxylic acid, capryllic diacids), sebacic acid (i.e., 1,10-decanedioic acid, decanedioic acid, decane-1,10-dioic acid,

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1595 of 3077

1,8-octanedicarboxylic acid, capric diacid), azelaic acid, undecanedioc acid, dodecanedioic acid (i.e., DDDA, 1,12-dodecanedioic acid, dodecanedioic acid, dodecane-1,12-dioic acid, 1,10-decanedicarboxylic acid, decamethylenedicaboxylic acid, 1,10-dicarboxydecane, lauric diacid), tetradecanedioic acid (i.e., TDDA, 1,14-tetradecanedioic acid, tetradecanedioic acid, tetradecane-1,14-dioic acid, 1,12-dodecanedicarboxylic acid, myristic diacid), thapsic acid (i.e., hexadecanedioic acid, 1,16-hexadecanedioic acid, hexadecanedioic acid, hexadecane-1,16-dioic acid, 1,14-tetradecanedicarboxylic acid, palmitic diacid), cis-9-hexadecenedioic acid (i.e., palmitoleic diacids), octanedioic acid (i.e., 1,18-octadecanedioic acid, octadecanedioic acid, octadecane-1,18-dioic acid, 1,16-hexadecanedicarboxylic acid, stearic diacid), cis-9-octadecenedioic acid (i.e., oleic diacids), cis-9,12-octadecenedioic acid (i.e., linoleic diacids), cis-9,12,15-octadecenedioic acid (i.e., linolenic diacids), arachidic diacid (i.e., eicosanoic diacid, icosanoic diacid), 11-eicosenoic diacid (i.e., cis-11-eicosenedioic acid), 13-eicosenoic diacids (i.e., cis-13-eicosenedioic acid), arachidonic diacid (i.e., cis-5,8,11,14-eicosatetraenedioic acid) and salts and esters of fatty acids, including, for example, any of the foregoing diacids.

The term “paraffin” as used herein refers to the common name for alkane hydrocarbons, independent of the source (e.g., plant derived, petroleum derived, chemically synthesized, fermented by a microorganism), or carbon chain length. A carbon source sometimes comprises a paraffin, and in some embodiments, a paraffin is predominant in a carbon source (e.g., about 75%, 80%, 85%, 90% or 95% paraffin). A paraffin sometimes is saturated (e.g., fully saturated), sometimes includes one or more unsaturations (e.g., about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 unsaturations) and sometimes is substituted with one or more non-hydrogen substituents. Non-limiting examples of non-hydrogen substituents include halo, acetyl, =O, =N-CN, =N-OR, =NR, OR, NR_2, SR, SO_2R, SO_2NR_2, $NRSO_2R$, $NRCONR_2$, NRCOOR, NRCOR, CN, COOR, $CONR_2$, OOCR, COR, and NO_2, where each R is independently H, C1-C8 alkyl, C2-C8 heteroalkyl, C1-C8 acyl, C2-C8 heteroacyl, C2-C8 alkenyl, C2-C8 heteroalkenyl, C2-C8 alkynyl, C2-C8 heteroalkynyl, C6-C10 aryl, or C5-C10 heteroaryl, and each R is optionally substituted with halo, =O, =N-CN, =N-OR’, =NR’, OR’, NR'_2, SR’, SO_2R', $SO_2NR'_2$, $NR'SO_2R'$, $NR'CONR'_2$, NR’COOR’, NR’COR’, CN, COOR’, $CONR'_2$, OOCR’, COR’, and NO_2, where each R’ is independently H, C1-C8 alkyl, C2-C8 heteroalkyl, C1-C8 acyl, C2-C8 heteroacyl, C6-C10 aryl or C5-C10 heteroaryl.

A carbon source sometimes comprises an alkyl, alkenyl or alkynyl compound or molecule (e.g., a compound that includes an alkyl, alkenyl or alkynyl moiety (e.g., alkane, alkene, alkyne)). In

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1596 of 3077

certain embodiments, an alkyl, alkenyl or alkynyl molecule, or combination thereof, is predominant in a carbon source (e.g., about 75%, 80%, 85%, 90% or 95% of such molecules). As used herein, the terms “alkyl,” “alkenyl” and “alkynyl” include straight-chain (referred to herein as "linear"), branched-chain (referred to herein as "non-linear"), cyclic monovalent hydrocarbyl radicals, and combinations of these, which contain only C and H atoms when they are unsubstituted. Non-limiting examples of alkyl moieties include methyl, ethyl, isobutyl, cyclohexyl, cyclopentylethyl, 2-propenyl, 3-butynyl, and the like. An alkyl that contains only C and H atoms and is unsubstituted sometimes is referred to as "saturated." An alkenyl or alkynyl generally is "unsaturated" as it contains one or more double bonds or triple bonds, respectively. An alkenyl can include any number of double bonds, such as 1, 2, 3, 4 or 5 double bonds, for example. An alkynyl can include any number of triple bonds, such as 1, 2, 3, 4 or 5 triple bonds, for example. Alkyl, alkenyl and alkynyl molecules sometimes contain between about 2 to about 60 carbon atoms (C). For example, an alkyl, alkenyl and alkynyl molecule can include about 1 carbon atom, about 2 carbon atoms, about 3 carbon atoms, about 4 carbon atoms, about 5 carbon atoms, about 6 carbon atoms, about 7 carbon atoms, about 8 carbon atoms, about 9 carbon atoms, about 10 carbon atoms, about 12 carbon atoms, about 14 carbon atoms, about 16 carbon atoms, about 18 carbon atoms, about 20 carbon atoms, about 22 carbon atoms, about 24 carbon atoms, about 26 carbon atoms, about 28 carbon atoms, about 30 carbon atoms, about 32 carbon atoms, about 34 carbon atoms, about 36 carbon atoms, about 38 carbon atoms, about 40 carbon atoms, about 42 carbon atoms, about 44 carbon atoms, about 46 carbon atoms, about 48 carbon atoms, about 50 carbon atoms, about 52 carbon atoms, about 54 carbon atoms, about 56 carbon atoms, about 58 carbon atoms or about 60 carbon atoms. In some embodiments, paraffins can have a mean number of carbon atoms of between about 8 to about 18 carbon atoms (e.g., about 8 carbon atoms, about 9 carbon atoms, about 10 carbon atoms, about 11 carbon atoms, about 12 carbon atoms, about 13 carbon atoms, about 14 carbon atoms, about 15 carbon atoms, about 16 carbon atoms, about 17 carbon atoms and about 18 carbon atoms). A single group can include more than one type of multiple bond, or more than one multiple bond. Such groups are included within the definition of the term “alkenyl” when they contain at least one carbon-carbon double bond, and are included within the term “alkynyl” when they contain at least one carbon-carbon triple bond. Alkyl, alkenyl and alkynyl molecules include molecules that comprise an alkyl, alkenyl and/or alkynyl moiety, and include molecules that consist of an alkyl, alkenyl or alkynyl moiety (i.e., alkane, alkene and alkyne molecules).

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1597 of 3077

Alkyl, alkenyl and alkynyl substituents sometimes contain 1-20C (alkyl) or 2-20C (alkenyl or alkynyl). They can contain about 8-14C or about 10-14C in some embodiments. A single group can include more than one type of multiple bond, or more than one multiple bond. Such groups are included within the definition of the term "alkenyl" when they contain at least one carbon-carbon double bond, and are included within the term "alkynyl" when they contain at least one carbon-carbon triple bond.

Alkyl, alkenyl and alkynyl groups or compounds sometimes are substituted to the extent that such substitution can be synthesized and can exist. Typical substituents include, but are not limited to, halo, acetyl, =O, =N-CN, =N-OR, =NR, OR, NR_2, SR, SO_2R, SO_2NR_2, $NRSO_2R$, $NRCONR_2$, NRCOOR, NRCOR, CN, COOR, $CONR_2$, OOCR, COR, and NO_2, where each R is independently H, C1-C8 alkyl, C2-C8 heteroalkyl, C1-C8 acyl, C2-C8 heteroacyl, C2-C8 alkenyl, C2-C8 heteroalkenyl, C2-C8 alkynyl, C2-C8 heteroalkynyl, C6-C11 aryl, or C5-C11 heteroaryl, and each R is optionally substituted with halo, =O, =N-CN, =N-OR', =NR', OR', NR'_2, SR', SO_2R', $SO_2NR'_2$, $NR'SO_2R'$, $NR'CONR'_2$, NR'COOR', NR'COR', CN, COOR', $CONR'_2$, OOCR', COR', and NO_2, where each R' is independently H, C1-C8 alkyl, C2-C8 heteroalkyl, C1-C8 acyl, C2-C8 heteroacyl, C6-C10 aryl or C5-C10 heteroaryl. Alkyl, alkenyl and alkynyl groups can also be substituted by C1-C8 acyl, C2-C8 heteroacyl, C6-C10 aryl or C5-C10 heteroaryl, each of which can be substituted by the substituents that are appropriate for the particular group.

"Acetylene" substituents are 2-10C alkynyl groups that are optionally substituted, and are of the formula -C≡C-Ri, where Ri is H or C1-C8 alkyl, C2-C8 heteroalkyl, C2-C8 alkenyl, C2-C8 heteroalkenyl, C2-C8 alkynyl, C2-C8 heteroalkynyl, C1-C8 acyl, C2-C8 heteroacyl, C6-C10 aryl, C5-C10 heteroaryl, C7-C12 arylalkyl, or C6-C12 heteroarylalkyl, and each Ri group is optionally substituted with one or more substituents selected from halo, =O, =N-CN, =N-OR', =NR', OR', NR'2, SR', SO_2R', $SO_2NR'_2$, $NR'SO_2R'$, $NR'CONR'_2$, NR'COOR', NR'COR', CN, COOR', $CONR'_2$, OOCR', COR', and NO_2, where each R' is independently H, C1-C6 alkyl, C2-C6 heteroalkyl, C1-C6 acyl, C2-C6 heteroacyl, C6-C10 aryl, C5-C10 heteroaryl, C7-12 arylalkyl, or C6-12 heteroarylalkyl, each of which is optionally substituted with one or more groups selected from halo, C1-C4 alkyl, C1-C4 heteroalkyl, C1-C6 acyl, C1-C6 heteroacyl, hydroxy, amino, and =O; and where two R' can be linked to form a 3-7 membered ring optionally containing up to three heteroatoms selected from N, O and S. In some embodiments, Ri of -C≡C-Ri is H or Me.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1598 of 3077

A carbon source sometimes comprises a heteroalkyl, heteroalkenyl and/or heteroalkynyl molecule or compound (e.g., comprises heteroalkyl, heteroalkenyl and/or heteroalkynyl moiety (e.g., heteroalkane, heteroalkene or heteroalkyne)). "Heteroalkyl", "heteroalkenyl", and "heteroalkynyl" and the like are defined similarly to the corresponding hydrocarbyl (alkyl, alkenyl and alkynyl) groups, but the 'hetero' terms refer to groups that contain one to three O, S or N heteroatoms or combinations thereof within the backbone; thus at least one carbon atom of a corresponding alkyl, alkenyl, or alkynyl group is replaced by one of the specified heteroatoms to form a heteroalkyl, heteroalkenyl, or heteroalkynyl group. The typical and sizes for heteroforms of alkyl, alkenyl and alkynyl groups are generally the same as for the corresponding hydrocarbyl groups, and the substituents that may be present on the heteroforms are the same as those described above for the hydrocarbyl groups. For reasons of chemical stability, it is also understood that, unless otherwise specified, such groups do not include more than two contiguous heteroatoms except where an oxo group is present on N or S as in a nitro or sulfonyl group.

The term "alkyl" as used herein includes cycloalkyl and cycloalkylalkyl groups and compounds, the term "cycloalkyl" may be used herein to describe a carbocyclic non-aromatic compound or group that is connected via a ring carbon atom, and "cycloalkylalkyl" may be used to describe a carbocyclic non-aromatic compound or group that is connected to a molecule through an alkyl linker. Similarly, "heterocyclyl" may be used to describe a non-aromatic cyclic group that contains at least one heteroatom as a ring member and that is connected to the molecule via a ring atom, which may be C or N; and "heterocyclylalkyl" may be used to describe such a group that is connected to another molecule through a linker. The sizes and substituents that are suitable for the cycloalkyl, cycloalkylalkyl, heterocyclyl, and heterocyclylalkyl groups are the same as those described above for alkyl groups. As used herein, these terms also include rings that contain a double bond or two, as long as the ring is not aromatic.

A carbon source sometimes comprises an acyl compound or moiety (e.g., compound comprising an acyl moiety). As used herein, "acyl" encompasses groups comprising an alkyl, alkenyl, alkynyl, aryl or arylalkyl radical attached at one of the two available valence positions of a carbonyl carbon atom, and heteroacyl refers to the corresponding groups where at least one carbon other than the carbonyl carbon has been replaced by a heteroatom chosen from N, O and S. Thus heteroacyl includes, for example, -C(=O)OR and $–C(=O)NR_2$ as well as –C(=O)-heteroaryl.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1599 of 3077

Acyl and heteroacyl groups are bonded to any group or molecule to which they are attached through the open valence of the carbonyl carbon atom. Typically, they are C1-C8 acyl groups, which include formyl, acetyl, pivaloyl, and benzoyl, and C2-C8 heteroacyl groups, which include methoxyacetyl, ethoxycarbonyl, and 4-pyridinoyl. The hydrocarbyl groups, aryl groups, and heteroforms of such groups that comprise an acyl or heteroacyl group can be substituted with the substituents described herein as generally suitable substituents for each of the corresponding component of the acyl or heteroacyl group.

A carbon source sometimes comprises one or more aromatic moieties and/or heteroaromatic moieties. "Aromatic" moiety or "aryl" moiety refers to a monocyclic or fused bicyclic moiety having the well-known characteristics of aromaticity; examples include phenyl and naphthyl. Similarly, "heteroaromatic" and "heteroaryl" refer to such monocyclic or fused bicyclic ring systems which contain as ring members one or more heteroatoms selected from O, S and N. The inclusion of a heteroatom permits aromaticity in 5 membered rings as well as 6 membered rings. Typical heteroaromatic systems include monocyclic C5-C6 aromatic groups such as pyridyl, pyrimidyl, pyrazinyl, thienyl, furanyl, pyrrolyl, pyrazolyl, thiazolyl, oxazolyl, and imidazolyl and the fused bicyclic moieties formed by fusing one of these monocyclic groups with a phenyl ring or with any of the heteroaromatic monocyclic groups to form a C8-C10 bicyclic group such as indolyl, benzimidazolyl, indazolyl, benzotriazolyl, isoquinolyl, quinolyl, benzothiazolyl, benzofuranyl, pyrazolopyridyl, quinazolinyl, quinoxalinyl, cinnolinyl, and the like. Any monocyclic or fused ring bicyclic system which has the characteristics of aromaticity in terms of electron distribution throughout the ring system is included in this definition. It also includes bicyclic groups where at least the ring which is directly attached to the remainder of the molecule has the characteristics of aromaticity. Typically, the ring systems contain 5-12 ring member atoms. The monocyclic heteroaryls sometimes contain 5-6 ring members, and the bicyclic heteroaryls sometimes contain 8-10 ring members.

Aryl and heteroaryl moieties may be substituted with a variety of substituents including C1-C8 alkyl, C2-C8 alkenyl, C2-C8 alkynyl, C5-C12 aryl, C1-C8 acyl, and heteroforms of these, each of which can itself be further substituted; other substituents for aryl and heteroaryl moieties include halo, OR, NR_2, SR, SO_2R, SO_2NR_2, $NRSO_2R$, $NRCONR_2$, NRCOOR, NRCOR, CN, COOR, $CONR_2$, OOCR, COR, and NO_2, where each R is independently H, C1-C8 alkyl, C2-C8 heteroalkyl, C2-C8 alkenyl, C2-C8 heteroalkenyl, C2-C8 alkynyl, C2-C8 heteroalkynyl, C6-C10

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1600 of 3077

aryl, C5-C10 heteroaryl, C7-C12 arylalkyl, or C6-C12 heteroarylalkyl, and each R is optionally substituted as described above for alkyl groups. The substituent groups on an aryl or heteroaryl group may be further substituted with the groups described herein as suitable for each type of such substituents or for each component of the substituent. Thus, for example, an arylalkyl substituent may be substituted on the aryl portion with substituents typical for aryl groups, and it may be further substituted on the alkyl portion with substituents as typical or suitable for alkyl groups.

Similarly, "arylalkyl" and "heteroarylalkyl" refer to aromatic and heteroaromatic ring systems, which are stand-alone molecules (e.g., benzene or substituted benzene, pyridine or substituted pyridine), or which are bonded to an attachment point through a linking group such as an alkylene, including substituted or unsubstituted, saturated or unsaturated, cyclic or acyclic linkers. A linker often is C1-C8 alkyl or a hetero form thereof. These linkers also may include a carbonyl group, thus making them able to provide substituents as an acyl or heteroacyl moiety. An aryl or heteroaryl ring in an arylalkyl or heteroarylalkyl group may be substituted with the same substituents described above for aryl groups. An arylalkyl group sometimes includes a phenyl ring optionally substituted with the groups defined above for aryl groups and a C1-C4 alkylene that is unsubstituted or is substituted with one or two C1-C4 alkyl groups or heteroalkyl groups, where the alkyl or heteroalkyl groups can optionally cyclize to form a ring such as cyclopropane, dioxolane, or oxacyclopentane. Similarly, a heteroarylalkyl group often includes a C5-C6 monocyclic heteroaryl group optionally substituted with one or more of the groups described above as substituents typical on aryl groups and a C1-C4 alkylene that is unsubstituted. A heteroarylalkyl group sometimes is substituted with one or two C1-C4 alkyl groups or heteroalkyl groups, or includes an optionally substituted phenyl ring or C5-C6 monocyclic heteroaryl and a C1-C4 heteroalkylene that is unsubstituted or is substituted with one or two C1-C4 alkyl or heteroalkyl groups, where the alkyl or heteroalkyl groups can optionally cyclize to form a ring such as cyclopropane, dioxolane, or oxacyclopentane.

Where an arylalkyl or heteroarylalkyl group is described as optionally substituted, the substituents may be on the alkyl or heteroalkyl portion or on the aryl or heteroaryl portion of the group. The substituents optionally present on the alkyl or heteroalkyl portion sometimes are the same as those described above for alkyl groups, and the substituents optionally present on the aryl or heteroaryl portion often are the same as those described above for aryl groups generally.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1601 of 3077

"Arylalkyl" groups as used herein are hydrocarbyl groups if they are unsubstituted, and are described by the total number of carbon atoms in the ring and alkylene or similar linker. Thus a benzyl group is a C7-arylalkyl group, and phenylethyl is a C8-arylalkyl.

"Heteroarylalkyl" as described above refers to a moiety comprising an aryl group that is attached through a linking group, and differs from "arylalkyl" in that at least one ring atom of the aryl moiety or one atom in the linking group is a heteroatom selected from N, O and S. The heteroarylalkyl groups are described herein according to the total number of atoms in the ring and linker combined, and they include aryl groups linked through a heteroalkyl linker; heteroaryl groups linked through a hydrocarbyl linker such as an alkylene; and heteroaryl groups linked through a heteroalkyl linker. Thus, for example, C7-heteroarylalkyl includes pyridylmethyl, phenoxy, and N-pyrrolylmethoxy.

"Alkylene" as used herein refers to a divalent hydrocarbyl group. Because an alkylene is divalent, it can link two other groups together. An alkylene often is referred to as $–(CH_2)_n-$ where n can be 1-20, 1-10, 1-8, or 1-4, though where specified, an alkylene can also be substituted by other groups, and can be of other lengths, and the open valences need not be at opposite ends of a chain. Thus –CH(Me)- and $–C(Me)_2-$ may also be referred to as alkylenes, as can a cyclic group such as cyclopropan-1,1-diyl. Where an alkylene group is substituted, the substituents include those typically present on alkyl groups as described herein.

In certain embodiments, the feedstock contains a mixed set of aliphatic molecules from which diacids may be produced. In some embodiments, an aliphatic molecule in the feedstock is the predominant aliphatic species and sometimes a particular fatty acid produced from that aliphatic molecule is the predominant fatty acid species produced. A predominant species generally is 51% or more by weight of aliphatic molecule species in a feedstock or 51% or more by weight of diacid species in a product (e.g., about 55% or more, 60% or more, 65% or more, 70% or more, 75% or more, 80% or more, 85% or more, 90% or more or 95% or more).

Target production, isolation and yield

Provided herein are methods for producing a target molecule or one or more target molecules. For example, in some embodiments, a method for producing one or more target molecules includes culturing a modified cell, microorganism or organism such as any of the modified cells

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1602 of 3077

and organisms described herein, under conditions in which the cell, organism or microorganism produces one or more target molecules. In some embodiments, a method for producing one or more target molecules includes contacting an engineered cell, organism, or microorganism with a carbon source (e.g., a feedstock, including a feedstock containing a fatty acid or alkane) under conditions whereby one or more target molecules is produced. In various embodiments of the methods provided herein, a target molecule, e.g., a fatty acid, including, for example, a dicarboxylic acid (e.g., adipic acid), and salts and/or esters thereof, is isolated or purified from the culture media or extracted from the engineered cells or organisms. Target molecule yield may be expressed as percent (%) theoretical yield, percent (%) maximum theoretical yield, units of target molecule produced per unit of feedstock added (e.g., grams of target molecule produced per gram of feedstock added), units of target molecule per volume of culture (e.g., grams of target molecule per liter of culture), units of target molecule per volume of cells (e.g., grams of target molecule per liter of cells), units of target molecule per weight of cells (e.g., grams of target molecule per dry cell weight (DCW) of cells), units of target molecule per volume of culture per unit of time (e.g., grams of target molecule per liter of culture per hour), and/or fold change (increase or decrease) when comparing target molecule production by a modified cell to production by an unmodified cell.

In some embodiments, fermentation of feedstocks by methods described herein can produce a target molecule product at a level of about 10% to about 100% of theoretical yield (e.g., about 15%, about 20%, about 25% or more of theoretical yield (e.g., 25% or more, 26% or more, 27% or more, 28% or more, 29% or more, 30% or more, 31% or more, 32% or more, 33% or more, 34% or more, 35% or more, 36% or more, 37% or more, 38% or more, 39% or more, 40% or more, 41% or more, 42% or more, 43% or more, 44% or more, 45% or more, 46% or more, 47% or more, 48% or more, 49% or more, 50% or more, 51% or more, 52% or more, 53% or more, 54% or more, 55% or more, 56% or more, 57% or more, 58% or more, 59% or more, 60% or more, 61% or more, 62% or more, 63% or more, 64% or more, 65% or more, 66% or more, 67% or more, 68% or more, 69% or more, 70% or more, 71% or more, 72% or more, 73% or more, 74% or more, 75% or more, 76% or more, 77% or more, 78% or more, 79% or more, 80% or more, 81% or more, 82% or more, 83% or more, 84% or more, 85% or more, 86% or more, 87% or more, 88% or more, 89% or more, 90% or more, 91% or more, 92% or more, 93% or more, 94% or more, 95% or more, 96% or more, 97% or more, 98% or more, or 99% or more of theoretical yield). The term “theoretical yield” as used herein refers to the amount of product that could be made from a starting material if the reaction is 100% complete. Theoretical yield

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1603 of 3077

is based on the stoichiometry of a reaction and ideal conditions in which starting material is completely consumed, undesired side reactions do not occur, the reverse reaction does not occur, and there are no losses in the work-up procedure. Culture media may be tested for target product concentration and drawn off when the concentration reaches a predetermined level. Detection methods are known in the art, including but not limited to chromatographic methods (e.g., gas chromatography) or combined chromatographic/mass spectrometry (e.g., GC-MS) methods. Target product may be present at a range of levels as described herein.

A target product sometimes is retained within an engineered cell or organism after a culture process is completed, and in certain embodiments, the target product is secreted out of the cell or organism into the culture medium. For example, in some embodiments in which a target molecule is secreted, culture media may be drawn from the culture system and fresh medium may be supplemented, and/or target product may be extracted from the culture media during or after the culture process is completed. Engineered cells or organisms may be cultured on or in solid, semi-solid or liquid media. In some embodiments media is drained from cells adhering to a plate. In certain embodiments, a liquid-cell mixture is centrifuged at a speed sufficient to pellet the cells but not disrupt the cells and allow extraction of the media, as known in the art. The cells may then be resuspended in fresh media. Target product may be purified from culture media according to methods known in the art.

Provided herein are non-limiting examples of methods useful for recovering target product from fermentation broth and/or isolating/partially purifying a target product from non-target products when utilizing mixed chain length feedstocks. Recovery of a fatty dicarboxylic acid (e.g., adipic acid, sebacic acid, suberic acid, octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid) from fermentation broth can be accomplished using a variety of methods. Generally, for example, the cells are filtered away from the media, and the target molecule is extracted with a water-imiscible solvent appropriate for target chemical. Optionally, a centrifugation step can first be employed to separate cell mass and a fatty dicarboxylic acid from the aqueous phase.

A fatty dicarboxylic acid has limited solubility in water under fermentation conditions, and can have a density similar to that of water. In some embodiments, upon centrifugation, the majority of fatty dicarboxylic acid may be pulled away from the water stream, and be concentrated in the cell mass stream. The concentrated fatty dicarboxylic acid stream can then be further

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1604 of 3077

concentrated via filtration steps (e.g., solid dodecanedioic acid will be retained on a filter, allowing water to pass through, concentrating the product). Once the fatty dicarboxylic acid is concentrated to the desired level, the temperature can be increased to above its melting point of 130°C. After the fatty dicarboxylic acid is melted, the remaining impurities can be removed via filtration; the final product may be recovered by decreasing the temperature, allowing the fatty dicarboxylic acid to solidify, and collecting the solid product.

Alternatively, a fatty dicarboxylic acid can be recovered from fermentation broth by first extracting the broth with an organic solvent in which a fatty dicarboxylic acid is soluble. The organic solvent phase can then be filtered through various membranes to further purify the fatty dicarboxylic acid. Subsequent extractions with the same or a different organic solvent can then be performed and each round of extraction can be followed by membrane filtration to further concentrate the fatty dicarboxylic acid. The organic solvent can be evaporated, leaving the fatty dicarboxylic acid behind as a residue and the residue can be dried to provide the fatty dicarboxylic acid in solid form.

In certain embodiments, target product is extracted from the cultured engineered cells or organisms. The cells may be concentrated through centrifugation at a speed sufficient to shear the cell membranes. In some embodiments, the cells may be physically disrupted (e.g., shear force, sonication) or chemically disrupted (e.g., contacted with detergent or other lysing agent). The phases may be separated by centrifugation or other method known in the art and target product may be isolated according to known methods.

Commercial grade target product sometimes is provided in substantially pure form (e.g., 90% pure or greater, 95% pure or greater, 99% pure or greater or 99.5% pure or greater). In some embodiments, target product may be modified into any one of a number of downstream products. For example, a fatty dicarboxylic acid (e.g., octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid) may be polycondensed with hexamethylenediamine to produce nylon. Nylon may be further processed into fibers for applications in carpeting, automobile tire cord and clothing. A fatty dicarboxylic acid can also be used for manufacturing plasticizers, lubricant components and polyester polyols for polyurethane systems. Various esters of food grade fatty dicarboxylic acids can be used as components in fragrance manufacture, gelling aids, flavorings, acidulant, leavening and buffering agent. A fatty dicarboxylic acid has two carboxylic

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1605 of 3077

acid (-COOH) groups, which can yield two kinds of salts. Its derivatives, acyl halides, anhydrides, esters, amides and nitriles, are used in making a variety of downstream products through further reactions of substitution, catalytic reduction, metal hydride reduction, diborane reduction, keto formation with organometallic reagents, electrophile bonding at oxygen, and condensation.

Target product may be provided within cultured cells and organisms containing target product, and cultured cells may be supplied fresh or frozen in a liquid media or dried. Fresh or frozen cells or organisms may be contained in appropriate moisture-proof containers that may also be temperature controlled as necessary. Target product sometimes is provided in culture medium that is substantially cell-free. In some embodiments target product or modified target product purified from cells or organisms is provided, and target product sometimes is provided in substantially pure form. In certain embodiments crystallized or powdered target product is provided. For example, dodecanedioic acid (1,12 dodecanedioic acid; DDDA) is a white powder or crystal with a melting point of between 260°F and 266°F. Sebacic acid (1,8 ocatanedicarboxylic acid) is also a white powder or crystal with a melting point of between 268°F and 274°F. A crystallized or powdered fatty dicarboxylic acid may be transported in a variety of containers including one ton cartons, drums, 50 pound bags and the like.

In certain embodiments, a target product is produced with a yield of about 0.50 grams of target product per gram of feedstock or carbon source added, or greater; 0.51 grams of target product per gram of feedstock or carbon source added, or greater; 0.52 grams of target product per gram of feedstock or carbon source added, or greater; 0.53 grams of target product per gram of feedstock or carbon source added, or greater; 0.54 grams of target product per gram of feedstock or carbon source added, or greater; 0.55 grams of target product per gram of feedstock or carbon source added, or greater; 0.56 grams of target product per gram of feedstock or carbon source added, or greater; 0.57 grams of target product per gram of feedstock or carbon source added, or greater; 0.58 grams of target product per gram of feedstock or carbon source added, or greater; 0.59 grams of target product per gram of feedstock or carbon source added, or greater; 0.60 grams of target product per gram of feedstock or carbon source added, or greater; 0.61 grams of target product per gram of feedstock or carbon source added, or greater; 0.62 grams of target product per gram of feedstock or carbon source added, or greater; 0.63 grams of target product per gram of feedstock or carbon source added, or greater; 0.64 grams of target product per gram of

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1606 of 3077

feedstock or carbon source added, or greater; 0.65 grams of target product per gram of feedstock or carbon source added, or greater; 0.66 grams of target product per gram of feedstock or carbon source added, or greater; 0.67 grams of target product per gram of feedstock or carbon source added, or greater; 0.68 grams of target product per gram of feedstock or carbon source added, or greater; 0.69 grams of target product per gram of feedstock or carbon source added, or greater; 0.70 grams of target product per gram of feedstock or carbon source added or greater; 0.71 grams of target product per gram of feedstock or carbon source added, or greater; 0.72 grams of target product per gram of feedstock or carbon source added, or greater; 0.73 grams of target product per gram of feedstock or carbon source added, or greater; 0.74 grams of target product per gram of feedstock or carbon source added, or greater; 0.75 grams of target product per gram of feedstock or carbon source added, or greater; 0.76 grams of target product per gram of feedstock or carbon source added, or greater; 0.77 grams of target product per gram of feedstock or carbon source added, or greater; 0.78 grams of target product per gram of feedstock or carbon source added, or greater; 0.79 grams of target product per gram of feedstock or carbon source added, or greater; 0.80 grams of target product per gram of feedstock or carbon source added, or greater; 0.81 grams of target product per gram of feedstock or carbon source added, or greater; 0.82 grams of target product per gram of feedstock or carbon source added, or greater; 0.83 grams of target product per gram of feedstock or carbon source added, or greater; 0.84 grams of target product per gram of feedstock or carbon source added, or greater; 0.85 grams of target product per gram of feedstock or carbon source added, or greater; 0.86 grams of target product per gram of feedstock or carbon source added, or greater; 0.87 grams of target product per gram of feedstock or carbon source added, or greater; 0.88 grams of target product per gram of feedstock or carbon source added, or greater; 0.89 grams of target product per gram of feedstock or carbon source added, or greater; 0.90 grams of target product per gram of feedstock or carbon source added, or greater; 0.91 grams of target product per gram of feedstock or carbon source added, or greater; 0.92 grams of target product per gram of feedstock or carbon source added, or greater; 0.93 grams of target product per gram of feedstock or carbon source added, or greater; 0.94 grams of target product per gram of feedstock or carbon source added, or greater; 0.95 grams of target product per gram of feedstock or carbon source added, or greater; 0.96 grams of target product per gram of feedstock or carbon source added, or greater; 0.97 grams of target product per gram of feedstock or carbon source added, or greater; 0.98 grams of target product per gram of

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1607 of 3077

feedstock or carbon source added, or greater; 0.99 grams of target product per gram of feedstock or carbon source added, or greater; 1.0 grams of target product per gram of feedstock or carbon source added, or greater; 1.1 grams of target product per gram of feedstock or carbon source added, or greater; 1.2 grams of target product per gram of feedstock or carbon source added, or greater; 1.3 grams of target product per gram of feedstock or carbon source added, or greater; 1.4 grams of target product per gram of feedstock or carbon source added, or greater; 1.5 grams of target product per gram of feedstock or carbon source added, or greater; 1.6 grams of target product per gram of feedstock or carbon source added, or greater; 1.7 grams of target product per gram of feedstock or carbon source added, or greater; 1.8 grams of target product per gram of feedstock or carbon source added, or greater; 1.9 grams of target product per gram of feedstock or carbon source added, or greater; or about 2.0 grams of target product per gram of feedstock or carbon source added, or greater.

Y_{max} is maximum theoretical yield. It is the amount of product that can be produced for a given biochemical pathway given a certain amount of consumed feedstock (e.g., grams adipic acid/grams oleic acid). Programs are available to calculate Y_{max} values that can include algorithms used to calculate flux balance (see, e.g., COBRA; Becker et al. (2007) *Nature Protocols 2*:727-738). In some embodiments, the maximum theoretical yield (Y_{max}) of adipic acid in an engineered cell or organism is about 0.92 grams of adipic acid produced per gram of oleic acid added. In some embodiments, the maximum theoretical yield (Y_{max}) of suberic acid in an engineered cell or organism is about 0.96 grams of suberic acid produced per gram of oleic acid added. In some embodiments, the maximum theoretical yield (Y_{max}) of sebacic acid in an engineered cell or organism is about 0.99 grams of sebacic acid produced per gram of oleic acid added. In some embodiments, the maximum theoretical yield (Y_{max}) of docecanedioic acid in an engineered cell or organism is about 1.02 grams of dodecanedioic acid produced per gram of oleic acid added. In some embodiments, the maximum theoretical yield (Y_{max}) of 3-hydroxy-propionic acid in an engineered cell or organism is about 1.91 grams of 3-hydroxy-propionic acid produced per gram of oleic acid added. In some embodiments, the maximum theoretical yield (Y_{max}) of triacetic acid lactone in an engineered cell or organism is about 1.00 grams of triacetic acid lactone produced per gram of oleic acid added. In some embodiments, the maximum theoretical yield (Y_{max}) of lycopene in an engineered cell or organism is about 1.20 grams of lycopene produced per gram of oleic acid added. In some embodiments of the engineered cells and organisms and target production methods provided herein in which a carbon recycle loop is included in the engineered cell or organism, the Y_{max} for production of a

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1608 of 3077

particular target molecule using a particular carbon source(s) is greater than it is for the same cell or organism that does not include an engineered carbon recycle loop. In some embodiments, for example, the Y_{max} may be at least about 5% to about 100% greater for a modified cell or organism engineered to include a carbon recycle loop than for the same cell or organism that does not contain an engineered recycle loop. For example, in some embodiments, the Y_{max} can be at least about 5%, 6%, 7%, 8%, 9%, 10%, 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19%, 20%, 21%, 22%, 23%, 24%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95% or 100% greater for a modified cell or organism engineered to include a carbon recycle loop than for the same cell or organism that does not contain an engineered recycle loop.

The percentage of Y_{max} for the engineered cell or organism under conditions in which a target molecule is produced is calculated as $(\%Y_{max}) = Y_{p/s} / Y_{max} * 100$, where $(Y_{p/s})$ = [target molecule (g/L)] * final volume of culture in flask (L)] / [feedstock added to flask (g)]. In some embodiments, the engineered cell or organism produces target molecule at about 10% to about 100% of maximum theoretical yield. In some embodiments, the engineered cell or organism produces target molecule at about 10% or greater of maximum theoretical yield, 15% or greater of maximum theoretical yield, 20% or greater of maximum theoretical yield, 25% or greater of maximum theoretical yield, 30% or greater of maximum theoretical yield, 35% or greater of maximum theoretical yield, 40% or greater of maximum theoretical yield, 45% or greater of maximum theoretical yield, 50% or greater of maximum theoretical yield, 55% or greater of maximum theoretical yield, 60% or greater of maximum theoretical yield, 65% or greater of maximum theoretical yield, 70% or greater of maximum theoretical yield, 75% or greater of maximum theoretical yield, 80% or greater of maximum theoretical yield, 85% or greater of maximum theoretical yield, 90% or greater of maximum theoretical yield, 95% or greater of maximum theoretical yield, or 100% of maximum theoretical yield.

In certain embodiments, a target molecule product (e.g., adipic acid, suberic acid, sebacic acid, dodecanedioic acid, 3-hydroxy-propionic acid, triacetic acid lactone, terpene) is produced with a yield of greater than about 0.15 grams per gram of the feedstock or carbon source (e.g., dodecane, mixed chain length alkanes, lauric acid, mixed chain length fatty acids, oil, the like or combinations of the foregoing). In some embodiments, a target molecule product is produced at between about 10% and about 100% of maximum theoretical yield of any introduced feedstock or carbon source ((e.g., about 15%, about 20%, about 25% or more of theoretical yield (e.g.,

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1609 of 3077

25% or more, 26% or more, 27% or more, 28% or more, 29% or more, 30% or more, 31% or more, 32% or more, 33% or more, 34% or more, 35% or more, 36% or more, 37% or more, 38% or more, 39% or more, 40% or more, 41% or more, 42% or more, 43% or more, 44% or more, 45% or more, 46% or more, 47% or more, 48% or more, 49% or more, 50% or more, 51% or more, 52% or more, 53% or more, 54% or more, 55% or more, 56% or more, 57% or more, 58% or more, 59% or more, 60% or more, 61% or more, 62% or more, 63% or more, 64% or more, 65% or more, 66% or more, 67% or more, 68% or more, 69% or more, 70% or more, 71% or more, 72% or more, 73% or more, 74% or more, 75% or more, 76% or more, 77% or more, 78% or more, 79% or more, 80% or more, 81% or more, 82% or more, 83% or more, 84% or more, 85% or more, 86% or more, 87% or more, 88% or more, 89% or more, 90% or more, 91% or more, 92% or more, 93% or more, 94% or more, 95% or more, 96% or more, 97% or more, 98% or more, or 99% or more of theoretical maximum yield).

In certain embodiments, a target molecule product is produced in a concentration range of between about 50 g/L to about 1000g/L of culture media (e.g., about 50 g/L, about 55 g/L, about 60 g/L, about 65 g/L, about 70 g/L, about 75 g/L, about 80 g/L, about 85 g/L, about 90 g/L, about 95 g/L, about 100 g/L, about 110 g/L, about 120 g/L, about 130 g/L, about 140 g/L, about 150 g/L, about 160 g/L, about 170 g/L, about 180 g/L, about 190 g/L, about 200 g/L, about 225 g/L, about 250 g/L, about 275 g/L, about 300 g/L, about 325 g/L, about 350 g/L, about 375 g/L, about 400 g/L, about 425 g/L, about 450 g/L, about 475 g/L, about 500 g/L, about 550 g/L, about 600 g/L, about 650 g/L, about 700 g/L, about 750 g/L, about 800 g/L, about 850 g/L, about 900 g/L, about 950 g/L, or about 1000 g/L).

In some embodiments, a target molecule product is produced at a rate of between about 0.5 g/L/hour to about 5 g/L/hour (e.g., about 0.5 g/L/hour, about 0.6 g/L/hour, about 0.7 g/L/hour, about 0.8 g/L/hour, about 0.9 g/L/hour, about 1.0 g/L/hour, about 1.1 g/L/hour, about 1.2 g/L/hour, about 1.3 g/L/hour, about 1.4 g/L/hour, about 1.5 g/L/hour, about 1.6 g/L/hour, about 1.7 g/L/hour, about 1.8 g/L/hour, about 1.9 g/L/hour, about 2.0 g/L/hour, about 2.25 g/L/hour, about 2.5 g/L/hour, about 2.75 g/L/hour, about 3.0 g/L/hour, about 3.25 g/L/hour, about 3.5 g/L/hour, about 3.75 g/L/hour, about 4.0 g/L/hour, about 4.25 g/L/hour, about 4.5 g/L/hour, about 4.75 g/L/hour, or about 5.0 g/L/hour.)

In certain embodiments, the engineered cell or organism produces between about 5-fold to about 500-fold more target molecule (a particular fatty acid, dicarboxylic acid, or other target

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1610 of 3077

molecule) compared to the amount produced by a wild-type or partially engineered cell or organism of the same strain, under identical fermentation conditions (e.g., about a 5-fold increase, about a 10-fold increase, about a 15-fold increase, about a 20-fold increase, about a 25-fold increase, about a 30-fold increase, about a 35-fold increase, about a 40-fold increase, about a 45-fold increase, about a 50-fold increase, about a 55-fold increase, about a 60-fold increase, about a 65-fold increase, about a 70-fold increase, about a 75-fold increase, about a 80-fold increase, about a 85-fold increase, about a 90-fold increase, about a 95-fold increase, about a 100-fold increase, about a 125-fold increase, about a 150-fold increase, about a 175-fold increase, about a 200-fold increase, about a 250-fold increase, about a 300-fold increase, about a 350-fold increase, about a 400-fold increase, about a 450-fold increase, or about a 500-fold increase).

In some embodiments, the engineered cell or organism produces a yield of target molecule (e.g., a particular fatty acid, dicarboxylic acid, or other target molecule), in terms of the percentage of the maximum theoretical yield (Y_{max}) that the yield is, that is greater than the yield (as a percentage of maximum theoretical yield) produced by a wild-type or partially engineered cell or organism of the same strain, under identical culture conditions. For example, the yield of a target molecule produced by an engineered or modified cell or organism provided herein can be a percent of maximum theoretical yield that is at least about 1 unit or more greater, 2 units or more greater, 3 units or more greater, 4 units or more greater, 5 units or more greater, 6 units or more greater, 7 units or more greater, 8 units or more greater, 9 units or more greater, 10 units or more greater, 11% units or more greater, 12 units or more greater, 13% units or more greater, 14 units or more greater, 15 units or more greater, 16 units or more greater, 17 units or more greater, 18 units or more greater, 19 units or more greater, 20 units or more greater, 21 units or more greater, 22 units or more greater, 23 units or more greater, 24 units or more greater, 25 units or more greater, 26 units or more greater, 27 units or more greater, 28 units or more greater, 29 units or more greater, 30 units or more greater, 31 units or more greater, 32 units or more greater, 33 units or more greater, 34 units or more greater, 35 units or more greater, 36 units or more greater, 37 units or more greater, 38 units or more greater, 39 units or more greater, 40 units or more greater, 41 units or more greater, 42 units or more greater, 43 units or more greater, 44 units or more greater, 45 units or more greater, 46 units or more greater, 47 units or more greater, 48 units or more greater, 49 units or more greater, 50 units or more greater, or more than 50 units greater than the yield (as a percentage of maximum theoretical yield) of the target molecule produced by a wild-type or partially engineered cell or

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1611 of 3077

organism of the same strain (i.e., a reference cell or organism) under the same culture conditions, wherein a unit is defined as 1% of the maximum theoretical yield. As an illustration, if the yield of a target molecule produced by an unmodified reference cell or organism is, for example, 80% of the maximum theoretical yield (Ymax), then a target molecule yield of an engineered cell or organism that is 1 unit greater would be 81% of the maximum theoretical yield, and a target molecule yield of an engineered cell or organism that is 5 units greater would be 85% of the maximum theoretical yield, and a target molecule yield of an engineered cell or organism that is 10 units greater would be 90% of the maximum theoretical yield, and so on.

Examples

The examples set forth below illustrate certain embodiments and do not limit the technology. Certain examples set forth below utilize standard recombinant DNA and other biotechnology protocols known in the art. Many such techniques are described in detail in Maniatis, T., E. F. Fritsch and J. Sambrook (1982) *Molecular Cloning: a Laboratory Manual*; Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. DNA mutagenesis can be accomplished using the Stratagene (San Diego, CA) "QuickChange" kit according to the manufacturer's instructions.

Non-limiting examples of recombinant DNA techniques and genetic manipulation of microorganisms are described herein. In some embodiments, strains of engineered organisms described herein can be mated to combine genetic backgrounds to further enhance carbon flux management through native and/or engineered pathways described herein, for the production of a desired target product (e.g., sebacic or dodecanedioic acid).

EXAMPLE 1: Construction of shuttle vectors and template expression vectors

Nucleotide sequences of nucleic acids and amino acid sequences of polypeptides referred to in the Examples that follow are provided in Examples 24 - 27.

Construction of the shuttle vector pAA061

Vector pAA061 was constructed from a pUC19 backbone to harbor the selectable marker *URA3* (e.g., nucleotide SEQ ID NO: 54 and amino acid SEQ ID NO: 1) from *Candida* strain ATCC 20336 as well as modifications to allow insertion of *Candida* strain ATCC 20336 promoters and

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1612 of 3077

terminators. A 1,507-bp DNA fragment containing the promoter (e.g., SEQ ID NO: 55), ORF (e.g., SEQ ID NO: 54), and terminator (e.g., SEQ ID NO: 56) of *URA3* from *Candida* strain ATCC 20336 was amplified using primers oAA0124 and oAA0125. The *URA3* PCR product was cut with NdeI/MluI and ligated into the 2,505-bp fragment of pUC19 cut with NdeI/BsmBI (an MluI compatible overhang was produced by BsmBI). In order to replace the *lac* promoter with a short 21-bp linker sequence, the resulting plasmid was opened with SphI/SapI and filled in with a linker produced by annealing oligos oAA0173 and oAA0174. The resulting plasmid was named pAA061 (Fig. 12).

Cloning of promoter and terminator pairs into vector pAA061

The multiple cloning site in pAA061 allows the insertion of promoter and terminator sequences from *Candida* strain ATCC 20336 that can subsequently be used to control the expression of an ORF (Fig. 12). The following example describes the insertion of two different promoter/terminator pairs into pAA061. This cloning method is general and can be applied to the insertion of other promoter and/or terminator sequences into pAA061.

Construction of vector pAA105

Vector pAA105 was constructed from base vector pAA061 to include the phosphoglycerate kinase (*PGK*) promoter (e.g., SEQ ID NO: 114) and terminator (e.g., SEQ ID NO: 115) regions from *Candida* strain ATCC 20336 with an intervening multiple cloning site for insertion of ORFs. The *PGK* promoter region was amplified by PCR from genomic DNA using primers oAA0347 and oAA0348. The 1,029-bp DNA fragment containing the *PGK* promoter was then cut with restriction enzymes PstI/XbaI. The *PGK* terminator region was amplified by PCR from genomic DNA using primers oAA0351 and oAA0352. The 396-bp DNA fragment containing the *PGK* terminator was then cut with restriction enzymes XbaI/NdeI. The 3,522-bp PstI/NdeI DNA fragment from pAA061 was used in a three-piece ligation reaction with the *PGK* promoter and terminator regions to produce plasmid pAA105 (Fig. 12 and Fig. 13). The sequence between the *PGK* promoter and terminator contains restriction sites for incorporating ORFs to be controlled by the constitutive *PGK* promoter. One example, plasmid pAA219, was constructed for expression of cytochrome P450 reductase (*CPRB*). The *CPRB* gene (e.g., coding nucleotide SEQ ID NO: 90) was amplified by PCR from *Candida* strain ATCC 20336 genomic

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1613 of 3077

DNA with primers oAA0570 and oAA0571. The resulting 2,071-bp amplicon was cut with BspQI and ligated into BspQI-cut plasmid pAA105 resulting in the expression vector pAA219 (Fig. 14).

Construction of vector pAA073

Vector pAA073 was constructed from base vector pAA061 to include the acyl-CoA oxidase (*POX4*) promoter (e.g., SEQ ID NO: 117) and terminator (e.g., SEQ ID NO 116) regions from *Candida* strain ATCC 20336 with an intervening multiple cloning site for insertion of ORFs (Fig. 15). The *POX4* promoter region was amplified by PCR from genomic DNA using primers oAA0208 and oAA0209. The 568-bp DNA fragment containing the *POX4* promoter was then cut with restriction enzymes PstI/XmaI. The *POX4* terminator region was amplified by PCR from genomic DNA using primers oAA0216 and oAA0217. The 214-bp DNA fragment containing the *POX4* terminator was then cut with restriction enzymes XmaI/NdeI. The 3,522-bp PstI/NdeI DNA fragment from pAA061 was used in a three-piece ligation reaction with the *POX4* promoter and terminator regions to produce plasmid pAA073 (Fig. 16 and SEQ ID NO: 110). The sequence between the *POX4* promoter and terminator contains restriction sites (including SapI and BspQI) for incorporating ORFs to be controlled by the fatty acid-inducible *POX4* promoter. One example, plasmid pAA153, was constructed for expression of cytochrome P450 monooxygenase A14 (*CYP52A14*). The *CYP52A14* gene (see, e.g., international patent application no. PCT/US2012/045615 published as WO2013/006730) was amplified by PCR from *Candida* strain ATCC 20336 genomic DNA with primers oAA0519 and oAA0520. The resulting 1,613-bp amplicon was cut with BspQI and ligated into BspQI-cut plasmid pAA073 resulting in the expression vector pAA153 (Fig. 17). In another example, plasmid pAA218, was constructed for expression of cytochrome P450 reductase (*CPRB*). The *CPRB* gene was amplified by PCR from *Candida* strain ATCC 20336 genomic DNA with primers oAA0572 and oAA0573. The resulting amplicon was cut with BspQI and ligated into BspQI-cut plasmid pAA073 resulting in the expression vector pAA218.

Table 3 lists the names of plasmids with promoter and terminator combinations that were cloned into the shuttle vector pAA061. All promoters and terminators listed were cloned from *Candida* strain ATCC 20336.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1614 of 3077

Table 3: *Candida* strain ATCC 20336 gene promoter/terminator combinations cloned into shuttle vector pAA061		
Plasmid	**Promoter**	**Terminator**
pAA072	*FAO1*	*FAO1*
pAA073	*POX4*	*POX4*
pAA074	*POX5*	*POX5*
pAA105	*PGK*	*PGK*
pAA332	*TEF1*	*TEF1*
pAA334	*GPD*	*GPD*
pAA335	*PEX11*	*PEX11*
pAA1617	*HDE*	*PEX11*

EXAMPLE 2: Genetic engineering methods

Genomic DNA preparation

Genomic DNA from yeast strains such as *Candida* strain ATCC 20336 can be prepared as follows. About 1.5 ml of an overnight culture of cells is centrifuged and the pellet is resuspended in about 200 µl of a solution containing 2% Triton X-100, 1% SDS, 100 mM NaCl, 10 MM Tris pH 8.0, and 1 mM EDTA. About 200 µl of acid washed glass beads are added with about 200 µl of phenol:chloroform:isoamyl alcohol (25:24:1) at a pH of about 8.0. The sample is vortexed for about 2 minutes after which about 200 µl of water is added. The sample is then centrifuged at 13,000 rpm for about 10 minutes. The aqueous layer is transferred to a new microcentrifuge tube and an equal volume of chloroform:isoamyl alcohol (24:1) solution is added. This sample is vortexed for 10 seconds and then centrifuged at 13,000 rpm for about 2 minutes. The aqueous layer is transferred to a new microfuge tube and 1 ml of ethanol is added. The tube is then placed at -80 °C for about 15 minutes and then spun at 13,000 rpm for 15 minutes to pellet the DNA. The DNA is washed with 70% ethanol and air-dried. The DNA is then resuspended in about 500 µl of water.

To calculate gene copy number, a qPCR method can be used as described by Jin et a.l (*Appl. Environ. Microbiol.* January 2003 vol. 69, no. 1, 495-503). qPCR is performed according to the manufacturer's instructions using either the Brilliant III Ultra-Fast SYBR® Green QPCR Master Mix (Agilent Technologies, Englewood, CO USA) or the QuantiTect Multiplex PCR NoROX Kit

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1615 of 3077

(Qiagen). Genomic DNA from *Candida* strain ATCC 20336 or plasmid DNA containing the actin gene from ATCC 20336 and a gene of interest can be used as standards.

Primers and probes can be made via standard DNA synthesis techniques and be obtained, for example, from Integrated DNA Technologies (Coralville, IA, USA).

Amplification of gene copy number by single-crossover integration of linearized vectors

Linear DNA transformed into yeast can be integrated into the genome by homologous recombination. The localization of genomic integration is determined by the homologous sequence at the ends of the transformed DNA. In the case of vectors without a yeast autonomous replication sequence (ARS), integration may occur by so-called single-crossover integration of multiple tandem arrays. For example, as shown in Fig. 1, a nucleic acid, or gene, of interest (GOI) can be added to a host cell genome by recombination between homologous sequences in the transforming linear DNA and the host genome (sequences of the host cell nonfunctional *ura3* gene in the example shown in Fig. 1). A Ura^- auxotrophic mutant containing a non-functional gene for orotidine 5'-phosphate decarboxylase (*ura3*) may be "rescued" to the Ura^+ phenotype by the genomic integration of a functional *URA3* gene. A plasmid containing a nucleic acid of interest (GOI) to be added to the host genome can be linearized by cutting in the middle of a functional *URA3* gene in the plasmid and may integrate into the genome in multiple tandem arrays to generate a functional *URA3* gene. Integration of only one copy of such a plasmid by single-crossover integration may not always result in a functional *URA3* gene (Fig. 1, part A), however integration of tandem arrays ensures that a functional *URA3* gene is formed (Fig. 1, part B). Vector pAA153 linearized by cutting in the middle of the *URA3* ORF with *ClaI* is an example of a single-crossover vector (Fig. 18).

Construction of antibiotic-free single-crossover cassettes

When a vector is linearized for single-crossover integration, all elements of the vector are integrated into the genome of the transformed organism, including any genes for antibiotic resistance used for maintaining the plasmid in *E. coli*. To avoid the integration of genes encoding antibiotic resistance into yeast production strains, plasmids were designed to allow the PCR amplification of linear DNA cassettes for single-crossover integration that did not contain antibiotic resistance genes. One example of this is the rearrangement of plasmid pAA153 (Fig.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1616 of 3077

17) to produce the antibiotic-free cassette plasmid pAA367 (Fig. 19). Plasmid pAA153 was used as the template for PCR amplification with primer pairs oAA2206/oAA2207 and oAA2208/oAA2209 producing amplicons of 1,022 bp and 2,822 bp, respectively. The two amplicons contained overlapping sequence allowing overlap extension PCR with primer pairs oAA2206/oAA2209 producing a 3,804-bp amplicon that was cloned into pCR-BluntII-TOPO (Thermo Fisher Scientific) resulting in plasmid pAA367 (Fig. 19). An antibiotic-free linear DNA cassette was amplified by PCR using pAA367 as template with primer pair oAA2206/oAA2209 and gel purified before transformation to remove the plasmid template. The 3,804-bp antibiotic-free cassette (Fig. 20) contains all of the critical DNA elements of plasmid pAA153 without the antibiotic resistance gene. Other antibiotic-free single-crossover cassettes were generated using the same general method of splitting the *URA3* selectable marker on either side of the desired promoter, gene of interest, and terminator sequences.

Generation of a Ura$^-$ mutant of *Candida* strain ATCC 20962

Candida strain ATCC 20962 (*pox5::ura3/pox5::ura3 pox4a::ura3/pox4b:: URA3* is a beta-oxidation blocked (*pox4Δ, pox5Δ*) and Ura$^+$ derivative of *Candida* strain ATCC 20336. To reutilize the *URA3* marker for subsequent engineering, a Ura$^-$ derivative was generated and isolated via the following method. A single colony having the Ura$^+$ phenotype was inoculated into 3 mL YPD media (10 g/L Bacto Yeast Extract, 20 g/L Bacto Peptone, and 20 g/L dextrose) and grown overnight at 30°C with shaking. The overnight culture was then harvested by centrifugation and resuspended in 1 mL YNB+YE (6.7 g/L Yeast Nitrogen Broth, 3g/L Yeast Extract). The resuspension was then serially diluted in YNB+YE and 100 uL aliquots plated on YPD plates (incubation overnight at 30°C) to determine titer of the original suspension. Additionally, triplicate 100 uL aliquots of the undiluted suspension were plated on SC Dextrose (Bacto Agar 20g/L, Uracil 0.3 g/L, Dextrose 20 g/L, Yeast Nitrogen Broth 6.7 g/L, Amino Acid Dropout Mix 2.14 g/L) and 5-FOA at 3 different concentrations (0.5, 0.75, and 1 mg/mL). Plates were incubated for at least 5 days at 30°C. Colonies arising on the SC Dextrose + 5-FOA plates were picked into 50 uL sterile, distilled water and 5 uL struck out to YPD and SC–URA (SC Dextrose medium without Uracil). Colonies growing only on YPD and not on SC–URA plates were then inoculated into 3 mL YPD and grown overnight at 30°C with shaking. The overnight culture was then harvested by centrifugation and resuspended in 1.5 mL YNB (6.7 g/L Yeast Nitrogen Broth). The resuspension was then serially diluted in YNB and 100 uL aliquots plated on YPD plates (incubation overnight at 30°C) to determine initial titer. Also, for each

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1617 of 3077

undiluted suspension, 1 mL was plated on SC–URA and incubated for up to 7 days at 30°C. Colonies on the SC-URA plates were revertants, and the isolate with the lowest reversion frequency (< 10^{-7}) was designated sAA103 (*ura3*/*ura3 pox4*::*ura3*/*pox4*::*ura3 pox5*::*ura3*/*pox5*::*ura3*) and used for subsequent strain engineering.

Construction of double-crossover gene knock-out cassettes with recyclable *URA3* marker

Linear DNA transformed into yeast can also be integrated into the genome by double-crossover homologous recombination. Constructs for knocking out the function of a gene of interest (GOI) in *Candida* strain ATCC 20336 (or its descendants) were designed by placing a *URA3* selectable marker between 5' and 3' homologous sequences for the GOI (Fig. 21A). The *URA3* selectable marker was designed to have DNA sequence direct repeats at the beginning and at the end of the gene sequence that provide a mechanism for removing the *URA3* selectable marker from the genome by a second crossover homologous recombination event after integration of the linear DNA while retaining some of the linear DNA sequence for disruption of the GOI. This enables the transformed cell to subsequently receive additional transforming DNA using *URA3* as a selectable marker and is thus referred to as "recycling" of the *URA3* marker. The direct repeat can be the *URA3* promoter, the *URA3* terminator, or any other DNA sequence. After transformation of a double-crossover gene knock-out cassette into an Ura^- mutant, the *URA3* marker allows selection on SC-URA plates for colonies that have integrated the construct (Fig. 21B) disrupting the GOI and generating an Ura^+ phenotype. The Ura^+ transformant can be isolated and subsequently treated to remove the functional *URA3* gene so that the *URA3* marker can be reused for another round of gene knock-out. The Ura+ transformant (Fig. 21B) plated on media containing 5-fluoroorotic acid (5-FOA) will only grow by rendering the *URA3* gene non-functional since cells that have an intact *URA3* gene convert, through a series of reactions, 5-FOA to fluorodeoxyuridylate, which is toxic to the cells. The DNA sequence direct repeats provide a convenient means for the cell to "loop out" the *URA3* gene by homologous recombination (Fig. 21C), resulting in a Ura^- phenotype with the DNA sequence direct repeat as a "scar" left behind at the gene knock out site. The *URA3* selection marker may now be used again for further genetic modifications such as deletion of the second allele of a GOI (Fig. 22).

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1618 of 3077

Construction of double-crossover gene knock-in cassettes with recyclable *URA3* marker

As described in the previous example, linear DNA cassettes may integrate into the yeast genome by double crossover homologous recombination. The DNA cassettes from the previous example can be modified to knock out a first gene of interest, GOI1, while simultaneously integrating, or knocking in, a second gene of interest, GOI2 (Fig. 3A). The GOI2 may be placed under the control of any promoter and terminator desired, except for the URA3 promoter and terminator. After transformation of a Ura^- mutant, GOI1 is knocked out while GOI2 is knocked in and the *URA3* marker confers a Ura^+ phenotype (Fig. 3B). The *URA3* marker can be recycled as described previously by selection on media containing 5-FOA converting the phenotype back to Ura^- (Fig. 3C).

EXAMPLE 3: Modification of nucleic acids encoding carnitine acetyltransferase

In order to generate new cells and organisms with altered carbon flux for enhanced production of target molecules, nucleic acid constructs were developed to introduce desired genetic modifications into host cells. The following Examples describe the construction and use of exemplary nucleic acid constructs in making genetic modifications of cells and organisms. Carnitine O-acetyltransferase (e.g., EC 2.3.1.7), also referred to herein as carnitine acetyltransferase, is an enzyme that reversibly links acetyl units from acetyl-CoA to the carrier molecule carnitine. In some organisms, such as, for example, certain yeast species, the enzyme is dually targeted to mitochondria and peroxisomes by N-terminal and C-terminal targeting signals, respectively (see, e.g., Elgersma *et al.* (1995) *EMBO J.* 14: 3472-3479 and Kawachi *et al.* (1996) *Eur. J. Biochem.* 238: 845-852). An N-terminal sequence is referred to as the mitochondrial targeting signal (mts) and a C-terminal sequence is referred to as the peroxisomal targeting sequence (pts). Peroxisomal acetyl-CoA not destined for the glyoxylate cycle can be converted to acetyl-carnitine by carnitine O-acetyltransferase. Due to its smaller size compared to acetyl-CoA, acetyl-carnitine is able to diffuse through pores in the peroxisomal membrane, across the cytoplasm to mitochondria where it is converted back to acetyl-CoA by mitochondrial carnitine O-acetyltransferase (see, e.g., Strijbis *et al.* (2008) *Eukaryotic Cell* 7: 610-618 and Strijbis *et al.* (2010) *J. Biol. Chem.* 285: 24335-24346).

One method for increasing the amount of cytosolic acetyl-CoA in an organism is to modify the expression of carnitine O-acetyltransferase to provide increased carnitine O-acetyltransferase

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1619 of 3077

activity in the cytoplasm. In one embodiment, increased activity can be obtained by increasing the amount of carnitine O-acetyltransferase enzyme in the cytosol. For example, engineered carnitine O-acetyltransferase proteins lacking amino acid sequence targeting signals that direct the enzyme to one or more cellular locations (e.g., peroxisomes and mitochondria) other than the cytosol can be expressed in host cells thereby increasing the amount of carnitine O-acetyltransferase in the cytosol. Such engineered proteins will remain in the cytoplasm after being produced by the organism. A modified carnitine O-acetyltransferase protein ($Cat2p^{\Delta mts\Delta pts}$) lacking a mitochondrial targeting signal (mts) and a peroxisomal targeting signal (pts) converts acetyl-carnitine in transit from the peroxisome to the mitochondria into acetyl-CoA in the cytoplasm. Modified carnitine O-acetyltransferases can be produced upon expression of heterologous nucleic acids encoding the proteins that have been introduced into host cells.

Cloning of DNA encoding Cat2p from *Candida* strain ATCC 20336

DNA encoding carnitine O-acetyltransferase (Cat2p) from *Candida* strain ATCC 20336 was amplified by PCR from genomic DNA using primers oAA2586 and oAA2587. The PCR product containing the DNA encoding Cat2p was purified and ligated into pCR-BluntII-TOPO vector (Thermo Fisher Scientific), transformed into competent TOP10 *E. coli* cells (Thermo Fisher Scientific) and clones containing PCR inserts were sequenced to confirm the correct nucleic acid sequence. The resulting sequence-confirmed plasmid was named pAA426.

Construction of *CAT2* knock out cassettes

Two *CAT2* knock-out cassettes were constructed for disruption of the two *CAT2* gene alleles in *Candida* strain ATCC 20336 using homologous recombination methods known in the art and described herein. Each cassette contained a 5' *CAT2* DNA fragment and a 3' *CAT2* DNA fragment (referred to as gene homology fragments) to provide sequence used in a first single cross over homologous recombination event that results in the insertion of the entire cassette into an endogenous *CAT2* gene. Each cassette also contained a *URA3* selectable marker gene positioned between the homology regions in the cassette. The selectable marker included a direct repeat of sequence of the *URA3* gene promoter at the 3' end of the gene that facilitates a second cross over homologous recombination event that results in excision of most of the *URA3* selectable marker gene from the allele. However, a portion of the cassette remains in the *CAT2* gene allele, which thereby disrupts and effectively knocks out the function of the endogenous

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1620 of 3077

gene. The cassettes differed in the size of the *CAT2* gene homology regions and the orientation of the *URA3* selectable marker between the homology regions. Each deletion cassette was generated by combining three DNA fragments: a 5' *CAT2* DNA fragment (i.e., gene homology fragment), a 3' *CAT2* DNA fragment (i.e., gene homology fragment) and a *URA3* gene fragment originally cloned from *Candida* strain ATCC 20336 for use as a selectable marker. In constructing the first *CAT2* deletion cassette, a *CAT2* 5' homology DNA fragment was generated by amplifying a 5' region of the *CAT2* gene from plasmid pAA426 using primers oAA2372 and oAA2373. A *CAT2* 3' homology DNA fragment was generated by amplifying a 3' region of the *CAT2* gene from plasmid pAA426 using primers oAA2376 and oAA2377. A *URA3* selectable marker DNA fragment ($P_{URA3}URA3T_{URA3}P_{URA3}$; see, e.g., SEQ ID NO: 57) was amplified by PCR from plasmid pAA298 (Fig. 23) using primers oAA2374 and oAA2375. The three PCR products were purified and combined in the same reaction to generate the first full-length deletion cassette which was assembled by overlap extension PCR. The cassette was then ligated into pCR-BluntII-TOPO generating plasmid pAA1519 (Fig. 24). The linear *CAT2* deletion cassette could then be amplified from plasmid pAA1519 with primers oAA2372 and oAA2377.

In constructing the second *CAT2* deletion cassette, a *CAT2* 5' homology DNA fragment was generated by amplifying a 5' region of the *CAT2* gene from plasmid pAA426 using primers oAA2372 and oAA3312. A *CAT2* 3' homology DNA fragment was generated by amplifying a 3' region of the *CAT2* gene from plasmid pAA426 using primers oAA3315 and oAA2377. A *URA3* selectable marker DNA fragment was amplified from plasmid pAA298 using primers oAA3313 and oAA3314. The three PCR products were purified, and combined in the same reaction to generate the second full-length deletion cassette which was assembled by overlap extension PCR. The cassette was then ligated into pCR-Blunt II-TOPO generating plasmid pAA1520. The linear *CAT2* deletion cassette could then be amplified from plasmid pAA1520 (Fig. 24) with primers oAA2372/oAA2377.

Generation of nucleic acid constructs encoding modified carnitine acetyltransferase proteins

Engineered nucleic acids encoding carnitine O-acetyltransferase lacking targeting signals were generated as follows. A portion of the *CAT2* gene from *Candida* strain ATCC 20336 was PCR amplified from pAA426 using primers oAA4719 and oAA4720, generating a DNA fragment

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1621 of 3077

(*CAT2*$^{\Delta mts}$; SEQ ID NO: 60) encoding a truncated protein lacking the 18 N-terminal amino acids normally present in the wild-type enzyme (Cat2p$^{\Delta mts}$; Fig. 25A and Fig. 25B and SEQ ID NO: 3). A plasmid backbone was then amplified from plasmid pAA1164 (Fig. 26; SEQ ID NO: 112) using primers oAA4722 and oAA4723. This DNA fragment contained all the elements of the pCR-BluntII-TOPO vector, a split *URA3* selectable marker (i.e., a *URA3* DNA fragment that has been cleaved to yield a 5' URA3 fragment and a 3' *URA3* fragment (referred to as homology regions)) originally cloned from *Candida* strain ATCC 20336, and the *HDE1* gene promoter (SEQ ID NO: 113) and *POX4* gene terminator (SEQ ID NO: 116), both from *Candida* strain ATCC 20336. The nucleic acid products amplified from pAA426 and pAA1164 contained overlapping sequence at their termini allowing directional ligation of the products to generate plasmid pAA1610. Plasmid pAA1610 contains the nucleic acid encoding the modified Cat2p$^{\Delta mts}$ protein under the control of the *HDE1* gene promoter and *POX4* gene terminator positioned between the two separate fragments of the *URA3* gene marker.

The nucleic acid segment encoding the three C-terminal residues of the modified Cat2p$^{\Delta mts}$ protein was deleted from the CAT2$^{\Delta mts}$ DNA fragment by site-directed mutagenesis in order to remove the peroxisomal targeting sequence from the encoded protein. Plasmid pAA1610 was used as template and PCR amplified using primers oAA5319 and oAA5320 containing the desired mutation eliminating the C-terminal-encoding nucleic acids. The amplification reaction was digested with *Dpn*I to eliminate the template plasmid before transformation into *E. coli* DH5α cells. One sequence verified construct was saved and named plasmid pAA1667. Plasmid pAA1667 is identical to pAA1610, other than the removal of the codons for the last 3 amino acids (the peroxisomal targeting sequence) of the Cat2p$^{\Delta mts}$ enzyme, resulting in the formation of a construct containing a modified DNA encoding Cat2p$^{\Delta mts\Delta pts}$ (Fig. 25A and Fig. 25B; SEQ ID NO: 4). The modified nucleic acid encoding Cat2p$^{\Delta mts\Delta pts}$ was referred to as *CAT2*$^{\Delta mts\Delta pts}$ (SEQ ID NO: 61).

In order to generate a DNA fragment encoding a modified carnitine O-acetyltransferase protein containing a mitochondrial targeting sequence but lacking the peroxisomal targeting sequence (Cat2p$^{\Delta pts}$), genomic DNA from strain ATCC 20336 was PCR amplified using primers oAA6472 and oAA6473 to generate DNA fragment *CAT2*$^{\Delta pts}$ (see, e.g., SEQ ID NO: 62), which is a truncated *CAT2* DNA fragment missing codons for the peroxisomal targeting sequence (PTS). A plasmid backbone containing all the elements of the pCR-BluntII-TOPO vector, a split URA3 (*Candida* strain ATCC 20336) selectable marker, and the *HDE1* gene promoter and *POX4* gene

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1622 of 3077

terminator (both from *Candida* strain ATCC20336) was then amplified from plasmid pAA1164 (Fig. 26) using primers oAA6367 and oAA6368. The nucleic acid products amplified from the genomic DNA and pAA1164 contained overlapping sequence at their termini allowing directional ligation of the products to generate plasmid pAA1876. Plasmid pAA1876 contains the nucleic acid encoding the modified Cat2p$^{\Delta pts}$ protein under the control of the *HDE1* gene promoter and *POX4* gene terminator positioned between the two separate fragments of the *URA3* gene marker.

Cloning and modification of DNA encoding Yat1p from Candida strain ATCC 20336

Another method of increasing the amount or concentration of cytosolic acetyl-CoA in an organism is to modify the expression of carnitine acetyltransferase in the organism such that there is decreased carnitine acetyltransferase activity in cellular organelles. For example, by decreasing the activity level of carnitine acetyltransferase in the mitochondria, there is a corresponding decrease in mitochondrial conversion of acetyl-carnitine to acetyl-CoA. This can introduce a bottleneck in acetyl-carnitine processing in the mitochondria which can have the effect of diverting acetyl-carnitine from entering the mitochondria from the cytoplasm. Methods of decreasing carnitine acetyltransferase activity in cellular compartments, such as the mitochondria, of a *Candida* yeast strain include replacing the wild-type promoter of an endogenous *CAT2* gene in the yeast with a weaker heterologous promoter and/or replacing or modifying a gene encoding a wild-type carnitine acetyltransferase such that the encoded modified or substituted carnitine acetyltransferase protein has a reduced enzyme activity.

The *YAT1* gene from *Candida* strain ATCC 20336 encodes a cytoplasmic-targeted enzyme with a reduced carnitine acetyltransferase activity relative to the activity of the enzyme encoded by the endogenous CAT2 gene. DNA encoding the YAT1p protein (SEQ ID NO: 6) was amplified from genomic DNA using primers oAA9946 and oAA9947 and the PCR product was purified and the sequence was verified. A plasmid backbone containing all the elements of the pCR-BluntII-TOPO vector, a split *URA3* (from *Candida* strain ATCC 20336) selectable marker, and the *HDE1* gene promoter and *POX4* gene terminator (both from *Candida* strain ATCC20336) was then amplified from plasmid pAA1164 (Fig. 26) using primers oAA4722 and oAA4723. The nucleic acid products amplified from *Candida* strain ATCC 20336 genomic DNA and pAA1164 contained overlapping sequence at their termini allowing directional ligation of the products to generate plasmid pAA1817. Plasmid pAA1817 contains the nucleic acid (see, e.g., SEQ ID NO:

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1623 of 3077

63) encoding the YAT1 protein under the control of the *HDE1* gene promoter and *POX4* gene terminator positioned between the two separate fragments of the *URA3* gene marker.

The Yat1p enzyme with carnitine acetyl transferase activity is normally targeted to the cytoplasm in *Candida* strain ATCC 20336. Modification of Yat1p to target the enzyme to the mitochondrial compartment of a yeast cell involved the addition of amino acids at the N-terminus of the wild-type protein to act as a mitochondrial targeting signal (MTS). The DNA sequences encoding the MTS for three genes encoding mitochondrial-targeted enzymes in *Candida* strain ATCC 20336 were cloned separately and added to the 5' end of the *YAT1* gene coding sequence providing DNA encoding the required N-terminal amino acids for a functional MTS. DNA encoding the predicted MTS from the Cox4p (SEQ ID NO: 7), Cit1p (SEQ ID NO: 8) and Cat2p (SEQ ID NO: 9) proteins was added to the 5' end of the coding sequence of the *YAT1* gene such that the resulting encoded protein contained the heterologous MTS in place of the native Yat1p initiating methionine (Fig. 27). Plasmid pAA1817 containing the DNA encoding the YAT1 protein was PCR amplified with primers oAA6771 and oAA6669 to yield an amplification product of the entire plasmid backbone and *YAT1* gene without the codon for the initiating methionine. The DNA sequences encoding the Cox4p MTS (SEQ ID NO: 64), Cit1p MTS (SEQ ID NO: 65) and Cat2p MTS (SEQ ID NO: 66) sequences were amplified from genomic DNA of *Candida* strain ATCC 20336 using primer pairs oAA6772/oAA6773, oAA6774/oAA6775, and oAA6776/oAA6777, respectively. The PCR products containing the modified YAT1p-encoding DNA with plasmid backbone, and the MTS-encoding sequences were designed with overlapping sequence at their termini allowing directional ligation, generating plasmids pAA1967 (containing $YAT1^{+COX4mts}$; SEQ ID NO: 67), pAA1968 (containing $YAT1^{+CIT1mts}$; SEQ ID NO: 68) and pAA1969 (containing $YAT1^{+CAT2mts}$; SEQ ID NO: 69). The modified YAT1p+mts-encoding DNA of each of plasmids pAA1967, pAA1968 and pAA1969 is under the control of the *HDE1* gene promoter and *POX4* gene terminator positioned between the two separate fragments of the *URA3* gene marker.

EXAMPLE 4: Modification of nucleic acids encoding thioesterase proteins

One approach to altering cellular carbon flux to recapture cytoplasmic acetyl-CoA that would have been lost to growth and energy generation processes and divert it to use in target molecule production is through an engineered acetyl group carbon recycle loop as described herein. In an exemplary recycle loop, acetyl moieties generated in the breakdown of fatty acids

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1624 of 3077

in peroxisomal β-oxidation are diverted from their usual mitochondrial destination and to cytosolic fatty acid synthesis to regenerate a fatty acid that can be subjected to another cycle of peroxisomal β-oxidation. The cytosolic fatty acid that is synthesized is in the form of an acyl-CoA that can be utilized in non-target molecule processes, such as lipid synthesis, by the cell. To divert the synthesized acyl-CoA away from such processes and toward ω- and β-oxidation, it can be converted to a free fatty acid through the action of a thioesterase (e.g., EC 3.1.2.20) enzyme which catalyzes the release of coenzyme A from long-chain acyl-CoA. In yeast, thioesterases (TES) are present in the peroxisomal compartment of the cells to ensure that free coenzyme A is available for β-oxidation. In one method of producing free fatty acids in the cytoplasm of yeast, a peroxisomal thioesterase with activity on long chain fatty acids is heterologously expressed in the cytoplasm. In one embodiment, this can be accomplished by modifying nucleic acids encoding a yeast thioesterase to delete the portion of the nucleic acid encoding the peroxisomal targeting signal at the C-terminus of the protein.

Cloning DNA encoding Tesp proteins from *Candida* strain ATCC 20336

Candida strain ATCC 20336 contains eight genes encoding peroxisomal thioesterases. DNA encoding proteins encoded by each gene was amplified by PCR from genomic DNA using the primers indicated in Table 4. The PCR products were purified and cloned into pCR-BluntII-TOPO. Sequence confirmed plasmids were named as listed in Table 4.

Table 4: Primers and plasmids used in amplification cloning of DNA encoding *Candida* thioesterase proteins*

Thioesterase	Primers	Plasmid
TES1-1 (nucleotide SEQ ID NO: 78; amino acid SEQ ID NO: 22)	oAA3582 oAA3583	pAA895
TES2-1 (nucleotide SEQ ID NO: 79; amino acid SEQ ID NO: 23)	oAA3584 oAA3585	pAA896
TES2-2 (nucleotide SEQ ID NO: 80; amino acid SEQ ID NO: 24)	oAA3584 oAA3585	pAA897
TES3-1 (nucleotide SEQ ID NO: 81; amino acid SEQ ID NO: 25)	oAA3586 oAA3587	pAA898
TES4-1 (nucleotide SEQ	oAA3588	pAA899

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1625 of 3077

Table 4: Primers and plasmids used in amplification cloning of DNA encoding *Candida* thioesterase proteins*

Thioesterase	Primers	Plasmid
ID NO: 82; amino acid SEQ ID NO: 26)	oAA3589	
TES5-1 (nucleotide SEQ ID NO: 83; amino acid SEQ ID NO: 27)	oAA3590 oAA3591	pAA900
TES6-1 (nucleotide SEQ ID NO: 84; amino acid SEQ ID NO: 28)	oAA3592 oAA3593	pAA902
TES7-1 (nucleotide SEQ ID NO: 85; amino acid SEQ ID NO: 29)	oAA3594 oAA3595	pAA903
TES7-2 (nucleotide SEQ ID NO: 86; amino acid SEQ ID NO: 30)	oAA3594 oAA3595	pAA908
TES8-1 (nucleotide SEQ ID NO: 87; amino acid SEQ ID NO: 31)	oAA3596 oAA3597	pAA904

*the "-1" and "-2" designations for the thioesterases indicate which member of the sequence-confirmed allelic pair is contained in the plasmid

Construction of a *TES3*$^{\Delta pts}$ mutant

A mutant of Tes3p lacking a peroxisomal targeting sequence was constructed to retarget the Tes3p enzyme to the cytoplasm. DNA encoding Tes3p lacking the C-terminal 3 residues (SEQ ID NO: 88) was amplified from plasmid pAA898 with primers oAA6449 and oAA6450. A plasmid backbone containing all the elements of the pCR-BluntII-TOPO vector, a split URA3 selectable marker, and the *HD1E* gene promoter and *POX4* gene terminator was then amplified from plasmid pAA1164 (Fig. 26) using primers oAA6367 and oAA6368. The nucleic acid products amplified from plasmids pAA898 and pAA1164 contained overlapping sequence at their termini allowing directional ligation of the products to generate plasmid pAA1609. Plasmid pAA1609 contains a nucleic acid encoding the Tes3p$^{\Delta pts}$ protein (SEQ ID NO: 32) under the control of the *HDE1* gene promoter and *POX4* gene terminator situated between the two separate fragments of the *URA3* gene marker.

Modified forms of the proteins encoded by the other TES genes lacking a peroxisomal targeting sequence can be constructed as described for the construction of the DNA encoding Tes3p$^{\Delta pts}$.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1626 of 3077

First, DNA encoding thioesterases lacking the C-terminal 3 residues is PCR amplified from each of pAA895, pAA896, pAA897, pAA899, pAA900, pAA902, pAA908 and pAA904. Then, to obtain a plasmid backbone containing all the elements of the pCR-BluntII-TOPO vector, a split URA3 selectable marker, the *HD1E* gene promoter and the *POX4* gene terminator, plasmid pAA1164 is amplified using primers oAA6367 and oAA6368. The final plasmid for each Tes protein-encoding DNA can then be obtained by ligating the amplicons with the backbone amplified plasmid pAA1164 through directional ligation.

EXAMPLE 5: Modification of nucleic acids encoding acetyl-CoA carboxylase proteins

Cytosolic acetyl-CoA can be converted to malonyl-CoA by the enzyme acetyl-CoA carboxylase (EC 6.4.1.2). This biochemical step represents a committed step in fatty acid biosynthesis and can be highly regulated. One approach to increasing fatty acid biosynthesis in an organism is to increase the acetyl-CoA carboxylase (ACC1p) activity in the cytosol. Generally, ACC1p is regulated by feedback inhibition of acyl-CoA and also by phosphorylation. Because the dephosphorylated state is the active state of the enzyme, one approach for increasing ACC1p activity is to reduce or eliminate phosphorylation of the protein. In one embodiment, the endogenous *ACC1* coding sequence is modified to encode an enzyme in which the phosphorylatable serine residues have been substituted with alanine residues thereby relieving the regulation by phosphorylation and dephosphorylation.

Cloning DNA encoding Acc1p from *Candida* strain ATCC 20336

Primers oAA0784 and oAA0785 were used to amplify DNA encoding Acc1p from genomic DNA of *Candida* strain ATCC 20336, and the resulting 6869-bp amplicon was cloned into pCR-BluntII-TOPO. The sequence-verified plasmid was named pAA245 (Fig. 28). The DNA sequence cloned into plasmid pAA245 included the entire open reading frame of the *ACC1* gene and the partial sequence for an intron at the 5' end of the *ACC1* coding sequence.

Mutation of *ACC1* to replace serine-encoding with alanine-encoding sequences

An *ACC1* gene fragment and 5' partial intron (see, e.g., SEQ ID NO: 75) were restriction cloned from pAA245 by cutting with BspQI and ligated into BspQI-cut plasmid pAA105 (Fig. 13) to put the gene under the control of the *PGK* promoter and terminator (SEQ ID NOS: 114 and 115,

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1627 of 3077

respectively) from *Candida* strain ATCC 20336. The resulting sequence-confirmed plasmid was named pAA326 (Fig. 29). A truncated version of the *ACC1* gene encoding R643 to the stop codon was generated by cutting pAA326 with SpeI/XbaI and was cloned into plasmid pAA061 (Fig. 12) cut with the same enzymes. The resulting plasmid was named pAA1634 (Fig. 30) and provided a smaller plasmid template for performing site-directed mutagenesis. Site-directed mutagenesis was performed on pAA1634 using the primers shown in Table 5 to introduce the mutations listed in the Table and resulting in plasmids pAA1647, pAA1648, pAA1649, pAA1650, and pAA1900.

In order to generate mutant *ACC1* gene fragments containing the entire protein coding sequence of nucleic acids, PCR products containing the *HDE1* promoter from *Candida* strain ATCC 20336 and the 5'-end 1,926 bp of the *ACC1* gene encoding M1 to S642 were cloned into each of the plasmids containing a truncated mutant *ACC1* fragment resulting in plasmids pAA1906-pAA1909, and pAA1915. Additionally, plasmid pAA1910 was generated in the same manner, although the *ACC1* gene in that plasmid did not contain any mutations. The S652A mutation was introduced into all of plasmids pAA1906-pAA1910 and pAA1915 by site-directed mutagenesis with primers oAA6691 and oAA6692 generating plasmids pAA1957-pAA1962.

Table 5: Primers used in generating nucleic acid sequences encoding ACC1p mutant proteins and plasmids containing the nucleic acids

Truncated Plasmid	Mutation(s)	Primers	Full-Length Plasmid	IGR5 Integration Plasmid
pAA1634	None	None	pAA1910	pAA2249
pAA1647	S1138A	oAA5668 oAA5669	pAA1906	pAA2245
pAA1648	S1153A	oAA5670 oAA5671	pAA1907	pAA2246
pAA1649	S1158A	oAA5672 oAA5673	pAA1908	pAA2247
pAA1650	S1153A, S1158A	oAA5690 oAA5691	pAA1909	pAA2248
pAA1900	S1138A, S1153A, S1158A	oAA6145 oAA6146	pAA1915	pAA2250
	S652A, S1138A	oAA6691 oAA6692	pAA1957	pAA2251
	S652A,	oAA6691	pAA1958	pAA2252

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1628 of 3077

Table 5: Primers used in generating nucleic acid sequences encoding ACC1p mutant proteins and plasmids containing the nucleic acids				
Truncated Plasmid	**Mutation(s)**	**Primers**	**Full-Length Plasmid**	**IGR5 Integration Plasmid**
	S1153A	oAA6692		
	S652A, S1158A	oAA6691 oAA6692	pAA1959	pAA2253
	S652A, S1153A, S1158A	oAA6691 oAA6692	pAA1960	pAA2254
	S652A	oAA6691 oAA6692	pAA1961	pAA2255
	S652A, S1138A, S1153A, S1158A	oAA6691 oAA6692	pAA1962	pAA2256

Construction of double crossover integration cassettes for *ACC1* mutants

The *ACC1* mutant genes under the control of the *HDE1* promoter and *PGK* terminator (P_{HDE}-*ACC1*-T_{PGK}) were PCR amplified from plasmids pAA1906-pAA1909, pAA1915, and pAA1957-pAA1962, as well as the unmodified *ACC1* gene from pAA1910, with primers oAA7257 and oAA7258. The PCR products were cut with SbfI/MluI and ligated into plasmid pAA2153 cut with the same restriction enzymes. Plasmid pAA2153 contains DNA encoding a *URA3* selectable marker with a direct repeat of the URA3 terminator sequence (T_{URA3}) sequence, i.e., T_{URA3}-P_{URA3}-*URA3*-T_{URA3} (see, e.g., SEQ ID NO: 58). The *URA3* selectable marker in pAA2153 is placed between genomic DNA sequence elements from *Candida* strain ATCC 20336 which are named IGR5. The IGR5-5' (SEQ ID NO: 125; 446 bp) and IGR5-3' (SEQ ID NO: 126; 500 bp) homology regions target the integration of the intervening DNA into genomic DNA by homologous recombination. Plasmids constructed for the integration of *ACC1* mutants at IGR5 were named pAA2245 – pAA2256 (see, e.g., Fig. 31). Linear DNA used for transformation was generated by PCR with primers oAA7259 and oAA7260 using plasmids pAA2245 – pAA2256 as template DNA.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1629 of 3077

EXAMPLE 6: Modification of nucleic acids encoding acetyl-carnitine transporter proteins

Another approach to increasing the amount of cytosolic acetyl-CoA in an organism is to retain acetyl-carnitine within the cytoplasm that would otherwise be transported out of the cytoplasm and into organelles, such as the mitochondria. One method for increasing the amount of acetyl-carnitine within the cytoplasm is to reduce the transport of acetyl-carnitine into the mitochondria or other organelles. Mitochondria possess two membranes with the outer membrane generally allowing free diffusion of metabolites and the inner membrane controlling metabolite transport with multiple membrane transport proteins. One mitochondrial inner-membrane transport protein (Crc1p) may function as an acetyl-carnitine transporter providing for transport of acetyl-carnitine into the mitochondria (see, e.g., van Roermund *et al*. (1999) *EMBO J*. 18: 5843-5852 and Palmieri *et al*. (1999) *FEBS Lett.* 462: 472-476). Without being bound or limited by theory, reducing the amount of this acetyl-carnitine transporter in the mitochondrial membrane can create a bottleneck in the transport of acetyl-carnitine into the mitochondria resulting in accumulation of acetyl-carnitine in the cytoplasm. The cytoplasmic acetyl-carnitine can then be converted to acetyl-CoA by a cytoplasmic carnitine O-acetyltransferase (e.g.,Cat2p$^{\Delta mts\Delta pts}$) as described herein. Reduction in the amount of Crc1p can be achieved, for example, by modifying the promoter controlling expression of DNA encoding Crc1p in a cell through introduction of heterologous nucleic acids encoding the protein into host cells that do not express a functional Crc1p transporter.

Cloning DNA encoding Crc1p from *Candida* strain ATCC 20336

The amount of CRC1p transporter expressed in the mitochondrial membrane in *Candida* host cells can be reduced by disrupting or deleting the endogenous *CRC1* gene and replacing it with a heterologous transporter-encoding construct under the control of a promoter that is weaker than the native *CRC1* promoter as described herein. DNA encoding a mitochondrial inner membrane acetyl-carnitine transport protein Crc1p (amino acid SEQ ID NO: 14; nucleotide SEQ ID NO: 71) was amplified from genomic DNA of *Candida* strain ATCC 20336 with primers oAA5511 and oAA5512. The amplified DNA was gel purified and cloned into the pCR-Blunt II-TOPO vector. Plasmids isolated from transformants were sequenced and one sequence-verified plasmid was saved as plasmid pAA1564.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1630 of 3077

Construction of *CRC1* knock-out cassettes

CRC1 gene knock-out cassettes were constructed for use in generating host cells that could then be transformed with heterologous nucleic acids for modified expression of Crc1p. Two *CRC1* knock-out cassettes were constructed for disruption of the two endogenous *CRC1* gene alleles in *Candida* strain ATCC 20336 using homologous recombination methods known in the art and described herein. Each cassette contained a 5' *CRC1* DNA fragment and a 3' *CRC1* DNA fragment (referred to as gene homology fragments) to provide sequence used in a first crossover homologous recombination event that results in integration of the entire cassette into an endogenous *CRC1* gene. Each cassette also contained a *URA3* selectable marker gene positioned between the homology regions in the cassette. The selectable marker included a direct repeat of sequence of the *URA3* gene promoter at the 3' end of the gene that facilitates a second cross over homologous recombination event that results in excision of most of the URA3 gene from the allele. However, a portion of the cassette remains in the *CRC1* gene which thereby disrupts and effectively knocks out the function of the endogenous gene. The cassettes differed in the sizes of the *CRC1* gene homology regions and the orientation of the *URA3* selectable marker between the homology regions. Each deletion cassette was generated by combining three DNA fragments: a 5' *CRC1* DNA fragment (i.e., gene homology fragment), a 3' *CRC1* DNA fragment (i.e., gene homology fragment) and a *URA3* selectable marker.

In constructing the first *CRC1* deletion cassette, a *CRC1* 5' homology DNA fragment was generated by amplifying a 5' region of the *CRC1* gene from plasmid pAA1564 using primers oAA5511 and oAA5553. A *CRC1* 3' homology DNA fragment was generated by amplifying a 3' region of the *CRC1* gene from plasmid pAA1564 using primers oAA5512 and oAA5554. A *URA3* selectable marker DNA fragment ($P_{URA3}URA3T_{URA3}P_{URA3}$) was amplified by PCR from plasmid pAA298 (Fig. 23) using primers oAA5555 and oAA5556. The three PCR products were purified and combined in the same reaction to generate the first full-length deletion cassette which was assembled by overlap extension PCR. The cassette was then ligated into pCR-BluntII-TOPO generating plasmid pAA1613 (Fig. 32). The linear *CRC1* deletion cassette could then be amplified from plasmid pAA1613 with primers oAA5511 and oAA5512.

A second *CRC1* deletion cassette was constructed by inserting the *URA3* selectable marker ($P_{URA3}URA3T_{URA3}P_{URA3}$) at a different location and in the opposite orientation as was done for plasmid pAA1613. A first PCR product was generated using plasmid pAA1564 as template and

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1631 of 3077

primers oAA5698 and oAA5699. This product contained 5' *CRC1* and 3' *CRC1* DNA fragments and all the elements of the pCR-Blunt II-TOPO vector. A second PCR product containing a *URA3* selectable marker ($P_{URA3}URA3T_{URA3}P_{URA3}$) was generated using plasmid pAA298 as template and primers oAA5700 and oAA5701. The two PCR products contained overlapping sequence at their termini allowing directional ligation, generating plasmid pAA1701 (Fig. 32). The linear *CRC1* deletion cassette could then be amplified from plasmid pAA1701 with primers oAA5511 and oAA5512.

Construction of modified *CRC1* genes with a heterologous promoter

The promoter for the glucose-6-phosphate isomerase gene (*G6PI*) from *Candida* strain ATCC 20336 drives low level expression of the *G6PI* gene. The *G6PI* promoter (SEQ ID NO: 118) was amplified from genomic DNA with primers oAA7409 and oAA7410. The 745-bp amplified DNA fragment was gel purified and cloned into the pCR-Blunt II-TOPO vector. Plasmids isolated from transformants were sequenced, and one sequence-verified plasmid was saved as plasmid pAA2218.

A DNA fragment containing the *G6PI* promoter, *CRC1* coding region, and *POX4* terminator (P_{G6PI}-*CRC1*-T_{POX4}) from *Candida* strain ATCC 20336 was constructed by overlap extension PCR. The *G6PI* promoter was amplified from plasmid pAA2218 with primers oAA7403 and oAA7404, the *CRC1* gene was amplified from plasmid pAA1564 using primers oAA7381 and oAA7382, and the *POX4* terminator was amplified from plasmid pAA073 using primers oAA7379 and oAA7380. The three amplicons were used in a subsequent overlap extension PCR reaction with primers oAA7379 and oAA7403 to build the desired DNA fragment, P_{G6PI}-*CRC1*-T_{POX4}. A plasmid backbone containing the IGR5 homology regions and *URA3* selectable marker (with T_{URA3} repeat) was amplified from plasmid pAA2247 (Fig. 31) using primers oAA7377 and oAA7378. The resulting 6,223-bp amplified plasmid backbone and the 1,816-bp DNA fragment P_{G6PI}-*CRC1*-T_{POX4} encoded overlapping sequence allowing directional ligation, resulting in plasmid pAA2214 SEQ ID NO: 122)6. A linear double crossover DNA cassette for genomic integration could then be amplified from plasmid pAA2214 using primers oAA7259 and oAA7260 (Fig. 33).

A plasmid containing an antibiotic-free single-crossover cassette encoding the *CRC1* gene driven by the low expression *G6PI* promoter was also constructed. A DNA fragment encoding

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1632 of 3077

the elements P_{G6PI}-*CRC1* was amplified from plasmid pAA2214 using primers oAA7624 and oAA7625. A plasmid backbone encoding all of the elements of the pCR-BluntII-TOPO, a split *URA3* marker, and *POX4* terminator was then amplified using primers oAA7265 and oAA7512 from plasmid pAA1164 (Fig. 26; SEQ ID NO: 112). In certain instances, a plasmid backbone was amplified from plasmid pAA1116 (SEQ ID NO: 123). The two PCR products encoded overlapping sequence at their termini allowing directional ligation placing the *CRC1* gene under the control of the *G6PI* promoter and *POX4* terminator, generating plasmid pAA2311 (Fig. 34; SEQ ID NO: 124). A 3,307-bp antibiotic-free linear DNA cassette was amplified by PCR using pAA2311 as template with primer pair oAA2206/oAA2209 and gel purified before transformation to remove the plasmid template.

EXAMPLE 7: Modification of nucleic acids encoding acetyl-CoA hydrolase proteins

Another approach to increasing the amount of cytosolic acetyl-CoA in an organism is to facilitate the generation and transport of acetate from acetyl-CoA-containing organelles, such as peroxisomes, into the cytoplasm. Acetyl-CoA produced via beta-oxidation in the peroxisome may be converted to acetate by the action of acetyl-CoA hydrolase (e.g. EC 3.1.2.1). This enzyme is typically not localized to the peroxisomal compartment. In order to introduce acetyl-CoA hydrolase into peroxisomes, a heterologous nucleic acid construct was developed in which the enzyme coding sequence was modified to include DNA encoding a peroxisomal targeting signal at the C-terminus of the hydrolase and exclude a mitochondrial targeting sequence at the N-terminus. The acetate produced freely diffuses out of the peroxisome into the cytoplasm where it can be converted back to acetyl-CoA by the enzyme acetyl-CoA synthetase. The expression of acetyl-CoA synthetase can also be amplified by increasing the genomic copy number and by promoter replacement. The acetyl-CoA generated from acetate in the cytosol is not able to enter the mitochondria, and thus is not subject to loss due to use in mitochondrial metabolic pathways such as the TCA cycle, unlike acetyl groups in acetyl-carnitine which can be taken up by mitochondria. To decrease peroxisomal generation of acetyl-carnitine, the carnitine acetyltransferase gene (e.g., *CAT2* in *Candida*) can be disrupted in cells that have been modified to produce peroxisomal acetate using, for example, *CAT2* knock-out cassettes as described herein. To provide mitochondria in such *CAT2*-disrupted cells with acetyl-CoA for use in cellular energy production for cell viability, the cell can be transformed with nucleic acid encoding a carnitine acetyltransferase (e.g., a reduced activity version of the enzyme) targeted

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1633 of 3077

to mitochondria (e.g., *YAT1*$^{+mts}$) to restore a minimum amount of acetyl-CoA generation in the mitochondria.

The enzyme encoded by the *ACH1* gene has acetyl-CoA hydrolase activity and is targeted to the mitochondria in *Candida* strain ATCC 20336. DNA encoding Ach1p (amino acid SEQ ID NO: 15 and nucleotide SEQ ID NO: 72) was amplified from genomic DNA of *Candida* strain ATCC 20336 by PCR using primers oAA6369 and oAA6370. The primers were designed to amplify a portion of the *ACH1* gene such that the amplification product, *ACH1*$^{\Delta mts+pts}$, (1) lacked nucleic acid sequence encoding the N-terminal 11 amino acids that form the mitochondrial targeting sequence, and (2) included a nucleic acid sequence encoding a peroxisomal targeting sequence (Gly-Arg-Arg-Ala-Lys-Leu) at the C-terminus. A plasmid backbone containing all the elements of the pCR-BluntII-TOPO vector, a 5' *URA3* homology region and a 3' *URA3* homology region selectable marker, and the *HDE* gene promoter and *POX4* gene terminator was then amplified from plasmid pAA1164 (Fig. 26) using primers oAA6367 and oAA6368. The nucleic acid products amplified from *Candida* strain ATCC 20336 and pAA1164 contained overlapping sequence at their termini allowing directional ligation of the products to generate plasmid pAA1846. Plasmid pAA1846 contains the nucleic acid encoding the ACH1$^{\Delta mts+pts}$ protein (amino acid SEQ ID NO: 16 and nucleotide SEQ ID NO: 73) under the control of the *HDE1* gene promoter and *POX4* gene terminator situated between the two separate fragments of the *URA3* gene marker.

EXAMPLE 8: Modification of nucleic acids encoding acetyl-CoA synthetase proteins

Another approach to increasing the amount of cytosolic acetyl-CoA in an organism is to increase the acetyl-CoA synthetase activity in the cytoplasm. Acetyl-CoA synthetase (e.g., EC 6.2.1.1; also referred to as acetate-CoA ligase) can catalyze the conversion of acetate to acetyl-CoA. Thus, for example, acetate that diffuses into the cytoplasm after being produced in the peroxisome by the action of heterologous peroxisomal acetyl-CoA hydrolase, as described herein, can be converted to acetyl-CoA by the enzyme acetyl-CoA synthetase. One method for increasing the acetyl-CoA synthetase activity in the cytoplasm is to increase the expression of acetyl-CoA synthetase by, for example, increasing the genomic copy number of acetyl-CoA synthetase-encoding DNA (e.g., *ACS*1 and *ACS*2 genes) and/or by replacing the endogenous *ACS*1 and/or *ACS*2 promoter(s) with a stronger, high-expression heterologous promoter (e.g.,

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1634 of 3077

the fatty acid-inducible hydratase-dehydrogenase-epimerase (*HDE*) gene promoter from *Candida*).

Cloning DNA encoding Acs1p from *Candida* strain ATCC 20336

DNA encoding an acetyl-CoA synthetase (acetate-CoA ligase; Acs1p) from *Candida* strain ATCC 20336 was amplified from genomic DNA using primers oAA6371 and oAA6372. A plasmid backbone containing all the elements of the pCR-BluntII-TOPO vector, a 5' *URA3* and a 3' *URA3* homology region selectable marker, and the *HDE* gene promoter and *POX4* gene terminator from *Candida* strain ATCC 20336 was then amplified from plasmid pAA1164 (Fig. 26) using primers oAA6367 and oAA6368. The nucleic acid products amplified from *Candida* strain ATCC 20336 and pAA1164 contained overlapping sequence at their termini allowing directional ligation of the products to generate plasmid pAA1847. Plasmid pAA1847 contains a nucleic acid encoding the ACS1 protein (amino acid SEQ ID NO: 20 and nucleotide SEQ ID NO: 76) under the control of the *HDE1* gene promoter and *POX4* gene terminator situated between the two separate fragments of the *URA3* gene marker.

Cloning DNA encoding Acs2p from *Candida* strain ATCC 20336

DNA encoding an acetyl-CoA synthetase (acetate-CoA ligase; Acs2p) from *Candida* strain ATCC 20336 was amplified from genomic DNA using primers oAA6470 and oAA6471. A plasmid backbone containing all the elements of the pCR-BluntII-TOPO vector, a 5' *URA3* and a 3' *URA3* homology region selectable marker, and the *HDE1* gene promoter and *POX4* gene terminator from *Candida* strain ATCC20336 was then amplified from plasmid pAA1164 (Fig. 26) using primers oAA6367 and oAA6368. The nucleic acid products amplified from *Candida* strain ATCC 20336 and pAA1164 contained overlapping sequence at their termini allowing directional ligation of the products to generate plasmid pAA1875. Plasmid pAA1875 contains a nucleic acid encoding the ACS2 protein (amino acid SEQ ID NO: 21 and nucleotide SEQ ID NO: 77) under the control of the *HDE1* gene promoter and *POX4* gene terminator situated between the two separate fragments of the *URA3* gene marker.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1635 of 3077

EXAMPLE 9: Modification of nucleic acids encoding ATP citrate lyase proteins

Another approach to increasing cytosolic acetyl-CoA is to increase the activity of ATP citrate lyase (e.g., EC 2.3.3.8) in the cytoplasm of host cells. This enzyme can catalyze the conversion of citrate to acetyl-CoA and oxaloacetate. For example, in oleaginous yeast, ATP citrate lyase typically is a heterodimeric enzyme (Acl1p/Acl2p) that converts citric acid from the TCA cycle that overflows into the cytoplasm into acetyl-CoA and oxaloacetate at the expense of an ATP. The acetyl-CoA can then used for the production of fatty acids and eventually triacylglycerides as a carbon-storage mechanism. One method of increasing ATP citrate lyase activity in the cytosol of host cells is to increase the amount of the enzyme in the cytosol. This can be accomplished in a number of ways. For example, heterologous DNA encoding Acl1p and Acl2p can be introduced into host cells and/or the regulation of expression of Acl1p and Acl2p in a cell can be modified.

Construction of synthetic nucleic acids encoding ATP citrate lyase subunits

The amino acid sequences from NCBI accession numbers CAG80394 and XP_503231 for Acl1p (SEQ ID NO: 42) and Acl2p (SEQ ID NO: 43), respectively, from *Yarrowia lipolytica* CLIB122 were used to design DNA sequences (SEQ ID NOS: 96 and 97) encoding each enzyme subunit that were codon optimized for expression in *Candida* strain ATCC 20336. DNA fragments 500 bp in length with 50 bp overlaps (synthetic Gene blocks from Integrated DNA Technologies) were amplified in a PCR reaction without primers: 95°C for 3 minutes for 1 cycle, 95°C for 30 seconds, 60°C for 30 seconds, 72°C for 30 seconds for a total of 15 cycles, and final extension at 72°C for 2 minutes. 5µL of the PCR product was used as template for TouchDown PCR reaction (see, for example, Don *et al*. (1991) *Nucleic Acids Res.* 19:4008) including the primers: 95°C for 3 minutes for 1 cycle, 95°C for 30 seconds, 68°C for 30 seconds, decreasing by 0.4°C per cycle, 72°C for 30 seconds for a total of 30 cycles, and final extension at 72°C for 2 minutes. The primers used, oAA3051 and oAA3052 for *ACL1* and oAA3053 and oAA3052 for *ACL2*, incorporated *Bsp*QI restriction sites outside of the coding sequence. The PCR products were purified and cloned into pCR-BluntII-TOPO. Sequence-confirmed plasmids were named pAA709 (*ACL1*) and pAA710 (*ACL2*).

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1636 of 3077

Construction of expression cassettes for synthetic ACL-encoding nucleic acids

Synthetic DNA expression constructs encoding Acl1p and Acl2p were generated using different promoter and terminator elements contained in pAA073 (Fig. 16) and pAA1617. Plasmid pAA073 contains the *POX4* promoter and terminator from *Candida* strain ATCC 20336. Plasmid pAA1617 was constructed from pAA061 (Fig. 12) as described herein, in the same manner that pAA073 was constructed from pAA061 except that it contains the *HDE1* promoter and *PEX11* terminator from *Candida* strain ATCC 20336 in place of the *POX4* promoter and terminator as shown for pAA073.

Plasmids pAA709 and pAA710 were cut with *Bsp*QI and the DNA fragments containing coding sequences were purified and cloned into *Bsp*QI-cut plasmid pAA073, placing the genes under the control of the *POX4* promoter and terminator. Sequence-confirmed plasmids were named pAA731 (*ACL1*) and pAA732 (*ACL2*). Similarly, the genes were PCR amplified with primers oAA6658 and oAA6659 for *ACL1* and oAA6660 and oAA6661 for *ACL2*, cut with *Bsp*QI and cloned into *Bsp*QI-cut plasmid pAA1617 placing them under the control of the *HDE1* promoter and *PEX11* terminator. Sequence-confirmed plasmids were named pAA1980 (*ACL1*) and pAA1981 (*ACL2*).

EXAMPLE 10: Modification of nucleic acids encoding acyl-CoA synthetase and cytochrome P450 reductase proteins

Acyl-CoA synthetases (e.g., EC 6.2.1.3) are enzymes that can catalyze the activation of free fatty acids in the cytoplasm into CoA esters (fatty acyl-CoA) which can be involved in several metabolic pathways. When free fatty acids are activated and then used in other cellular processes, they generally are not available for the fatty acid and acetyl-CoA generation that can occur through β-oxidation in peroxisomes. Therefore, because acyl-CoA synthetases can reduce the amount of free fatty acids in the cytoplasm of an organism, such as a yeast, one approach to enhancing fatty acid and acetyl-CoA generation in an organism involves reduction or elimination of the activity of acyl-CoA synthetase activity in the cytoplasm. When acyl-CoA synthetase activity is decreased or eliminated, free fatty acids that accumulate in the cytoplasm can then enter the ω- and β-oxidation pathways.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1637 of 3077

A cytochrome P450 reductase (e.g., EC 1.6.2.4) is an enzyme that can catalyze the reduction of the heme thiolate moiety of cytochrome P450 by transferring an electron to the cytochrome P450. This activity recycles cytochrome P450 and makes it available for further use in catalyzing reactions that occur in ω-oxidation of fatty acids. This, in turn, can provide dicarboxylic acids that are able to undergo β-oxidation in peroxisomes which can result in generation of shorter-chain dicarboxylic acids and acetyl-CoA. Another approach to enhancing dicarboxylic acid and acetyl-CoA production is to increase the activity of cytochrome P450 reductase in an organism. This can be accomplished in a number of ways aimed at increasing the amount of cytochrome P450 reductase protein in the cells. For example, a method for simultaneously decreasing or eliminating acyl-CoA synthetase activity and increasing cytochrome P450 reductase activity in an organism is to disrupt a gene encoding an acyl-CoA synthetase and replace it with nucleic acid encoding a cytochrome P450 reductase.

Cloning nucleic acid encoding the FAA1p from ATCC 20336

The *Candida* strain ATCC 20336 *FAA1* and *CPRB* genes encode an acyl-CoA synthetase (amino acid SEQ ID NO: 35 and nucleotide SEQ ID NO: 91) and cytochrome P450 reductase (amino acid SEQ ID NO: 34 and nucleotide SEQ ID NO: 90), respectively. A double-crossover knock-in cassette was constructed for use in disrupting a *Candida* host strain *FAA*1 gene by integrating into the host genome at the *FAA1* locus a DNA encoding a CPRB protein under the control of the *Candida* strain ATCC 20336 *POX4* promoter and terminator. A DNA fragment encoding the Faa1p acyl-CoA synthetase was amplified from genomic DNA of *Candida* strain ATCC 20336 using primers oAA951 and oAA952. The PCR product was gel purified and ligated into pCR-Blunt II-TOPO (Thermo Fisher Scientific), transformed into competent TOP10 *E. coli* cells (Thermo Fisher Scientific) and clones containing PCR inserts were sequenced to confirm the DNA sequence. One such plasmid was designated pAA275.

Construction of a cassette for disruption of *FAA1* with DNA encoding CPRBp

A DNA fragment from a 5' region of the *FAA*1 gene was amplified from plasmid pAA275 using primers oAA3557 and oAA3558, and a DNA fragment from a 3' region of the *FAA1* gene was amplified from plasmid pAA275 using primers oAA3563 and oAA3564. The *URA3* selectable marker ($P_{URA3}URA3T_{URA3}P_{URA3}$) was amplified from plasmid pAA298 (Fig. 23; SEQ ID NO: 111) using primers oAA3559 and oAA3560. A DNA fragment containing an ORF of a *Candida* strain

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1638 of 3077

ATCC 20336 *CPRB* gene surrounded by the *POX4* promoter and terminator was amplified from plasmid pAA218 using primers oAA3561 and oAA3562. The two PCR products from amplification of pAA275, and the PCR products from amplification of pAA298 and pAA218 were purified, assembled by overlap extension PCR with primers oAA3557 and oAA3564 and ligated into pCR-BluntII-TOPO generating plasmid pAA879 (Fig. 35). The linear *FAA1* deletion cassette could then be amplified from plasmid pAA879 with primers oAA3557 and oAA3564.

EXAMPLE 11: Modification of nucleic acids encoding cytochrome P450 monooxygenase proteins

A cytochrome P450 enzyme (e.g., monooxygenase activity, EC 1.14.14.1) often catalyzes the insertion of one atom of oxygen bound to the heme group in cytochrome P450 into an organic substrate (RH) while the other oxygen atom is reduced to water. This activity can occur in the initial step in the ω-oxidation pathway in which a fatty acid can be converted to a corresponding fatty alcohol. Dicarboxylic acids that are ultimately produced after completion of all steps in the ω-oxidation pathway generally are able to undergo β-oxidation in peroxisomes which results in generation of shorter-chain dicarboxylic acids and acetyl-CoA. Another approach to enhancing dicarboxylic acid and acetyl-CoA production is to increase the activity of cytochrome P450 monooxygenase in a cell. One method for increasing the cytochrome P450 monooxygenase activity in a cell is to increase the expression of cytochrome P450 monooxygenase by, for example, increasing the genomic copy number of cytochrome P450 monooxygenase-encoding DNA (e.g., *CYP52A17* gene) and/or by replacing an endogenous cytochrome P450 monooygenase promoter with a stronger, high-expression heterologous promoter (e.g., the fatty acid-inducible hydratase-dehydrogenase-epimerase (*HDE*) gene promoter from *Candida*).

DNA encoding cytochrome P450 monooxygenase A17 (*CYP52A17*; amino acid SEQ ID NO: 45 and nucleotide SEQ ID NO: 99) from *Candida* strain ATCC 20336 was amplified from genomic DNA using primers oAA5770 and oAA5771. A plasmid backbone containing all the elements of the pCR-BluntII-TOPO vector, a split *URA3* marker and the *HDE* gene promoter and *POX4* gene terminator was amplified from plasmid pAA1164 (Fig. 26) using primers oAA5768 and oAA5769. The nucleic acid products amplified from *Candida* strain ATCC 20336 and pAA1164 contained overlapping sequence at their termini allowing directional ligation of the products to generate plasmid pAA1712. Plasmid pAA1712 contains nucleic acid encoding the CYP52A17

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1639 of 3077

protein under the control of the *HDE1* gene promoter and *POX4* gene terminator situated between the two separate fragments of the *URA3* selectable marker.

EXAMPLE 12: Modification of nucleic acids encoding acyl-CoA oxidase and peroxisomal biogenesis proteins

Acyl-CoA oxidases can be involved in the β-oxidation pathway that occurs in the peroxisomes of many cells and organisms in which long-chain acyl-CoA molecules can be broken down into acetyl-CoA molecules and shorter chain acyl-CoA molecules. The β-oxidation pathway generally includes four main reaction steps resulting in an acyl-CoA that is shortened by two carbon atoms which are released as acetyl-CoA. The shortened acyl-CoA molecule can re-enter the pathway after each cycle and be subjected to another removal of two carbons from the acyl carbon chain. As such, each cycle of the β-oxidation pathway generates a shorter-chain fatty acid and acetyl-CoA, and is a major source of acetyl-CoA in cells. There are multiple acyl-CoA oxidases expressed in yeast and other cells and organisms. These acyl-CoA oxidases can have differing substrate specificities. For example, the peroxisomal acyl-CoA oxidases encoded by the *POX4* and *POX5* genes of *Candida* strain ATCC 20336 exhibit differing activities. The Pox4p enzyme has broad chain length fatty acid specificity whereas the Pox5p enzyme is not very active on fatty acids with short chain lengths. The Pox5p enzyme displays optimal activity on fatty acid substrates of 12 to 18 carbons and a decreased activity on substrates below C10 with a low activity on C6 and C8 substrates.

The catabolism of fatty acids through β-oxidation can be manipulated to serve, in effect, as a process for the production of shorter-chain fatty acids from long-chain acyl-CoA molecules. This can be accomplished by modifying enzymatic activities of the β-oxidation pathway to reduce or eliminate activities that break down fatty acids once they reach a certain short-chain length. For example, one approach to targeting production of short chain dicarboxylic acids (e.g., adipic acid) through the β-oxidation pathway is to delete or disrupt the *POX4* gene in a cell or organism, such as a yeast, e.g., *Candida*. Elimination or reduction of Pox4p acyl-CoA oxidase activity prevents breakdown of long-chain dicarboxylic acids to dicarboxylic acid products with fewer than eight carbon atoms (i.e., chain length shorter than C8) by peroxisomal β-oxidation. One method of disrupting the *POX4* gene and simultaneously introducing a desired nucleic acid for enhancing generation of shorter-chain dicarboxylic acids and acetyl-CoA through modification of β-oxidation is to replace the *POX4* gene with nucleic acid encoding a

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1640 of 3077

peroxisomal membrane protein that promotes peroxisome biogenesis, such as the protein encoded by a yeast *PEX11* (peroxisomal biogenesis factor 11) gene from *Candida* strain ATCC 20336.

Cloning DNA encoding Pex11p from *Candida* strain ATCC 20336

DNA encoding a Pex11p (amino acid SEQ ID NO: 33 and nucleotide SEQ ID NO: 89) from *Candida* strain ATCC 20336 was PCR amplified from genomic DNA using primers oAA2127 and oAA2128. The PCR product was purified, cut with *Bsp*QI, and ligated into *Bsp*QI-cut plasmid pAA073 to operably position the *PEX11* gene fragment under the control of the *POX4* promoter and terminator. The resulting sequence-confirmed plasmid was named pAA336.

Construction of a cassette for disruption of *POX4* by insertion of nucleic acid encoding *P*ex11p

A double-crossover knock-in cassette was constructed to integrate into a host cell genome at the *POX4* locus, DNA encoding *Candida* strain ATCC 20336 Pex11p under the control of the *POX4* promoter and terminator. A DNA fragment containing a 5' homology region of the *Candida* strain ATCC 20336 *POX4* gene and a *Candida* strain ATCC 20336 *URA3* gene selectable marker ($P_{URA3}URA3T_{URA3}P_{URA3}$) was amplified by PCR from plasmid pAA208 (Fig. 36; SEQ ID NO: 127) with primers oAA3156 and oAA3354. A DNA fragment containing a 3' homology region of the *Candida* strain ATCC 20336 *POX4* gene was amplified by PCR from plasmid pAA208 with primers oAA3159 and oAA3356. A DNA fragment containing the coding sequence of the *PEX11* gene under the control of the *POX4* promoter and terminator was amplified by PCR from plasmid pAA336 with primers oAA3157 and oAA3158. All three DNA fragments were assembled by overlap PCR and primers oAA3355 and oAA3357. The 5,190-bp PCR product was then cloned into pCR-BluntII-TOPO and the resulting plasmid was named plasmid pAA850 (Fig. 37).

Modification of DNA encoding Pox5p

In a cell, such as a *Candida* yeast cell, in which Pox5p is the primary or only functional acyl-CoA oxidase, long chain fatty acyl-CoA or diacyl-CoA substrates can be shortened to about 8 carbons and do not typically enter another cycle of β-oxidation for further shortening of the

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1641 of 3077

carbon chain length. The shorter substrates (e.g., a C8 fatty dicarboxylic acid) are not typically recognized as a substrate by Pox5p. In such cells, shorter chain fatty-acyl-CoA molecules are then acted on by peroxisomal thioesterases resulting in removal of Co-A, and the released fatty dicarboxylic acid (e.g., an α,ω - dicarboxylic acid) product is secreted from the cell. In this instance, the acyl-CoA oxidase chain-length substrate specificity effectively controls the chain length of the fatty acid, e.g., a diacid, being produced.

Protein engineering of Pox5p was performed to identify mutants correlated with increased adipic acid production in yeast cells expressing the modified Pox5p enzymes during the fermentation process. A structural model of a Pox5p was created using SWISS-MODEL to identify amino acid positions to mutate. The SWISS-MODEL system has been described, for example, by Arnold *et al.* ((2006) *Bioinformatics* 22: 195-201), Guex *et al.* ((2009) *Electrophoresis* 30 Supplement 1: S162-S173) and Kiefer *et al*. ((2009) *Nucleic Acids Res.* 37 (Database issue): D387-D392). The crystal structure of rat peroxisomal acyl-CoA oxidase II (PDB ID: 1IS2, chain A) (see, for example, Nakajima *et al*. (2002) *J. Biochem.* 131: 376-374) was used as the template to model Pox5p. The resulting structural model was analyzed using the HotSpot Wizard 1.7 program (Worldwide web Uniform Resource Locator (URL) loschmidt.chemi.muni.cz/hotspotwizard/index.jsp) to determine amino acids to target for mutagenesis. HotSpot Wizard is a tool for identifying sites for engineering of substrate specificity and/or activity of enzymes using a combination of structural, functional and sequence analysis and has been described by Pavelka *et al*. ((2009) *Nucleic Acids Res*. 37 (Web Server issue): W376-W383). HotSpot Wizard identified several amino acid positions, or "hotspots," of Pox5p to mutate, with each position given a score from 1 (cold) to 9 (hot) (Fig. 38). The residues identified as hotspots (a score of 6 to 9) are generally found in the portions of the protein that form the walls of the catalytic site pocket and tunnels for entry or exit of the substrate. Part of the HotSpot Wizard analysis is the identification of homologs by a BLAST search (see, e.g., Johnson *et al*. (2008) *Nucleic Acids Res*. 36 (Web Server issue): W5-W9) and their alignment using MUSCLE as described, for example, by Edgar ((2004) *BMC Bioinformatics* 5: 113 and *Nucleic Acids Res*. 32: 1792-1797). The multiple sequence alignment revealed the variety of amino acids found at each position and their relative frequency amongst all the sequences. In the structural model of Pox5p, positions 98 and 429 are located near the back of the catalytic site pocket and are near the catalytic glutamate at position 436. These residues were identified as hotspots and were selected for substitution. Amino acid residue 98, phenylalanine, was identified as a hotspot (with a score of 9) and the residue most frequently

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1642 of 3077

found at this position amongst other homologs is glycine. In Pox5p of *Candida* ATCC 20336, the residue at position 98 is phenylalanine. Therefore, the phenylalanine codon at position 98 was changed to a glycine codon in one modified Pox5p-encoding DNA construct that was generated. The most common residue at amino acid position 429 (with a score of 6) in the analyzed homologs was a tryptophan, which is also the amino acid residue at position 429 in Pox5p of *Candida* ATCC 20336. Therefore, the tryptophan codon at position 429 was changed to the second most common residue, phenylalanine, in another modified Pox5p-encoding DNA construct that was generated. The HotSpot scores for residues 98 and 429 and the amino acids found at that position in the multiple sequence alignment are shown in Table 6.

Table 6: HotSpot Scores and comparison of amino acids at residues 98 and 429 in *Candida* Pox5p

Amino Acid Position	HotSpot Score	Amino Acid in Pox5p of *Candida* ATCC 20336	Residues at the Position from the Alignment
98	9	Phenylalanine	12xG,8xF,5xA,4xY,3xP,3xS,3xW,2xD,2xQ,2xR,1xE,1xH, 1xL,1xN,1Xv
429	6	Tryptophan	23xW,7xF,7xY,4xA,4xV,1xC,1xI,1xL,1xN,1xQ

Site-directed mutagenesis of *POX5*

The *POX5* mutants F98G and W429F were constructed by overlap extension PCR of the template pAA166 containing DNA encoding a wild-type Pox5p. To construct pAA166, oligonucleotides oAA540 and oAA541 were used to amplify DNA encoding Pox5p (amino acid SEQ ID NO: 36 and nucleotide SEQ ID NO: 92) from genomic DNA of *Candida* strain ATCC 20336. The resulting PCR fragment was digested with the restriction enzyme *Bsp*QI and ligated into *Bsp*QI-cut pAA073. The resulting plasmid pAA166 contained DNA encoding Pox5p positioned between the *POX4* promoter and terminator from *Candida* strain ATCC 20336 and also contained a *Candida* strain ATCC 20336 *URA3* gene selectable marker. The nucleic acids encoding *POX5* mutants F98G and W429F were constructed by overlap extension PCR (Fig. 39) of pAA166 using the primers shown in the table shown in Fig. 39 as follows: PCR #1, using oligos A and B, and PCR #2, using oligos C and D were used to amplify the *POX5* coding sequence. Oligos B and C contain the desired point mutation(s). The resulting PCR fragments from PCR #1 and #2 were then used as template and primer for each other to form the full-length coding sequence of the *POX5* mutant in an overlap extension PCR. The resulting overlap extension PCR products were cloned into plasmid pAA073 via the BspQI restriction

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1643 of 3077

sites encoded by oligos A and D. The sequence-confirmed plasmids containing the Pox5p(F98G) (amino acid SEQ ID NO: 37) and Pox5p(W429F) (amino acid SEQ ID NO: 38) mutants were named pAA1055 and pAA1129, respectively. Each plasmid contained DNA encoding a Pox5p mutant (F98G or W429F) positioned between the *POX4* promoter and terminator from *Candida* strain ATCC 20336 and also contained a *Candida* strain ATCC 20336 *URA3* gene selectable marker. Plasmid pAA1164 (Fig. 26), containing DNA encoding a Pox5p mutant (F98G) positioned between, and operably linked to, the *HDE1* promoter and *POX4* terminator from *Candida* strain ATCC 20336 and also containing a *Candida* strain ATCC 20336 *URA3* gene selectable marker, was also generated for expression of Pox5p(F98G).

EXAMPLE 13: Modification of nucleic acids encoding peroxisomal transport proteins

Another approach to enhancing the production of fatty acids, such as dicarboxylic acids, and increasing the amount of cytosolic acetyl-CoA is to reduce or eliminate the transport of fatty acyl-CoA from the cytoplasm of cells directly to peroxisomes. For example, fatty acyl-CoA generated in the cytoplasm can move into peroxisomes by ATP-binding cassette transporters responsible for transporting long-chain fatty acyl-CoA from the cytoplasm across the peroxisomal membrane and into the peroxisomal matrix. Exemplary peroxisomal transport proteins are encoded by yeast *PXA*1 and *PXA*2 (peroxisomal ATP-binding cassette transporter complex) genes. One method of reducing or eliminating the transport of fatty acyl-CoA from the cytoplasm into the peroxisome is to decrease the activity of Pxa1p and Pxa2p transport proteins. This can be accomplished, for example, through deletion or disruption of the *PXA1* and/or *PXA2* genes in an organism. By blocking import of activated acyl-CoA into the peroxisome in, for example, *Candida* host cells, the fatty acyl-CoA generated in the cytosol can be converted to free fatty acid available to undergo ω-oxidation to generate additional dicarboxylic acids for entry into the peroxisome and generation of shorter-chain fatty acids and acetyl-CoA through peroxisomal β-oxidation.

Cloning DNA encoding Pxa1p from *Candida* strain ATCC 20336

DNA encoding a Pxa1p (amino acid SEQ ID NO: 40 and nucleotide SEQ ID NO: 94) from *Candida* strain ATCC 20336 was PCR amplified from genomic DNA using primers oAA2125 and oAA2126. The PCR product, containing DNA encoding a Pxa1p, was purified and cloned into pCR-BluntII-TOPO vector. The resulting sequence-confirmed plasmid was named pAA353.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1644 of 3077

Construction of *PXA1* deletion cassettes

Two *PXA1* deletion cassettes for use in disrupting host cell *PXA*1 genes were constructed that differed in the size of the *PXA*1 homology sequences contained in the cassettes and the orientation of the *URA3* selectable marker DNA positioned between the homology sequences. A first deletion cassette was constructed by amplifying 5' and 3' DNA fragments of *PXA1* from plasmid pAA353 using primers oAA2679 and oAA2680 for amplifying a 5' DNA fragment and oAA2683 and oAA2684 for amplifying a 3' DNA fragment. A *URA3* selectable marker ($P_{URA3}URA3T_{URA3}P_{URA3}$) was amplified from plasmid pAA298 using primers oAA2681 and oAA2682. The three PCR products were purified, assembled by overlap extension PCR and ligated into pCR-BluntII-TOPO generating plasmid pAA1117 (Fig. 40). A linear *PXA1* deletion cassette could then be amplified from plasmid pAA1117 with primers oAA2679 and oAA2684.

A second deletion cassette was constructed by amplifying 5' and 3' DNA fragments of *PXA1* from plasmid pAA353 using primers oAA2914 and oAA2915 for amplifying a 5' DNA fragment and oAA2918 and oAA2919 for amplifying a 3' DNA fragment. A *URA3* selectable marker was amplified from plasmid pAA298 using primers oAA2916 and oAA2917. The three PCR products were purified, assembled by overlap extension PCR and ligated into pCR-Blunt II-TOPO generating plasmid pAA1155 (Fig. 40). A linear *PXA1* deletion cassette could then be amplified from plasmid pAA1155 with primers oAA2914 and oAA2919.

Cloning DNA encoding Pxa2p from *Candida* strain ATCC 20336

DNA encoding Pxa2p (amino acid SEQ ID NO: 41 and nucleotide SEQ ID NO: 95) from *Candida* strain ATCC 20336 was PCR amplified from genomic DNA using primers oAA2159 and oAA2162. The PCR product containing DNA encoding Pxa2p was purified and cloned into pCR-BluntII-TOPO vector. The resulting sequence-confirmed plasmid was named pAA354.

PXA2 deletion cassettes for use in disrupting host *PXA2* genes can be constructed as described for the construction of *PXA1* genes.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1645 of 3077

EXAMPLE 14: Generation of host strains for use in development of genetically modified organisms

In some embodiments of exemplary methods for the development of genetically modified cells and organisms, host strains for transformation can carry beneficial mutations and/or amplifications. Examples include auxotrophic mutations (e.g., *ura3*) that facilitate selection of transformants, and mutations and/or amplifications yielding host cell alterations that serve to work in conjunction with subsequent engineered genetic modifications to enhance generation of cytosolic acetyl-CoA and/or production of fatty acids in the transformants. Alterations in an acyl-CoA oxidase (Pox4p, Pox5p), acyl-CoA synthetase (Faa1p, Fat1p), peroxisomal biogenesis factor 11 (Pex11p) and/or cytochrome P450 reductase (CPRBp) are non-limiting examples of host cell mutations and/or amplifications that can augment the subsequent genetic modifications introduced in the development of organisms for enhanced generation of cytosolic acetyl-CoA and/or production of fatty acids.

For example, in some embodiments, the activity of β- and/or ω-oxidation enzymes in host cells was amplified by placing the genes encoding them under the control of strong and/or inducible promoters and integrating them in additional copies in the genome. To allow the recycling of the *URA3* selectable marker for additional genetic modification, some of these amplified genes were integrated by double-crossover integration at genetic loci providing a simultaneous deletion of an unwanted enzyme activity. For example, in some host cells, the *PEX11* gene was integrated under the control of the *POX4* promoter at the *POX4* locus. This simultaneously provided the deletion of an unwanted activity (Pox4p, acyl-CoA oxidase with undesired substrate specificity) and amplification of a desired activity (Pex11p, peroxisome biogenesis factor). Examples of some possible host strains for use in developing genetically modified cells and organisms for enhanced production of cytosolic acetyl-CoA and/or production of fatty acids are described herein. Fig. 41 provides a flow diagram showing the parent-child relationship for some of the exemplary engineered yeast strains.

Cell transformation methods

Development of genetically modified cells and organisms can involve introduction of heterologous nucleic acids into cells using cell transformation methods described herein and/or known in the art. For example, generally, for the transformation of *Candida* yeast cells, 5 mL

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1646 of 3077

YPD start cultures are inoculated with a single colony of *Candida* cells and incubated overnight at 30°C, with shaking at about 200 rpm. The following day, fresh 25 mL YPD cultures, containing 0.05% Antifoam B, are inoculated to an initial OD_{600nm} of 0.4 and the culture incubated at 30°C, with shaking at about 200 rpm until an OD_{600nm} of 1.0-2.0 is reached. Cells are pelleted by centrifugation at 1,000 x *g*, 4°C for 10 minutes. Cells are washed by resuspending in 10 mL sterile water, pelleted, resuspended in 1 mL sterile water and transferred to a 1.5 mL microcentrifuge tube. The cells are then washed in 1 mL sterile TE/LiOAC solution, pH 7.5, pelleted, resuspended in 0.25 mL TE/LiOAC solution and incubated with shaking at 30°C for 30 minutes.

The cell solution is divided into 50 µL aliquots in 1.5 mL tubes to which is added 5-8 µg of linearized transforming DNA and 5 µL of carrier DNA (boiled and cooled salmon sperm DNA, 10 mg/mL). Sterile PEG solution (300 µL of 40% PEG 3500, 1X TE, 1X LiOAC) is added, mixed thoroughly and incubated at 30°C for 60 minutes with gentle mixing every 15 minutes. DMSO (40 µL) is added, mixed thoroughly and the cell solution is incubated at 42°C for 15 minutes. Cells are then pelleted by centrifugation at 1,000 x *g* 30 seconds, resuspended in 500 µL of YPD media and incubated at 30°C with shaking at about 200 rpm for 2 hours. Cells are then pelleted by centrifugation and resuspended in 1 mL 1X TE, cells are pelleted again, resuspended in 0.2 mL 1X TE and plated on selective media. Plates are incubated at 30°C for growth of transformants. Transformation methods using standard electroporation techniques are also described in U.S. Patent Nos. 5,648,247 and 5,204,252.

Strain sAA886 (*ura3/ura3 pox4a::ura3/pox4b::ura3 faa1::P_{URA3}/faa1::P_{URA3} fat1-Δ1::P_{URA3}/fat1-Δ2::P_{URA3}*)

Table 7 summarizes the genotypes and steps in the construction of *Candida* strain sAA886. Functional *POX5* alleles were restored in *Candida* strain ATCC 20962 (also referred to as sAA003; *ura3/ura3 pox4a::ura3/pox4b::URA3/pox5::ura3/pox5::ura3*; see U.S. Patent No. 5,254,466) by transformation of sAA003 with *POX5* linear DNA to replace the URA3-disrupted loci with a functional allele. The *POX5* DNA was obtained from pAA049 which was constructed by amplifying the entire sequence of NCBI accession number M12161 for the YSAPOX5 locus from genomic DNA prepared from *Candida* strain ATCC 20336 using primers oAA0179 and oAA0182. The 2,624-bp PCR product was cloned into the vector, pCR-BluntII-TOPO, sequenced and designated pAA049. A 2,624-bp DNA fragment containing the *POX5* ORF as

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1647 of 3077

well as 456 bp upstream and 179 bp downstream of the ORF was amplified by PCR of plasmid pAA049 using primers oAA0179 and oAA0182. The purified PCR product was used to transform competent sAA003 cells which were plated on YNB agar plates supplemented with dodecane vapor as the carbon source (e.g., by placing a filter paper soaked with dodecane in the lid of an inverted petri dish) and incubated at 30°C for 4-5 days. Colonies growing on dodecane as the sole carbon source were restreaked onto YPD-agar and incubated at 30°C. Single colonies were grown in YPD cultures and used for the preparation of genomic DNA. PCR analysis of the genomic DNA prepared from the transformants was performed with oligos oAA0179 and oAA0182. A ura3-disrupted *POX5* would produce a PCR product of 4,784 bp while a functional *POX5* would produce a PCR product of 2,624 bp. In a resulting strain, sAA235 (see also U.S. Patent No. 8,343,752), a PCR product 2,624 bp was amplified indicating that both *POX5* alleles had been functionally restored.

Table 7: Yeast strain genotypes and steps in constructing sAA886 and sAA2428

Strain	Genotype	Construction
sAA003	*ura3/ura3 pox4a::ura3/pox4b::URA3 pox5::ura3/pox5::ura3*	ATCC 20962
sAA235	*ura3/ura3 pox4a::ura3/pox4b::URA3* ***POX5/POX5***	Transformation of sAA003 with PCR product using pAA049 as template and growth on YNB-dodecane
sAA329	*ura3/ura3 pox4a::ura3/pox4b::**ura3** POX5/POX5*	Growth of sAA235 on 5-FOA
sAA722	*ura3/ura3 pox4a::ura3/pox4b::ura3 POX5/POX5* ***faa1::URA3****/FAA1*	Transformation of sAA239 with BamHI/XhoI fragment of pAA276
sAA741	*ura3/ura3 pox4a::ura3/pox4b::ura3 POX5/POX5* ***faa1::*** $\mathbf{P_{URA3}}$*/FAA1*	Growth of sAA722 on 5-FOA
sAA776	*ura3/ura3 pox4a::ura3/pox4b::ura3 POX5/POX5 faa1::* P_{URA3}***/faa1::URA3***	Transformation of sAA741 with BamHI/XhoI fragment of pAA282
sAA779	*ura3/ura3 pox4a::ura3/pox4b::ura3 POX5/POX5 faa1::* P_{URA3}***/faa1::*** $\mathbf{P_{URA3}}$	Growth of sAA776 on 5-FOA
sAA865	*ura3/ura3 pox4a::ura3/pox4b::ura3 POX5/POX5 faa1::* P_{URA3}*/faa1::*P_{URA3} ***fat1-Δ1::URA3****/FAT1*	Transformation of sAA779 with deletion cassette generated by OE-PCR
sAA869	*ura3/ura3 pox4a::ura3/pox4b::ura3 POX5/POX5 faa1::* P_{URA3}*/faa1::*P_{URA3} ***fat1-Δ1::*** $\mathbf{P_{URA3}}$*/FAT1*	Growth of sAA865 on 5-FOA

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1648 of 3077

Table 7: Yeast strain genotypes and steps in constructing sAA886 and sAA2428		
Strain	**Genotype**	**Construction**
sAA875	*ura3/ura3 pox4a::ura3/pox4b::ura3 POX5/POX5 faa1::* P_{URA3}*/faa1::*P_{URA3} *fat1-Δ1::*P_{URA3}***/fat1-Δ2::URA3***	Transformation of sAA869 with deletion cassette generated by OE-PCR
sAA886	*ura3/ura3 pox4a::ura3/pox4b::ura3 POX5/POX5 faa1::* P_{URA3}*/faa1::*P_{URA3} *fat1-Δ1::*P_{URA3}***/fat1-Δ2::**P_{URA3}	Growth of sAA875 on 5-FOA
sAA2291	*ura3/ura3 pox4a::ura3/pox4b::ura3* ***pox5::URA3****/POX5 faa1::*P_{URA3}*/faa1::*P_{URA3} *fat1-Δ1::*P_{URA3}*/fat1-Δ2::*P_{URA3}	Transformation of sAA886 with a *Pac*I-digested fragment of pAA918 (SEQ ID NO: 128)
sAA2310	*ura3/ura3 pox4a::ura3/pox4b::ura3* ***pox5Δ::***T_{URA3} */POX5 faa1::*P_{URA3}*/faa1::*P_{URA3} *fat1-Δ1::*P_{URA3}*/fat1-Δ2::*P_{URA3}	Growth of sAA2291 on 5-FOA
sAA2399	*ura3/ura3 pox4a::ura3/pox4b::ura3 pox5Δ::*T_{URA3}*/* ***pox5::URA3*** *faa1::*P_{URA3}*/faa1::*P_{URA3} *fat1-Δ1::*P_{URA3}*/fat1-Δ2::*P_{URA3}	Transformation of sAA2310 with a *Pac*I-digested fragment of pAA918
sAA2428	*ura3/ura3 pox4a::ura3/pox4b::ura3 pox5Δ::*T_{URA3}*/* ***pox5Δ::***T_{URA3} *faa1::* P_{URA3}*/faa1::*P_{URA3} *fat1-Δ1::*P_{URA3}*/fat1-Δ2::*P_{URA3}	Growth of sAA2399 on 5-FOA

In order to delete the *FAA1* genes in the *POX5*-restored strain, sAA329 (generated by growing sAA235 on 5-FOA and selecting for Ura$^-$ cells) was transformed with a fragment of pAA276 (Fig. 42) containing DNA from an *FAA1* gene. *FAA1* DNA was PCR amplified from *Candida* ATCC 20336 genomic DNA using primers oAA951 and oAA952. The PCR product was gel purified and ligated into pCR-BluntII-TOPO, transformed into competent TOP10 *E. coli* cells and clones containing PCR inserts were sequenced to confirm the DNA sequence. One such plasmid was designated pAA275. Plasmid pAA280 was digested with BamHI to release a 2.0 kb P_{URA3}*URA3*$T_{URA3}P_{URA3}$ cassette. Plasmid pAA275 was digested with *Bgl*II and gel purified. The DNA cassette and the *Bgl*II fragment of pAA275 were ligated together to generate plasmids pAA276 and pAA282. Plasmids pAA276 and pAA282 have the P_{URA3}*URA3*$T_{URA3}P_{URA3}$ cassette inserted between the 5' (N-terminal-encoding) and 3' (C-terminal-encoding) fragments of the *FAA1* gene in opposite orientations.

Plasmid pAA276 was digested with *Bam*HI/*Xho*I and column purified. Strain sAA329 (*ura3/ura3 pox4a::ura3/pox4b::ura3 POX5/POX5*) was transformed with the linearized DNA and plated on SCD-URA plate. Several colonies were checked for *FAA1* disruption. One such strain was designated sAA722. Strain sAA722 was grown in YPD media overnight and plated on 5-FOA plate. Colonies that grew in the presence of 5-FOA were PCR screened for the looping out of

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1649 of 3077

the *URA3* gene leaving behind only the *URA3* promoter (P_{URA3}) in the *FAA1* site. Out of 30 colonies analyzed, only one strain showed the correct genetic modification. The strain was designated sAA741. Plasmid pAA282 was digested with *Bam*HI/*Xho*I and column purified. Strain sAA741 was transformed with the linearized DNA and plated on SCD-ura plate. Several colonies were checked for double *FAA1* knockout by insertional inactivation. One such strain was designated sAA776. Strain sAA776 was grown in YPD media overnight and plated on 5-FOA plate. Colonies that grew in the presence of 5-FOA were PCR screened for the looping out of the URA3 gene leaving behind only the *URA3* promoter (PURA3) in both *FAA1* gene alleles. One such strain was named sAA779.

Deletion of each *FAT1* allele was achieved by transforming sAA779 cells with linear DNA cassettes constructed by overlap extension PCR (OE-PCR). The deletion cassette for the first *FAT1* allele in sAA779 was created from three DNA fragments: one containing *FAT1* 5' DNA, one containing *FAT1* 3' DNA and one containing *URA3* marker DNA. The first two DNA fragments containing the *FAT1* DNA were obtained from DNA encoding Fat1p. The full-length coding sequence of the *FAT1* gene was amplified from *Candida* strain ATCC 20336 genomic DNA using primers oAA1023 and oAA1024. The 2,086-bp PCR product was gel purified and ligated into pCR-Blunt II-TOPO, transformed into competent TOP10 E. coli cells and clones containing PCR inserts were sequenced to confirm correct DNA sequence. One such plasmid was named pAA296.

The first DNA fragment (*FAT1* 5' homology) was amplified from plasmid pAA296 using primers oAA2055 and oAA2056. The second DNA fragment (FAT1 3' homology) was amplified from plasmid pAA296 using primers oAA2069 and oAA2060. The third DNA fragment (P_{URA3}*URA3*$T_{URA3}P_{URA3}$) was amplified from plasmid pAA298 using primers oAA2057 and oAA2068. A map of plasmid pAA298 is set forth in Fig. 23. All three DNA fragments were combined in the same reaction to generate the full-length deletion cassette by overlap extension PCR using primers oAA2055 and oAA2060. Strain sAA779 was transformed with the full-length deletion cassette and plated on SCD-URA plate. Several colonies were screened by PCR for integration of the deletion cassette at the first *FAT1* allele. One such strain was named sAA865. Strain sAA865 was grown in YPD media overnight and plated on 5-FOA plate. Colonies that grew in the presence of 5-FOA were PCR screened for the looping out of the *URA3* gene leaving behind only the URA3 promoter (PURA3) in the first *FAT1* allele. One such strain was named sAA869. The deletion of the second *FAT1* allele in sAA869 was performed by

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1650 of 3077

transformation with a deletion cassette created by overlap extension PCR. The deletion cassette for the second *FAT1* allele was constructed from three DNA fragments. The first DNA fragment (FAT1 5' homology) was amplified from plasmid pAA296 using primers oAA2070 and oAA2075. The second DNA fragment (FAT1 3' homology) was amplified from plasmid pAA296 using primers oAA2074 and oAA2075. The third DNA fragment (P_{URA3}*URA3*$T_{URA3}P_{URA3}$) was amplified from plasmid pAA298 using primers oAA2072 and oAA2073. All three DNA fragments were combined in the same reaction to create a full-length deletion cassette by overlap extension PCR using primers oAA2070 and oAA2071. Strain sAA869 was transformed with the full-length deletion cassette and plated on SCD-URA plate. Several colonies were screened by PCR for integration of the deletion cassette at the second *FAT1* allele. One such strain was named sAA875. Strain sAA875 was grown overnight in YPD media and then streaked on to 5-FOA containing plates. Single colonies were tested for *URA3* reversion frequency, and the isolate with least reversion frequency was named sAA886.

Strain sAA2428 (*ura3/ura3 pox4a::ura3/pox4b::ura3 pox5Δ::*T_{URA3}*/pox5Δ::*T_{URA3} *faa1::*P_{URA3}*/faa1::*P_{URA3} *fat1-Δ1::*P_{URA3}*/fat1-Δ2::*P_{URA3})

Strain sAA2428 was generated by knocking out both alleles of *POX5* from the genome of the starting strain sAA886 (*ura3/ura3 pox4a::ura3/pox4b::ura3 POX5/POX5 faa1::* P_{URA3}*/faa1::* P_{URA3} *fat1-Δ1::*P_{URA3}*/fat1-Δ2::*P_{URA3}) in a series of homologous recombination events as outlined in Table 7 above and described herein. Both *POX5* alleles were disrupted by transforming the strain with the double-crossover knock out cassette from plasmid pAA918.

Knock out plasmid pAA918 (SEQ ID NO: 128) was constructed by combining DNA fragments containing *POX5* 5', *POX5* 3' and a *URA3* marker gene. Approximately 600 bp of the 5' flanking region of the *Candida* strain ATCC 20336 *POX5* gene was amplified from genomic DNA using oligos oAA2656 and oAA2657. This fragment was gel purified and cloned into a pCR-Blunt II-TOPO vector to create plasmid pAA494. Approximately 500 bp of the 3' flanking region of the *POX5* gene was amplified with primers oAA2658 and oAA2659 using genomic DNA of *Candida* strain ATCC 20336 and this fragment was cloned into a pCR-Blunt II-TOPO vector to create plasmid pAA495. *EcoRI*/*BamHI*-digested fragment of pAA494, *HindIII*/*BamHI* fragment of pAA495 and *HindIII*/*EcoRI* fragment of pUC19 were ligated together to construct plasmid pAA496. Subsequently, a DNA fragment containing the $T_{URA3}P_{URA3}$*URA3*T_{URA3} cassette flanked by *NotI* restriction sites was cloned into the *NotI* site of pAA496 to create the *POX5* knockout

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1651 of 3077

construct plasmid pAA918 (Fig. 43). Strain sAA886 was transformed with *PacI*-digested fragment of pAA918 and plated on SCD-URA plate. Several colonies were screened by PCR for integration of the deletion cassette in the first *POX5* allele. A positive colony was named sAA2291. Strain sAA2291 was grown in YPD media overnight and plated on a 5-FOA plate. Colonies that grew in the presence of 5-FOA were PCR screened for the looping out of the *URA3* gene to leave behind only the *URA3* terminator (T_{URA3}) in the first *POX5* allele. This strain was named sAA2310. The deletion of the second *POX5* allele was performed by transformation of *PacI*-digested fragment of pAA918 into sAA2310 and plating on a SCD-URA plate. Colonies that grew on the SCD-URA plate were PCR screened for integration of the knockout cassette in the second *POX5* allele. A positive colony was named sAA2399. Strain sAA2399 was grown in YPD media overnight and plated on a 5-FOA plate. Colonies that grew in the presence of 5-FOA were PCR screened for the looping out of the *URA3* gene to leave behind only the *URA3* terminator (T_{URA3}) in the second *POX5* allele. This strain was named sAA2428.

Strains sAA2570 and sAA2571

Host strains containing DNA encoding modified Pox5p enzymes were generated by transformation of strain sAA2428. Strains sAA2570 and sAA2571 were generated by transformation of sAA2428 with DNA encoding a modified Pox5p enzyme (Pox5p F98G) in which the codon for amino acid at position 98 in the protein is a glycine codon instead of the phenylalanine codon at position 98 in the wild-type Pox5p enzyme. Plasmid pAA1055, containing DNA encoding Pox5p (F98G) flanked by the *POX4* promoter and terminator, and also containing a *URA3* selectable marker, was digested with BstZ17I. The resulting linear DNA fragment was transformed into sAA2428. Transformants were selected by growth on SCD-URA plates. Clones that grew well on minimal media containing 0.1% Tween 80 and 0.1% oleic acid as the sole carbon source were then grown in shake flask fermentations and analyzed for dicarboxylic acid production. Two of the selected strains were sAA2570 and sAA2571.

Strains sAA2782, sAA2783 and sAA2784

Strains sAA2782, sAA2783 and sAA2784 were generated by transformation of sAA2428 with DNA encoding a modified Pox5p enzyme (Pox5p W429F) in which the codon for amino acid at position 429 in the protein is a phenylalanine codon instead of the tryptophan codon at position 429 in the wild-type Pox5p enzyme. Plasmid pAA1129, containing DNA encoding Pox5p

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1652 of 3077

(W429F) flanked by the *POX4* promoter and terminator, and also containing a *URA3* selectable marker, was digested with BstZ17I. The resulting linear DNA fragment was transformed into sAA2428. Transformants were selected by growth on SCD-URA plates. Clones that grew well on minimal media containing 0.1% Tween 80 and 0.1% oleic acid as the sole carbon source were then grown in shake flask fermentations and analyzed for fatty acid production. Three of the selected strains were sAA2782, sAA2783 and sAA2784.

Analysis of adipic acid production by strains sAA2570, sAA2571, sAA2782, sAA2783 and sAA2784

Host strains containing DNA encoding modified Pox5p enzymes were grown in shake flask fermentations on oleic acid and analyzed for production of fatty acids. Cultures (5 ml) of each strain that had been grown overnight in YPD medium to an initial OD_{600nm} of 0.4 were used to inoculate 250-mL glass flasks containing 50mL of SP92 media (6.7g/L Difco yeast nitrogen base, 3.0g/L Difco yeast extract, 3.0g/L ammonium sulfate, 1.0g/L potassium phosphate monobasic, 1.0g/L potassium phosphate dibasic, 75g/L glycerol). After 24 h incubation at 30°C with shaking at 250 rpm (2" throw incubator) the cells were centrifuged and the cell pellet resuspended in 15 mL of HiP-TAB media (yeast nitrogen base without amino acids and without ammonium sulfate, 1.7 g/L; yeast extract, 3.0 g/L; potassium phosphate monobasic, 10.0 g/L; potassium phosphate dibasic, 10.0 g/L). The cultures were transferred to fresh 250-mL glass bottom-baffled flasks and 2% v/v oleic acid was added. Cultures were incubated at 30°C with shaking at 300 rpm to start fatty acid production. Samples of culture broth were taken every 24 hours for gas chromatographic (GC) analysis to determine the diacid composition of the culture. The results of analyses of the host strains containing DNA encoding modified Pox5p enzymes were compared to those of a strain containing DNA encoding wild-type Pox5p (sAA875: *pox4a::ura3/pox4b::ura3 POX5/POX5 faa1::* P_{URA3}*/faa1::* P_{URA3} *fat1-Δ1::*P_{URA3}*/fat1-Δ2::URA3*).

The results of the gas chromatographic determinations of fatty acid content of the culture broth of shake flask fermentations comparing strains containing DNA encoding wild-type Pox5p to strains containing DNA encoding mutant Pox5p enzymes are provided in Table 8 and Table 9. The strains containing DNA encoding Pox5p mutants exhibited higher adipic acid (C6 diacid) selectivity, productivity, and yield.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1653 of 3077

Table 8: Comparison of Diacids in Culture Broth from 72-hour Shake Flask Cultures of Strains Expressing Wild-Type Pox5p and an F98G Mutant

Strain name	Pox5p Acyl CoA-Oxidase	g/L C8 diacid	g/L C6 diacid (yield as %Ymax*)
sAA875	Wild-type Pox5p	3.93	10.66 (54)
sAA2570	Pox5p F98G	0.00	17.17 (78)
sAA2571	Pox5p F98G	0.00	15.50 (79)

*Ymax is maximum theoretical yield. It is the amount of product that can be produced for a given biochemical pathway given a certain amount of consumed feedstock (i.e., grams adipic acid/grams oleic acid). The empirical yield is calculated by: mass of product produced divided by the mass of substrate consumed.

Table 9: Comparison of Diacids in Culture Broth from 72-hour Shake Flask Cultures of Strains Expressing Wild-Type Pox5p and a W429F Mutant

Strain name	Pox5p Acyl CoA-Oxidase	g/L C8 diacid	g/L C6 diacid (yield as %Ymax)
sAA875	Wild-type Pox5p	5.75	8.80 (48)
sAA2782	Pox5p W429F	0.41	12.46 (67)
sAA2783	Pox5p W429F	0.62	13.24 (72)
sAA2784	Pox5p W429F	0.77	12.90 (71)

Strain sAA2323 (*ura3/ura3 pox4a::PEX11-P_{URA3}/pox4b::PEX11-P_{URA3} faa1::P_{URA3}/faa1::P_{URA3} fat1-Δ1::P_{URA3}/fat1-Δ2::P_{URA3}*)

Strain sAA2323 was generated by integration of a double-crossover knock-in expression cassette containing nucleic acid encoding *Candida* strain ATCC 20336 Pex11p into both alleles of *pox4* in the genome of the starting strain sAA886 in a series of successive double-crossover homologous recombination events as outlined in Fig. 41 and described herein. Integration of the expression cassette into each *pox4* allele was accomplished by transforming the strain with a double-crossover knock-in cassette from plasmid pAA850.

A linear nucleic acid fragment containing a *POX4* promoter-*PEX11* ORF-*POX4* terminator expression cassette, a *URA3* selectable marker ($P_{URA3}URA3T_{URA3}P_{URA3}$) and a POX4 5' sequence on one end of the fragment and *POX4* 3' sequence on the other end of the fragment was amplified from plasmid pAA850 (FIG. 37) using primers oAA3355 and oAA3357. The linear nucleic acid fragment was used to chemically transform strain sAA886, and transformants were

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1654 of 3077

plated onto SCD-URA plates. The resultant colonies were streaked onto YPD for isolation and characterization. Colony PCR was performed to confirm the presence of the knock-in cassette and one isolate was saved as strain sAA2046 (*ura3/ura3 pox4a::PEX11-URA3/pox4b::ura3 faa1::P_{URA3}/faa1::P_{URA3} fat1-Δ1::P_{URA3}/fat1-Δ2::P_{URA3}*).

Strain sAA2046 was grown overnight in YPD media and plated on 5-FOA plates. Colonies that grew in the presence of 5-FOA were PCR screened for the looping out of the *URA3* gene leaving behind the *URA3* promoter (P_{URA3}) and *PEX11* gene in the first *pox4* allele. The resultant strain was saved as strain sAA2122 (*ura3/ura3 pox4a::PEX11- P_{URA3}/pox4b::ura3 faa1::P_{URA3}/faa1::P_{URA3} fat1-Δ1::P_{URA3}/fat1-Δ2::P_{URA3}*).

A linear nucleic acid fragment amplified from plasmid pAA850 (Fig. 37) using primers oAA3355 and oAA3357 was also used to integrate the *POX4* promoter-*PEX11* ORF-*POX4* terminator expression cassette into the second *pox4* allele. Strain sAA2122 was chemically transformed with the linear nucleic acid, and transformants were plated onto SCD-URA plates. The resultant colonies were streaked onto YPD for isolation and characterization. Colony PCR was performed to confirm the presence of the knock-in cassette at both alleles and one isolate was saved as strain sAA2235 (*ura3/ura3 pox4a::PEX11- P_{URA3}/pox4b:: PEX11-URA3 faa1::P_{URA3}/faa1::P_{URA3} fat1-Δ1::P_{URA3}/fat1-Δ2::P_{URA3}*).

Strain sAA2235 was grown overnight in YPD media and plated on 5-FOA plates. Colonies that grew in the presence of 5-FOA were PCR screened for the looping out of the *URA3* gene leaving behind the *URA3* promoter (P_{URA3}) and *PEX11* gene in the second *pox4* allele. The resultant strain was saved as strain sAA2323 (*ura3/ura3 pox4a::PEX11- P_{URA3}/pox4b::PEX11-P_{URA3} faa1::P_{URA3}/faa1::P_{URA3} fat1-Δ1::P_{URA3}/fat1-Δ2::P_{URA3}*).

Strain sAA2697 (*ura3/ura3 pox4a::PEX11-P_{URA3}/pox4b::PEX11-P_{URA3} faa1::CPRB-P_{URA3}/faa1::CPRB-P_{URA3} fat1-Δ1::P_{URA3}/fat1-Δ2::P_{URA3}*)

Strain sAA2697 was generated by integration of a double-crossover knock-in expression cassette containing nucleic acid encoding *Candida* ATCC 20336 strain CPRBp into both alleles of *faa1* in the genome of the starting strain sAA2323 in a series of successive double-crossover homologous recombination events as outlined in Fig. 41 and described herein. Integration of the expression cassete into each *faa1* allele was accomplished by transforming the strain with the plasmid construct pAA879.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1655 of 3077

A linear nucleic acid fragment containing a *POX4* promoter-*CPRB* ORF-*POX4* terminator expression cassette, a *URA3* selectable marker ($P_{URA3}URA3T_{URA3}P_{URA3}$) and an *FAA1* 5' sequence on one end of the fragment and *FAA1* 3' sequence on the other end of the fragment was amplified from plasmid pAA879 (Fig. 35) using primers oAA3557 and oAA3564. The linear nucleic acid fragment was used to chemically transform strain sAA2323, and transformants were plated onto SCD-URA plates. The resultant colonies were streaked onto YPD for isolation and characterization. Colony PCR was performed to confirm the presence of the knock-in cassette and one isolate was saved as strain sAA2404 (*ura3/ura3 pox4a::PEX11-*P_{URA3}*/pox4b::PEX11-*P_{URA3} *faa1::CPRB-URA3/faa1::*P_{URA3} *fat1-Δ1::*P_{URA3}*/fat1-Δ2::*P_{URA3}). Strain sAA2404 was grown overnight in YPD media and plated on 5-FOA plates. Colonies that grew in the presence of 5-FOA were PCR screened for the looping out of the *URA3* gene leaving behind the *URA3* promoter (P_{URA3}) and *CPRB* gene in the first *FAA1* allele. The resultant strain was saved as strain sAA2539 (*ura3/ura3 pox4a::PEX11-*P_{URA3}*/pox4b::PEX11-*P_{URA3} *faa1::CPRB-*P_{URA3}*/faa1::*P_{URA3} *fat1-Δ1::*P_{URA3}*/fat1-Δ2::*P_{URA3}).

A linear nucleic acid fragment amplified from plasmid pAA879 (Fig. 35) using primers oAA3557 and oAA3564 was also used to integrate the *POX4* promoter-*CPRB* ORF-*POX4* terminator expression cassette into the second *faa1* allele. Strain sAA2539 was chemically transformed with the linear nucleic acid, and transformants were plated onto SCD-URA plates. The resultant colonies were streaked onto YPD for isolation and characterization. Colony PCR was performed to confirm the presence of the knock-in cassette at both alleles and one isolate was saved as strain sAA2622 (*ura3/ura3 pox4a::PEX11-* P_{URA3}*/pox4b::PEX11-*P_{URA3} *faa1::CPRB-*P_{URA3}*/faa1::CPRB-URA3 fat1-Δ1::*P_{URA3}*/fat1-Δ2::*P_{URA3}).

Strain sAA2622 was grown overnight in YPD media and plated on 5-FOA plates. Colonies that grew in the presence of 5-FOA were PCR screened for the looping out of the *URA3* gene leaving behind the *URA3* promoter (P_{URA3}) and *CPRB* gene in the second *FAA1* allele. The resultant strain was saved as strain sAA2697 (*ura3/ura3 pox4a::PEX11-*P_{URA3}*/pox4b::PEX11-*P_{URA3} *faa1::CPRB-*P_{URA3}*/faa1::CPRB-*P_{URA3} *fat1-Δ1::*P_{URA3}*/fat1-Δ2::*P_{URA3}).

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1656 of 3077

Strain sAA2800 (*ura3/ura3 pox4a::PEX11-P_{URA3}/pox4b::PEX11-P_{URA3} faa1::CPRB-P_{URA3}/faa1::CPRB-P_{URA3} fat1-Δ1::P_{URA3}/fat1-Δ2::P_{URA3} pox5Δ::T_{URA3}/pox5Δ::T_{URA3}*)

Strain sAA2800 was generated by knocking out both alleles of *POX5* from the genome of the starting strain sAA2697 in a series of successive double-crossover homologous recombination events as outlined in Fig. 41 and described herein. Both *POX5* alleles were disrupted by transforming the strain with the double-crossover knock-out cassette from plasmid pAA918.

A linear nucleic acid fragment containing a *URA3* selectable marker ($T_{URA3}P_{URA3}$*URA3*T_{URA3}) and a *POX5* 5' sequence on one end of the fragment and *POX5* 3' sequence on the other end of the fragment was obtained by cutting plasmid pAA918 (Fig. 43) with PacI and gel purifying the 2,893-bp DNA cassette. The linear nucleic acid fragment was used to chemically transform strain sAA2697, and transformants were plated onto SCD-URA plates. The resultant colonies were streaked onto YPD for isolation and characterization. Colony PCR was performed to confirm the presence of the knock-out cassette and one isolate was saved as strain sAA2704 (*ura3/ura3 pox4a::PEX11-P_{URA3}/pox4b::PEX11-P_{URA3} faa1::CPRB-P_{URA3}/faa1::CPRB-P_{URA3} fat1-Δ1::P_{URA3}/fat1-Δ2::P_{URA3} pox5Δ::URA3/ POX5*).

Strain sAA2704 was grown overnight in YPD media and plated on 5-FOA plates. Colonies that grew in the presence of 5-FOA were PCR screened for the looping out of the *URA3* gene leaving behind the *URA3* terminator (T_{URA3}) in the first *POX5* allele. The resultant strain was saved as strain sAA2739 (*ura3/ura3 pox4a::PEX11-P_{URA3}/pox4b::PEX11-P_{URA3} faa1::CPRB-P_{URA3}/faa1::CPRB-P_{URA3} fat1-Δ1::P_{URA3}/fat1-Δ2::P_{URA3} pox5Δ::T_{URA3}/ POX5*).

The knock-out of the second *POX5* allele was performed by chemical transformation of strain sAA2739 with a double-crossover knock-out cassette generated by cutting plasmid pAA918 (Fig. 43) with PacI and gel purifying the 2,893-bp DNA cassette. The transformation reaction was plated onto SCD-URA plates. The resultant colonies were streaked onto YPD for isolation and characterization. Colony PCR was performed to confirm the presence of the knock-out cassette at both alleles and one isolate was saved as strain sAA2748 (*ura3/ura3 pox4a::PEX11-P_{URA3}/pox4b::PEX11-P_{URA3} faa1::CPRB-P_{URA3}/faa1::CPRB-P_{URA3} fat1-Δ1::P_{URA3}/fat1-Δ2::P_{URA3} pox5Δ::T_{URA3}/pox5Δ::URA3*).

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1657 of 3077

Strain sAA2748 was grown overnight in YPD media and plated on 5-FOA plates. Colonies that grew in the presence of 5-FOA were PCR screened for the looping out of the *URA3* gene leaving behind the *URA3* terminator (T_{URA3}) in the second *POX5* allele. The resultant strain was saved as strain sAA2800 (*ura3/ura3 pox4a::PEX11-*P_{URA3}*/pox4b::PEX11-*P_{URA3} *faa1::CPRB-*P_{URA3}*/faa1::CPRB-*P_{URA3} *fat1-Δ1::*P_{URA3}*/fat1-Δ2::*P_{URA3} *pox5Δ::*T_{URA3}*/pox5Δ::*T_{URA3}).

Strain sAA3656 (*ura3/ura3 pox4a::PEX11-*P_{URA3}*/pox4b::PEX11-*P_{URA3} *faa1::CPRB-*P_{URA3}*/faa1::CPRB-*P_{URA3} *fat1-Δ1::*P_{URA3}*/fat1-Δ2::*P_{URA3} *pox5Δ::*T_{URA3}*/pox5Δ::*T_{URA3} *pxa1-Δ1::*P_{URA3}*/pxa1-Δ2::*P_{URA3})

Strain sAA3656 was generated by disrupting both alleles of *PXA1* in the genome of the starting strain sAA2800 (*ura3/ura3 pox4a::PEX11-*P_{URA3}*/pox4b::PEX11-*P_{URA3} *faa1::CPRB-*P_{URA3}*/faa1::CPRB-*P_{URA3} *fat1-Δ1::*P_{URA3}*/fat1-Δ2::*P_{URA3} *pox5Δ::*T_{URA3}*/pox5Δ::*T_{URA3}) in a series of successive double-crossover homologous recombination events as outlined in Fig. 41 and described herein. Both *PXA1* alleles were disrupted by transforming the strain with the double-crossover knock-out cassettes from plasmids pAA1117 and pAA1155.

A linear nucleic acid fragment containing a *URA3* selectable marker ($P_{URA3}URA3T_{URA3}P_{URA3}$) and a PXA1 5' sequence on one end of the fragment and *PXA1* 3' sequence on the other end of the fragment was amplified from plasmid pAA1117 (Fig. 40) using primers oAA2679 and oAA2684. The linear nucleic acid fragment was used to chemically transform strain sAA2800, and transformants were plated onto SCD-URA plates. The resultant colonies were streaked onto YPD for isolation and characterization. Colony PCR was performed to confirm the presence of the knock-out cassette and one isolate was saved as strain sAA3467 (*ura3/ura3 pox4a::PEX11-*P_{URA3}*/pox4b::PEX11-*P_{URA3} *faa1::CPRB-*P_{URA3}*/faa1::CPRB-*P_{URA3} *fat1-Δ1::*P_{URA3}*/fat1-Δ2::*P_{URA3} *pox5Δ::*T_{URA3}*/pox5Δ::*T_{URA3} *pxa1-Δ1::URA3/PXA1*).

Strain sAA3467 was grown overnight in YPD media and plated on 5-FOA plates. Colonies that grew in the presence of 5-FOA were PCR screened for the looping out of the *URA3* gene leaving behind the *URA3* promoter (P_{URA3}) in the first *PXA1* allele. The resultant strain was saved as strain sAA3522 (*ura3/ura3 pox4a::PEX11-*P_{URA3}*/pox4b::PEX11-*P_{URA3} *faa1::CPRB-*P_{URA3}*/faa1::CPRB-*P_{URA3} *fat1-Δ1::*P_{URA3}*/fat1-Δ2::*P_{URA3} *pox5Δ::*T_{URA3}*/pox5Δ::*T_{URA3} *pxa1-Δ1::*P_{URA3}*/PXA1*).

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1658 of 3077

The knock-out of the second *PXA1* allele was performed by chemical transformation of strain sAA3522 with a double-crossover knock-out cassette generated by PCR with primers oAA2914 and oAA2919 and plasmid pAA1155 (Fig. 40) as template. The transformation reaction was plated onto SCD-URA plates. The resultant colonies were streaked onto YPD for isolation and characterization. Colony PCR was performed to confirm the presence of the knock-out cassette at both alleles and one isolate was saved as strain sAA3584 (*ura3/ura3 pox4a::PEX11-P_{URA3}/pox4b::PEX11-P_{URA3} faa1::CPRB-P_{URA3}/faa1::CPRB-P_{URA3} fat1-Δ1::P_{URA3}/fat1-Δ2::P_{URA3} pox5Δ::T_{URA3}/pox5Δ::T_{URA3} pxa1-Δ1::P_{URA3}/pxa1-Δ2::URA3*).

Strain sAA3584 was grown overnight in YPD media and plated on 5-FOA plates. Colonies that grew in the presence of 5-FOA were PCR screened for the looping out of the *URA3* gene leaving behind the *URA3* promoter (P_{URA3}) in the second *PXA1* allele. The resultant strain was saved as strain sAA3656 (*ura3/ura3 pox4a::PEX11-P_{URA3}/pox4b::PEX11-P_{URA3} faa1::CPRB-P_{URA3}/faa1::CPRB-P_{URA3} fat1-Δ1::P_{URA3}/fat1-Δ2::P_{URA3} pox5Δ::T_{URA3}/pox5Δ::T_{URA3} pxa1-Δ1::P_{URA3}/pxa1-Δ2:: P_{URA3}*).

EXAMPLE 15: Analysis of the growth of cells and organisms

As described herein, modified organisms developed using compositions and methods provided herein were analyzed for growth in different media. The analysis method was generally carried out as follows. Overnight cultures of engineered transformant strains, as well as control strains, were grown in YPD medium at 30°C with shaking at 250 rpm. The OD_{600nm} of the overnight cultures was measured and the cultures were normalized to an OD_{600nm} of 0.1 by dilution with sterile water in sterile 1.5mL tubes. A serial dilution of the normalized cultures was prepared in 96-well microtiter plates by transferring 250 microliters of normalized cultures into a column 1 well and 200 microliters of water into wells of columns 2 through 6. Using a multichannel pipet, 50 microliters of normalized culture from a column 1 well was transferred to a column 2 well and pipetted up and down to mix. Then 50 microliters of the suspension in a column 2 well was transferred to a column 3 well and pipetted up and down to mix. This process was repeated for successive columns ending at column 6. Finally, 50 microliters of suspension from the column 6 well was removed and discarded, leaving 200 microliters in all wells. A sterile stamp tool formatted for use with 96 well plates was dipped into the 96 well plate containing the culture serial dilution and the adhering hanging drops were transferred to agar plates containing growth medium. The stamping was repeated for all growth medium plates desired. After allowing the

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1659 of 3077

transferred liquid to soak into the plates, they were incubated at 30°C for 1 to 5 days and inspected for growth.

EXAMPLE 16: Generation of cells and organisms expressing modified carnitine acetyltransferase

In one embodiment, increased carnitine O-acetyltransferase activity can be obtained by increasing the amount of carnitine O-acetyltransferase enzyme in the cytosol. For example, engineered carnitine O-acetyltransferase proteins lacking amino acid sequence targeting signals that direct the enzyme to one or more cellular locations (e.g., peroxisomes and mitochondria) other than the cytoplasm can be expressed in host cells in one method of increasing the amount of carnitine O-acetyltransferase in the cytoplasm. Such modified carnitine O-acetyltransferases can be produced upon expression of heterologous nucleic acids encoding the proteins that have been introduced into host cells. In some embodiments, carnitine O-acetyltransferase activity can be decreased in particular cellular locations. For example, host cells that have been modified to knock out endogenous carnitine acetyltransferase activity in mitochondria and peroxisomes can be transformed with nucleic acid encoding carnitine O-acetyltransferases that are less enzymatically active than the endogenous enzyme that target the protein to a particular cellular location(s) (e.g., mitochondria) other than the cytoplasm.

Generation of *CAT2* knock-out strains

Generation of cells in which endogenous *CAT2* genes are disrupted can be accomplished, for example, by successive double-crossover homologous recombination events using gene disruption methods known in the art and/or described herein. In one embodiment, the host is a strain of the yeast *Candida* that contains two alleles of the *CAT2* gene. The effective deletion via disruption of the first *CAT2* allele was performed using a double-crossover knock-out cassette that both deleted a portion of the *CAT2* coding region and disrupted the gene with a *URA3* selectable marker. A double-crossover knock-out cassette was obtained by PCR amplification of plasmid pAA1519 (Fig. 24) with primers oAA2372 and oAA2377. The PCR product was purified and chemically transformed into *Candida* strain ATCC 20913 (a Ura^- version of *Candida* strain ATCC 20336) and the cells were grown on plates containing synthetic complete media supplemented with dextrose and lacking uracil (SCD-URA) to select for cells

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1660 of 3077

that incorporated the *URA3*-containing deletion cassette. The resultant colonies were streaked onto SCD-URA for isolation and characterization. PCR of genomic DNA was performed to confirm the presence of the deletion cassette, and two isolates (*ura3/ura3 cat2-Δ1::URA3/CAT2*) were saved as sAA4477 and sAA4478.

In order to generate a uracil auxotrophic version of strain sAA4477 that could be used as a host to disrupt the second *CAT2* gene allele using uracil-based selection techniques, sAA4477 was grown overnight in YPD media and plated on plates containing 5-fluoroorotic acid (5-FOA). 5-FOA is a compound that is metabolized by *URA3* cells into a chemical that is toxic to the cells and causes cell death. Growth of cells containing the $P_{URA3}URA3T_{URA3}P_{URA3}$ cassette from pAA1519, such as sAA4477 cells, on 5-FOA promotes recombination between the homologous P_{URA3} sequences of the cassette resulting in "looping out" of the functional *URA3* gene and cells that are able to survive in the presence of 5-FOA. Colonies that grew in the presence of 5-FOA were PCR screened for the looping out of the *URA3* gene leaving behind only the *URA3* promoter (P_{URA3}) in the first *CAT2* allele. Two isolates with the correct construction (*ura3/ura3 cat2-Δ1::*P_{URA3}*/CAT2*) were saved as sAA4551 and sAA4552.

The deletion of the second *CAT2* allele was performed using a double-crossover knock-out cassette that both deleted a portion of the *CAT2* coding region and disrupted the gene with a *URA3* selectable marker. Plasmid pAA1520 (Fig. 24) was PCR amplified with primers oAA2372 and oAA2377. The PCR product was purified and chemically transformed into strain sAA4551 and the cells were plated onto SCD-URA plates. The resultant colonies were streaked onto SCD-URA for isolation and characterization. PCR of genomic DNA was performed to confirm the presence of the deletion cassette and two isolates (*ura3/ura3 cat2-Δ1::*P_{URA3}*/cat2-Δ2::URA3*) were saved as sAA4594 and sAA4595.

In order to generate a Ura$^-$ version of strain sAA4594 that could be used as a host to disrupt and/or introduce additional genes using uracil-based selection techniques, strain sAA4594 was grown overnight in YPD media and plated on 5-FOA plates. Colonies that grew in the presence of 5-FOA were PCR screened for the looping out of the *URA3* gene leaving behind only the *URA3* promoter (P_{URA3}) in the second *CAT2* allele. Two isolates with the correct construction (*ura3/ura3 cat2-Δ1::*P_{URA3}*/cat2-Δ2::*P_{URA3}) were saved as sAA4625 and sAA4626. Strains sAA4625 and sAA4626 have both *CAT2* alleles effectively deleted due to their disruption by the presence of the *URA3* promoter (P_{URA3}) at different positions of each *CAT2* allele.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1661 of 3077

Generation of strains containing nucleic acids encoding modified Cat2p

Yeast strains containing nucleic acids encoding modified Cat2 proteins lacking peroxisomal and/or mitochondrial targeting sequences were generated by transformation of a Cat2$^-$ strain sAA4625 with nucleic acid encoding Cat2p$^{\Delta mts}$, Cat2p$^{\Delta pts}$ or nucleic acid encoding Cat2p$^{\Delta mts\Delta pts}$. Plasmid pAA1610 containing nucleic acid encoding the modified Cat2p$^{\Delta mts}$ protein under the control of the *HDE1* gene promoter and *POX4* gene terminator positioned between two separate fragments of the *URA3* gene selectable marker was PCR amplified with primers oAA2206 and oAA2209. The PCR product was purified and chemically transformed into strain sAA4625 and the cells were grown on SC-URA plates. The resultant colonies were restreaked onto YPD for isolation and characterization. PCR of genomic DNA was performed to confirm the presence of the cassette, and an isolate was analyzed for growth as described herein.

Plasmid pAA1876 containing nucleic acid encoding the modified Cat2p$^{\Delta pts}$ protein under the control of the *HDE1* gene promoter and *POX4* gene terminator positioned between the two separate fragments of the *URA3* gene selectable marker was PCR amplified with primers oAA2206 and oAA2209. The PCR product was purified and chemically transformed into strain sAA4625 and the cells were grown on SC-URA plates. The resultant colonies were restreaked onto YPD for isolation and characterization. PCR of genomic DNA was performed to confirm the presence of the cassette, and an isolate was analyzed for growth as described herein.

EXAMPLE 17: Generation of cells and organisms expressing a modified mitochondrial carnitine acetyltransferase

As described in the following Example of spot-growth assays of yeast expressing modified carnitine acetyltransferase proteins, transformation of a Cat2$^-$ strain of *Candida* with nucleic acid encoding Cat2p$^{\Delta mts}$, which lacks a mitochondrial targeting sequence (MTS) but retains an intact peroxisomal targeting sequence (PTS) for targeting the enzyme to the peroxisome, was not sufficient to rescue the growth deficiency of the Cat2$^-$ strain when cultured using oleic acid as the sole carbon source. However, growth of the Cat2$^-$ strain on oleic acid could be restored through expression of a mitochondrial-targeted enzyme with carnitine acetyltransferase activity in conjunction with expression of the Cat2$^{\Delta mts}$ mutant. The mitochondrial carnitine acetyltransferase enzyme could be engineered to have a low activity level to restore minimal

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1662 of 3077

growth on fatty acids and provide a bottleneck resulting in increased cytoplasmic acetyl-carnitine concentrations. In order to generate yeast strains having a reduced carnitine acetyltransferase activity in the mitochondria, Cat2^{-} strains were transformed with DNA encoding modified Yat1p enzymes. In wild-type *Candida* strain ATCC 20336, Yat1p, which has lower enzymatic activity than Cat2p, is localized to the cytoplasm. Therefore, to express the lower-activity carnitine acetyltransferase activity in the mitochondria, the Yat1p-encoding nucleic acid used in the transformation was modified to include nucleic acid encoding mitochondrial targeting sequences (MTS). The transforming DNA encoded the predicted MTS from the Cox4p, Cit1p or Cat2p proteins of *Candida* strain ATCC 20336 which was added to the 5' end of the DNA coding sequence of the *YAT1* gene such that the resulting encoded proteins contained the heterologous MTS in place of the native Yat1p initiating methionine.

Plasmid pAA1967 containing nucleic acid encoding the modified YAT1p$^{+COX4mts}$ protein under the control of the *HDE1* gene promoter and *POX4* gene terminator positioned between the two separate fragments of the *URA3* gene marker was PCR amplified with primers oAA2206 and oAA2209. The PCR product was purified and chemically transformed into Ura^{-},Cat2^{-} strain sAA4625 and the cells were grown on SC-URA plates. The resultant colonies were restreaked onto YPD for isolation and characterization. PCR of genomic DNA was performed to confirm the presence of the cassette, and an isolate was analyzed for growth as described herein. Plasmid pAA1968 containing nucleic acid encoding the modified YAT1p$^{+CIT1mts}$ protein under the control of the *HDE1* gene promoter and *POX4* gene terminator positioned between the two separate fragments of the *URA3* gene marker was PCR amplified with primers oAA2206 and oAA2209. The PCR product was purified and chemically transformed into strain sAA4625 and the cells were grown on SC-URA plates. The resultant colonies were restreaked onto YPD for isolation and characterization. PCR of genomic DNA was performed to confirm the presence of the cassette, and an isolate was analyzed for growth as described herein.

Plasmid pAA1969 containing nucleic acid encoding the modified YAT1p$^{+CAT2mts}$ protein under the control of the *HDE1* gene promoter and *POX4* gene terminator positioned between the two separate fragments of the *URA3* gene marker was PCR amplified with primers oAA2206 and oAA2209. The PCR product was purified and chemically transformed into strain sAA4625 and the cells were grown on SC-URA plates. The resultant colonies were restreaked onto YPD for isolation and characterization. PCR of genomic DNA was performed to confirm the presence of the cassette, and an isolate was analyzed for growth as described herein.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1663 of 3077

EXAMPLE 18: Analysis of organisms in which carnitine acetyltransferase expression has been modified

Analysis of strains via growth assays

Transformant strains engineered for modified carnitine acetyltransferase expression were cultured in control media and in media containing oleic acid and inspected for growth using methods described herein in a previous Example. As shown in Fig. 44, the wild-type strain ATCC 20336 containing two complete *CAT2* genes was able to grow on the control plate containing synthetic complete media with dextrose minus uracil (SCD–URA) and on the fatty acid plate containing yeast nitrogen base without amino acids, plus phosphate and 2% oleic acid (YNBP + 2% oleic acid). A $Cat2^{-}$,Ura^{+} strain (*cat2-Δ1::*P_{URA3}*/cat2-Δ2::URA3*) sAA4594 was able to grow only on the control plate. The deletion of the *CAT2* genes resulted in a mutant strain unable to grow on oleic acid as the sole carbon source. Spot growth assays of serial dilutions were used to evaluate the ability of a genetic construct encoding a Cat2p mutant to rescue the fatty acid growth deficiency of a $Cat2^{-}$ strain. Transformation of a $Ura3^{-}$,$Cat2^{-}$ strain sAA4625 with a genetic construct containing nucleic acid encoding $Cat2p^{\Delta mts}$ and *URA3* selectable marker (amplicon from PCR amplification of pAA1610 with primers oAA2206 and oAA2209) resulted in a strain able to grow on only the control plate. Transformation of a $Ura3^{-}$,$Cat2^{-}$ strain sAA4625 with a genetic construct containing nucleic acid encoding $Cat2p^{\Delta pts}$ and *URA3* selectable marker (amplicon from PCR amplification of pAA1876 with primers oAA2206/oAA2209) resulted in strains (three different transformants designated as *cat2-Δ1::*P_{URA3}*/cat2-Δ2::* P_{URA3} + pAA1876 in Fig. 44) able to grow on both the control and oleic acid plates. Removal of the N-terminal MTS from the Cat2p prevented targeting of the enzyme to the mitochondrial compartment thereby preventing conversion of acetyl-carnitine to acetyl-CoA in the mitochondria and preventing growth on oleic acid. However, removal of the type 1 PTS from the C-terminus of Cat2p did not affect the ability of the enzyme to restore growth on fatty acids to a $Cat2^{-}$ strain. Without being limited or bound by theory, it is believed that the $Cat2p^{\Delta pts}$ is still targeted to the peroxisomal compartment via another mechanism and, with the intact MTS targeting it to the mitochondria, the $Cat2p^{\Delta pts}$ enzyme is able to restore growth on fatty acids.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1664 of 3077

Although expression of *CAT2*$^{\Delta mts}$, which includes nucleic acid encoding a peroxisomal targeting sequence, should result in a Cat2p$^{\Delta mts}$ that is targeted to peroxisomes, expression of *CAT2*$^{\Delta mts}$ alone was not sufficient to rescue the oleic acid growth deficiency in strains in which both *CAT2* genes were disrupted. *Candida* Cat2^{-} deletion strains transformed with DNA encoding additional modified mitochondrial carnitine acetyltransferase enzymes were also tested for growth on oleic acid to further investigate whether growth on fatty acids is restored by a mitochondrial targeted enzyme in conjunction with Cat2p$^{\Delta mts}$. Growth of the Cat2^{-} strains on oleic acid was restored when a carnitine acetyl transferase enzyme (i.e., Yat1p) targeted to the mitochondria was expressed in cells in conjunction with expression of a peroxisomal-targeted Cat2p (encoded by *CAT2*$^{\Delta mts}$). A Ura3^{-},Cat2^{-} mutant sAA4625 was transformed with two separate DNA constructs: one containing a *URA3* selectable marker and *CAT2*$^{\Delta mts}$ (an amplicon from PCR amplification of pAA1610 with primers oAA2206 and oAA2209) and another containing DNA encoding a Yat1p with the MTS of one of three different mitochondrial enzymes (Cox4p, Cit1p or Cat2p) fused to its N-terminus (three separate amplicons from PCR amplification of plasmids pAA1967, pAA1968 and pAA1969 with primers oAA2206 and oAA2209). Because the wild-type endogenous Yat1p is a cytoplasmic enzyme, DNA constructs pAA1967, pAA1968 and pAA1969 were engineered as described herein to encode Yat1p with an MTS from the Cox4p, Cit1p and Cat2p proteins, respectively, added to the N-terminus for mitochondrial targeting. As shown in Fig. 45, the co-transformation of a Cat2^{-} strain with DNA encoding Cat2p$^{\Delta mts}$ and a Yat1p^{+mts} resulted in growth on both the control and oleic acid plates demonstrating a rescue of the fatty acid growth deficiency of a Cat2^{-} deletion strain. Some of the transformants displayed slower growth which may be related to copy number of the transformed genetic constructs that integrated into the host cell genome and/or to the lower activity level of Yat1p. As shown in these results, a mitochondrial-targeted carnitine acetyltransferase enzyme (a Yat1p^{+mts}), which has a decreased activity level compared to wild-type, endogenous Cat2p enzymes, could restore minimal growth on oleic acid to a Cat2^{-} *Candida* strain.

In vitro carnitine acetyltransferase enzyme activity assays

Wild-type control and transformant strains engineered for modified carnitine acetyltransferase expression can also be analyzed for total cellular carnitine acetyltransferase activity. Cells are grown for 48 hours in HiP-TAB medium (yeast nitrogen base without amino acids and without ammonium sulfate, 1.7 g/L; yeast extract, 3.0 g/L; potassium phosphate monobasic, 10.0 g/L;

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1665 of 3077

potassium phosphate dibasic, 10.0 g/L) with 2% oleic acid. Approximately 100 mg of cells are harvested and washed with water then resuspended in 100 mM Tris-HCl, pH 8.0. Soluble cell extracts (in 100 mM Tris-HCl pH 8.0) are prepared by 3 rounds of bead beating cells using 1-minute pulses with 1-minute ice rests followed by pelleting insoluble debris at 13,000 x *g* for 10 min at 4°C. The assay reaction mixture contains 0.18 mM coenzyme A, 5.6 mM acetyl-DL-carnitine, and 92 mM Tris-HCl pH 8.0 buffer. Reactions are carried out at 25°C and enzyme activity is determined by monitoring an increase of absorption at 233 nm. One unit of activity represents the conversion of 1.0 µmole of acetyl-L-carnitine and CoA to L-carnitine and acetyl-coA per minute at pH 8.0 at 25°C.

EXAMPLE 19: Generation of cells and organisms expressing a modified acetyl Co-A hydrolase, acetyl-CoA synthetase and/or carnitine acetyltransferase

Peroxisomal beta-oxidation, which results in generation of acetyl Co-A and fatty acids, generally requires a supply of free coenzyme-A which, in a wild-type cell, can be supplied by peroxisomal Cat2p enzyme transferring acetyl units from acetyl-CoA to carnitine. Another enzyme for liberating coenzyme-A from acetyl-CoA is acetyl-CoA hydrolase (Ach1p). This enzyme can catalyze the conversion of acetyl-CoA to acetate. Because the Ach1p enzyme is targeted to the mitochondrial compartment in wild-type cells, engineering of Ach1p to reroute the enzyme to the peroxisome can involve removing the N-terminal mitochondrial targeting sequence (MTS) and adding a C-terminal peroxisomal targeting sequence (PTS) as described herein. The acetate produced in the peroxisomal compartment via the Ach1p reaction can diffuse out of the peroxisome into the cytoplasm where it can be converted back to acetyl-CoA by cytosolic acetyl-CoA synthetases (Acs1p, Acs2p), the expression of which can be amplified by increasing the copy number of ACS1/2p-encoding nucleic acids in cells and/or by replacing the promoter controlling transcription of the nucleic acids. Cytoplasmic acetyl-CoA can then be used in generating cytosolic malonyl-CoA and in fatty acid synthesis as described herein and/or converted to acetyl-carnitine by cytoplasmic Yat1p, transported into the mitochondria and converted back to acetyl-CoA by endogenous mitochondrial Cat2p or an engineered Yat1p with an added MTS expressed in mitochondria as also described herein.

A Ura3$^-$,Cat2$^-$ mutant strain sAA4625 of *Candida* was transformed with DNA constructs containing a *URA3* selectable marker, *ACH1*$^{\Delta mts+pts}$ encoding an Ach1p lacking an MTS and containing an added PTS, *ACS1*, *ACS2*, and *YAT1*$^{+mts}$ encoding a Yat1p containing an added

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1666 of 3077

MTS from either Cox4p, Cit1p or Cat2p fused to the N-terminus as follows. All transformation cassettes were generated by PCR amplification from their respective plasmids (plasmids pAA1846, pAA1847, pAA1875, pAA1967, pAA1968 and pAA1969) using primers oAA2206 and oAA2209 then purified before transformation. Transformation reactions were plated on SC-URA to select for transformants that had integrated the DNA cassettes containing the URA3 selectable marker(s). Individual colonies were restreaked onto YPD plates for isolation prior to further use.

EXAMPLE 20: Analysis of organisms in which acetyl Co-A hydrolase expression has been modified

To assess the growth capabilities of strains transformed with DNA encoding a modified Ach1p, the strains were analyzed in spot growth assays (using assay methods described herein in a previous Example) and the results were compared to the growth of control strains. A Ura3$^-$,Cat2$^-$ *Candida* strain sAA4625 was transformed with DNA constructs containing a *URA3* selectable marker, *ACH1*$^{\Delta mts+pts}$ encoding an Ach1p lacking an MTS and containing an added PTS, *ACS1*, *ACS2*, and *YAT1*$^{+mts}$ encoding a Yat1p containing an added MTS from either Cox4p, Cit1p or Cat2p fused to the N-terminus. As shown in Fig. 46, although the control Cat2$^-$ deletion strain was unable to grow on oleic acid, the strain co-transformed with *ACH1*$^{\Delta mts+pts}$, *ACS1* and/or *ACS2*, and *YAT1*$^{+mts}$ resulted in growth on both the control (SCD-URA) and oleic acid plates demonstrating a rescue of the fatty acid growth deficiency of the Cat2$^-$ mutant. These results demonstrated that carbon from peroxisomal beta-oxidation could be rerouted to the cytoplasmic compartment in modified host cells.

EXAMPLE 21: Generation of cells and organisms expressing heterologous nucleic acids for modified expression of acetyl-carnitine transport proteins, thioesterase, carnitine acetyltransferase and acetyl-CoA carboxylase

Recombinant microorganisms were developed for modified expression of acetyl-carnitine transport proteins, thioesterase, carnitine acetyltransferase and acetyl-CoA carboxylase through transformation with heterologous nucleic acids as described herein. The cells used in the development of these microorganisms were uracil auxotrophs in which multiple genes (e.g., *CRC1*, *POX4*, *FAA1*, *FAT1*, *PXA1*) had been disrupted and the expression of multiple proteins (e.g., Pox5p, Pex11p, CPRBp, P450A17p) had been amplified and/or altered for increased

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1667 of 3077

activity of the proteins (e.g., enzymes). In one example, *Candida* host yeast cells in which the endogenous *CRC1* genes had been disrupted were used in the development of the modified recombinant microorganisms.

Generation of *CRC1* knock-out strains

Generation of cells in which endogenous *CRC1* genes are disrupted can be accomplished, for example, by double-crossover homologous recombination events using gene disruption methods known in the art and/or described herein. In one embodiment, the host is a strain of the yeast *Candida* that contains two alleles of the *CRC1* gene.

Generation of *CRC1* deletion strain sAA9426

The effective deletion via disruption of the first *CRC1* allele was performed using a cassette that both deleted a portion of the *CRC1* coding region and disrupted the gene with a *URA3* selectable marker containing a P_{URA3} repeat ($P_{URA3}URA3T_{URA3}P_{URA3}$). The linear DNA construct for transformation was PCR amplified from plasmid pAA1613 (Fig. 32) with primers oAA5511 and oAA5512. The PCR product was purified and chemically transformed into *Candida* strain ATCC 20913 (a Ura^- derivative of *Candida* strain ATCC 20336) and the cells were plated onto SCD-URA plates. The resultant colonies were streaked onto YPD for isolation and characterization. Colony PCR was performed to confirm the presence of the deletion cassette and one isolate was saved as strain sAA4368 (*ura3/ura3 crc1-Δ1::URA3/CRC1)*.

Strain sAA4368 was grown overnight in YPD media and plated on 5-FOA plates. Colonies that grew in the presence of 5-FOA were PCR screened for the looping out of the *URA3* gene leaving behind only the *URA3* promoter (P_{URA3}) in the first *CRC1* allele. The resultant strain was saved as strain sAA4388 (*ura3/ura3 crc1-Δ1::*P_{URA3}*/CRC1)*.

The deletion of the second *CRC1* allele was performed using a cassette that deleted a shorter portion of the *CRC1* coding region than the first allele disruption cassette and disrupted the gene with a *URA3* selectable marker containing a P_{URA3} repeat ($P_{URA3}URA3T_{URA3}P_{URA3}$). The linear DNA construct for transformation was PCR amplified from plasmid pAA1701 (Fig. 32) with primers oAA5511 and oAA5512. The PCR product was purified and chemically transformed into strain sAA4388 and the cells were plated onto SCD-URA plates. The resultant colonies

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1668 of 3077

were streaked onto YPD for isolation and characterization. Colony PCR was performed to confirm the presence of the deletion cassette and one isolate was saved as strain sAA9398 (*ura3/ura3 crc1-Δ1::P_{URA3}/crc1-Δ2::URA3)*.

Strain sAA9398 was grown overnight in YPD media and plated on 5-FOA plates. Colonies that grew in the presence of 5-FOA were PCR screened for the looping out of the *URA3* gene leaving behind only the *URA3* promoter (P_{URA3}) in the second *CRC1* allele. The resultant strain was saved as strain sAA9426 (*ura3/ura3 crc1-Δ1::P_{URA3}/ crc1-Δ2::P_{URA3})*. Table 10 summarizes the steps, intermediate strains and genotypes in the development of strain sAA9426.

Table 10: Yeast strain genotypes and steps in constructing sAA9426

Name	Ura3 +/-	Plasmid	Genotype
ATCC 20913	-	NA	*ura3/ura3*
sAA4368	+	pAA1613	*ura3/ura3* ***crc1-Δ1::URA3****/CRC1*
sAA4388	-	NA	*ura3/ura3* ***crc1-Δ1::P_{URA3}****/CRC1*
sAA9398	+	pAA1701	*ura3/ura3 crc1-Δ1::P_{URA3}* ***/crc1-Δ2::URA3***
sAA9426	-	NA	*ura3/ura3 crc1-Δ1::P_{URA3}* ***/crc1-Δ2::P_{URA3}***

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1669 of 3077

Generation of *CRC1* deletion strain sAA4377

The effective deletion via disruption of the first *CRC1* allele was performed using a cassette that both deleted a portion of the *CRC1* coding region and disrupted the gene with a *URA3* selectable marker containing a P_{URA3} repeat ($P_{URA3}URA3T_{URA3}P_{URA3}$). The linear DNA construct for transformation was PCR amplified from plasmid pAA1613 (Fig. 32) with primers oAA5511 and oAA5512. The PCR product was purified and chemically transformed into strain sAA886 (*ura3/ura3 pox4a::ura3/pox4b::ura3 faa1::*P_{URA3}*/faa1::*P_{URA3} *fat1-Δ1::*P_{URA3}*/fat1-Δ2::*P_{URA3}), which was developed through a series of homologous recombination steps from *Candida* strain ATCC 20962 (sAA003) as described herein, and the cells were plated onto SCD-URA plates. The resultant colonies were streaked onto YPD for isolation and characterization. Colony PCR was performed to confirm the presence of the deletion cassette and one isolate was saved as strain sAA4057 (*ura3/ura3 pox4a::ura3/pox4b::ura3 faa1::*P_{URA3}*/faa1::*P_{URA3} *fat1-Δ1::*P_{URA3}*/fat1-Δ2::*P_{URA3} *crc1-Δ1::URA3/CRC1)*.

Strain sAA4057 was grown overnight in YPD media and plated on 5-FOA plates. Colonies that grew in the presence of 5-FOA were PCR screened for the looping out of the *URA3* gene leaving behind only the *URA3* promoter (P_{URA3}) in the first *CRC1* allele. The resultant strain was saved as strain sAA4096 (*ura3/ura3 pox4a::ura3/pox4b::ura3 faa1::*P_{URA3}*/faa1::*P_{URA3} *fat1-Δ1::*P_{URA3}*/fat1-Δ2::*P_{URA3} *crc1-Δ1::*P_{URA3}*/CRC1)*.

The deletion of the second *CRC1* allele was performed using a cassette that deleted a shorter portion of the *CRC1* coding region than the first allele disruption cassette and disrupted the gene with a *URA3* selectable marker containing a P_{URA3} repeat ($P_{URA3}URA3T_{URA3}P_{URA3}$). The linear DNA construct for transformation was PCR amplified from plasmid pAA1701 (Fig. 32) with primers oAA5511 and oAA5512. The PCR product was purified and chemically transformed into strain sAA4096 and the cells were plated onto SCD-URA plates. The resultant colonies were streaked onto YPD for isolation and characterization. Colony PCR was performed to confirm the presence of the deletion cassette and one isolate was saved as strain sAA4281 (*ura3/ura3 pox4a::ura3/pox4b::ura3 faa1::*P_{URA3}*/faa1::*P_{URA3} *fat1-Δ1::*P_{URA3}*/fat1-Δ2::*P_{URA3} *crc1-Δ1::*P_{URA3}*/crc1-Δ2::URA3)*.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1670 of 3077

Strain sAA4281 was grown overnight in YPD media and plated on 5-FOA plates. Colonies that grew in the presence of 5-FOA were PCR screened for the looping out of the *URA3* gene leaving behind only the *URA3* promoter (P_{URA3}) in the second *CRC1* allele. The resultant strain was saved as strain sAA4377 (*ura3/ura3 pox4a::ura3/pox4b::ura3 faa1::P_{URA3}/faa1::P_{URA3} fat1-Δ1::P_{URA3}/fat1-Δ2::P_{URA3} crc1-Δ1::P_{URA3}/ crc1-Δ2::P_{URA3})*. This *CRC1* deletion strain contains mutations in the *FAA1*, *FAT1*, *POX4* and *URA3* genes that inactivate cytoplasmic acyl-CoA synthetases (Faa1p and Fat1p), which prevents free fatty acids in the cytoplasm from being reactivated to the acyl-CoA form, a peroxisomal acyl-CoA oxidase (Pox4p), which blocks break down of short-chain acyl-CoA in β-oxidation, and orotidine-5-phosphate decarboxylase, which results in uracil auxotrophy of the strain. Table 11 summarizes the steps, intermediate strains and genotypes in the development of strain sAA4377.

Table 11: Yeast strain genotypes and steps in constructing sAA4377

Name	Ura3 +/-	Plasmid	Genotype
sAA886	-	NA	*ura3/ura3 pox4a::ura3/pox4b::ura3 faa1::P_{URA3}/faa1::P_{URA3} fat1-Δ1::P_{URA3}/fat1-Δ2::P_{URA3}*
sAA4057	+	pAA1613	*ura3/ura3 pox4a::ura3/pox4b::ura3 faa1::P_{URA3}/faa1::P_{URA3} fat1-Δ1::P_{URA3}/fat1-Δ2::P_{URA3}* ***crc1-Δ1::URA3****/CRC1*
sAA4096	-	NA	*ura3/ura3 pox4a::ura3/pox4b::ura3 faa1::P_{URA3}/faa1::P_{URA3} fat1-Δ1::P_{URA3}/fat1-Δ2::P_{URA3}* ***crc1-Δ1::P_{URA3}****/CRC1*
sAA4281	+	pAA1701	*ura3/ura3 pox4a::ura3/pox4b::ura3 faa1::P_{URA3}/faa1::P_{URA3} fat1-Δ1::P_{URA3}/fat1-Δ2::P_{URA3} crc1-Δ1::P_{URA3}/**crc1-Δ2::URA3***
sAA4377	-	NA	*ura3/ura3 pox4a::ura3/pox4b::ura3 faa1::P_{URA3}/faa1::P_{URA3} fat1-Δ1::P_{URA3}/fat1-Δ2::P_{URA3} crc1-Δ1::P_{URA3}/**crc1-Δ2::P_{URA3}***

Generation of *CRC1* deletion strain sAA6270

A *CRC1* deletion strain was constructed in host strain sAA3656 (described in a previous Example) that had been modified for enhanced ω-oxidation and altered β-oxidation. The effective deletion via disruption of the first *CRC1* allele was performed using a cassette that both deleted a portion of the *CRC1* coding region and disrupted the gene with a *URA3* selectable marker containing a P_{URA3} repeat ($P_{URA3}URA3T_{URA3}P_{URA3}$). The linear DNA construct for transformation was PCR amplified from plasmid pAA1613 (Fig. 32) with primers oAA5511 and oAA5512. The PCR product was purified and chemically transformed into strain sAA3656 and the cells were plated onto SCD-URA plates. The resultant colonies were streaked onto YPD for

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1671 of 3077

isolation and characterization. Colony PCR was performed to confirm the presence of the deletion cassette and one isolate was saved as strain sAA6087 (*ura3/ura3 pox4a::PEX11-P_{URA3}/pox4b::PEX11-P_{URA3} faa1::CPRB-P_{URA3}/faa1::CPRB-P_{URA3} fat1-Δ1::P_{URA3}/fat1-Δ2::P_{URA3} pox5Δ::T_{URA3}/pox5Δ::T_{URA3} pxa1-Δ1::P_{URA3}/pxa1-Δ2::P_{URA3} crc1-Δ1::URA3/CRC1*).

Strain sAA6087 was grown overnight in YPD media and plated on 5-FOA plates. Colonies that grew in the presence of 5-FOA were PCR screened for the looping out of the *URA3* gene leaving behind only the *URA3* promoter (P_{URA3}) in the first *CRC1* allele. The resultant strain was saved as strain sAA6151 (*ura3/ura3 pox4a::PEX11-P_{URA3}/pox4b::PEX11-P_{URA3} faa1::CPRB-P_{URA3}/faa1::CPRB-P_{URA3} fat1-Δ1::P_{URA3}/fat1-Δ2::P_{URA3} pox5Δ::T_{URA3}/pox5Δ::T_{URA3} pxa1-Δ1::P_{URA3}/pxa1-Δ2::P_{URA3} crc1-Δ1::P_{URA3}/CRC1*).

The deletion of the second *CRC1* allele was performed using a cassette that deleted a shorter portion of the *CRC1* coding region than the first allele disruption cassette and disrupted the gene with a *URA3* selectable marker containing a P_{URA3} repeat ($P_{URA3}URA3T_{URA3}P_{URA3}$). The linear DNA construct for transformation was PCR amplified from plasmid pAA1701 (Fig. 32) with primers oAA5511 and oAA5512. The PCR product was purified and chemically transformed into strain sAA6151 and the cells were plated onto SCD-URA plates. The resultant colonies were streaked onto YPD for isolation and characterization. Colony PCR was performed to confirm the presence of the deletion cassette and one isolate was saved as strain sAA6192 (*ura3/ura3 pox4a::PEX11-P_{URA3}/pox4b::PEX11-P_{URA3} faa1::CPRB-P_{URA3}/faa1::CPRB-P_{URA3} fat1-Δ1::P_{URA3}/fat1-Δ2::P_{URA3} pox5Δ::T_{URA3}/pox5Δ::T_{URA3} pxa1-Δ1::P_{URA3}/pxa1-Δ2::P_{URA3} crc1-Δ1::P_{URA3}/crc1-Δ2::URA3*).

Strain sAA6192 was grown overnight in YPD media and plated on 5-FOA plates. Colonies that grew in the presence of 5-FOA were PCR screened for the looping out of the *URA3* gene leaving behind only the *URA3* promoter (P_{URA3}) in the second *CRC1* allele. The resultant strain was saved as strain sAA6270 (*ura3/ura3 pox4a::PEX11-P_{URA3}/pox4b::PEX11-P_{URA3} faa1::CPRB-P_{URA3}/faa1::CPRB-P_{URA3} fat1-Δ1::P_{URA3}/fat1-Δ2::P_{URA3} pox5Δ::T_{URA3}/pox5Δ::T_{URA3} pxa1-Δ1::P_{URA3}/pxa1-Δ2::P_{URA3} crc1-Δ1::P_{URA3}/crc1-Δ2::P_{URA3}*). This *CRC1* deletion strain contains mutations in the *FAA1*, *FAT1*, *POX4, POX5, PXA1* and *URA3* genes that inactivate cytoplasmic acyl-CoA synthetases (thereby preventing free fatty acids in the cytoplasm from being reactivated to the acyl-CoA form), peroxisomal acyl-CoA oxidases (thereby blocking break down of acyl-CoA in β-oxidation), an ATP-binding cassette transporter (thereby blocking import

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1672 of 3077

of activated acyl-CoA into the peroxisomes) and orotidine-5-phosphate decarboxylase (resulting in uracil auxotrophy of the strain). Table 12 summarizes the steps, intermediate strains and genotypes in the development of strain sAA6270.

Table 12: Yeast strain genotypes and steps in constructing sAA6270

Name	Ura3 +/-	Plasmid	Genotype
sAA3656	-	NA	*ura3/ura3 pox4a::PEX11-P_{URA3}/pox4b::PEX11-P_{URA3} faa1::CPRB-P_{URA3}/faa1::CPRB-P_{URA3} fat1-Δ1::P_{URA3}/fat1-Δ2::P_{URA3} pox5Δ::T_{URA3}/pox5Δ::T_{URA3} pxa1-Δ1::P_{URA3}/**pxa1-Δ2::P_{URA3}***
sAA6087	+	pAA1613	*ura3/ura3 pox4a::PEX11-P_{URA3}/pox4b::PEX11-P_{URA3} faa1::CPRB-P_{URA3}/faa1::CPRB-P_{URA3} fat1-Δ1::P_{URA3}/fat1-Δ2::P_{URA3} pox5Δ::T_{URA3}/pox5Δ::T_{URA3} pxa1-Δ1::P_{URA3}/pxa1-Δ2::P_{URA3} **crc1-Δ1::URA3**/CRC1*
sAA6151	-	NA	*ura3/ura3 pox4a::PEX11-P_{URA3}/pox4b::PEX11-P_{URA3} faa1::CPRB-P_{URA3}/faa1::CPRB-P_{URA3} fat1-Δ1::P_{URA3}/fat1-Δ2::P_{URA3} pox5Δ::T_{URA3}/pox5Δ::T_{URA3} pxa1-Δ1::P_{URA3}/pxa1-Δ2::P_{URA3} **crc1-Δ1::P_{URA3}**/CRC1*
sAA6192	+	pAA1701	*ura3/ura3 pox4a::PEX11-P_{URA3}/pox4b::PEX11-P_{URA3} faa1::CPRB-P_{URA3}/faa1::CPRB-P_{URA3} fat1-Δ1::P_{URA3}/fat1-Δ2::P_{URA3} pox5Δ::T_{URA3}/pox5Δ::T_{URA3} pxa1-Δ1::P_{URA3}/pxa1-Δ2::P_{URA3} crc1-Δ1::P_{URA3}/**crc1-Δ2::URA3***
sAA6270	-	NA	*ura3/ura3 pox4a::PEX11-P_{URA3}/pox4b::PEX11-P_{URA3} faa1::CPRB-P_{URA3}/faa1::CPRB-P_{URA3} fat1-Δ1::P_{URA3}/fat1-Δ2::P_{URA3} pox5Δ::T_{URA3}/pox5Δ::T_{URA3} pxa1-Δ1::P_{URA3}/pxa1-Δ2::P_{URA3} crc1-Δ1::P_{URA3}/**crc1-Δ2::P_{URA3}***

Analysis of *CRC1* deletion strains

Growth assays of *CRC1* deletion strains

Transformant strains engineered for modified Crc1p expression were cultured in control media and in media containing oleic acid and inspected for growth using methods described herein in a previous Example. As shown in Fig. 47, yeast strains encoding at least one functional allele of *CRC1* are able to grow on both control media (SCD-URA) and on the fatty acid media (YNBP + 2% oleic acid). Yeast strains with both *CRC1* alleles knocked out are able to grow on control media but not on fatty acid media, demonstrating the key function of the Crc1p inner mitochondrial acetyl-carnitine transporter in fatty acid metabolism. The effect of the double *CRC1* knockout is demonstrated in two different genetic backgrounds. The single (sAA4368)

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1673 of 3077

and double (sAA9398) *CRC1* knockouts were generated from *Candida* strain ATCC 20913 (a Ura$^-$ derivative of *Candida* strain ATCC20336). The single (sAA4057) and double (sAA4281) *CRC1* knockouts were generated from strain sAA875, which is a partially beta-oxidation blocked strain described in a previous Example. Included in the growth assay are yeast strains (sAA5916, sAA5917 and sAA5918) generated by transformation of a double-crossover integration cassette into sAA4377 (the Ura- derivative of sAA4281) that was targeted to IGR5 and contained a *URA3* selectable marker and a *CRC1* gene linked to a weak promoter. The weak promoters used were from the genes encoding glucose-6-phosphate isomerase (*G6PI* in sAA5916), mitochondrial copper and phosphate carrier (*PIC2* in sAA5917), or high-affinity sulfate permease (*SUL2* in sAA5918). The transformants with *CRC1* driven by a weak promoter each contained one integrated copy of the double-crossover cassette. None of the weak promoters were sufficient, with one genomic copy, to drive the expression of enough *CRC1* to rescue the growth defect of the parental strain on fatty acid media.

Generation of strains sAA6317 and sAA6424-sAA6428 expressing heterologous nucleic acids for modified expression of Crc1p, Tes3p, Cat2p and Acc1p

CRC1 deletion strain sAA6270 was used as a host strain in the development of recombinant yeast strains expressing heterologous nucleic acids for modified expression of acetyl-carnitine transport proteins (Crc1p), thioesterase (Tes3p), carnitine acetyltransferase (Cat2p) and acetyl-CoA carboxylase (Acc1p). The final recombinant strains contain DNA encoding: (1) Crc1p under the control of the glucose-6-phosphate isomerase (G6PI) promoter for reduced expression of Crc1p, (2) Tes3p$^{\Delta pts}$ under the control of the hydratase-dehydrogenase-epimerase (*HDE*) promoter to provide for thioesterase enzyme activity in the cytoplasm, (3) Cat2p$^{\Delta mts\Delta pts}$ under the control of the HDE promoter to provide for increased carnitine O-acetyltransferase activity in the cytoplasm, and (4) modified Acc1p (S1153A) under the control of the HDE promoter to provide for increased acetyl-CoA carboxylase activity in the cytoplasm. These heterologous nucleic acids are expressed in the genetic background of a host cell that has been modified to express heterologous DNA encoding a protein (Pox5p(F98G)) for increased selective peroxisomal acyl-CoA oxidase activity as well as proteins for increased cytochrome P450 reductase (CPRBp) and monooxygenase (P450A17p) activities and peroxisomal biogenesis (Pex11p) activity. The host cell has also been modified to decrease or eliminate expression of cytosolic acyl-CoA synthetase, Pox4p acyl-CoA oxidase and peroxisomal Pxa1p acyl-CoA transport protein.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1674 of 3077

Linear DNA cassettes were generated from plasmids pAA1667 ($P_{HDE1}CAT2^{\Delta mts\Delta pts}T_{POX4}$), pAA1164 ($P_{HDE}POX5(F98G)T_{POX4}$), pAA1609 ($P_{HDE1}TES3^{\Delta pts}T_{POX4}$), pAA1712 ($P_{HDE1}P450A17T_{POX4}$), and pAA2311 ($P_{G6PI}CRC1T_{POX4}$) by PCR amplification using primers oAA2206 and oAA2209. A linear DNA cassette was also generated from plasmid pAA1907 (P_{HDE}-*ACC1(S1153A)*-T_{PGK}) by digestion with *Spe*I. The linear DNA cassettes were designed for homologous recombination into the host *ura3* locus. The DNA cassettes were purified and chemically transformed together into strain sAA6270, which was then spread onto plates lacking uracil (SC-URA). The resultant colonies were streaked onto SC-URA plates for isolation and colony PCR was used to verify the presence of the transformed genes. Growth on YNBP+oleic acid plates was used to confirm integration of DNA encoding Crc1p and Pox5p(F98G) into the host genome. Strains confirmed to be URA^+ and containing DNA encoding the recycling loop enzymes (Crc1p, $Tes3p^{\Delta pts}$, $Cat2p^{\Delta mts\Delta pts}$ and Acc1(S1153A)p) were selected and stocked as sAA6317, sAA6424, sAA6425, sAA6426, sAA6427, and sAA6428.

Diacid production by strains sAA6317 and sAA6424-sAA6428 from oleic acid in shake flasks

Transformant strains were grown in shake flask fermentations of oleic acid and analyzed for production of fatty acids. Cultures (5 ml) of each strain were grown overnight in YPD medium and used to inoculate 250-mL glass flasks containing 50mL of SP92 media to an initial OD_{600nm} of 0.4 (SP92 media: 6.7 g/L Difco yeast nitrogen base, 3.0 g/L Difco yeast extract, 3.0 g/L ammonium sulfate, 1.0 g/L potassium phosphate monobasic, 1.0 g/L potassium phosphate dibasic, 75 g/L glycerol). After 24 h incubation at 30°C with shaking at 250 rpm (2" throw incubator) the cells were centrifuged and the cell pellet resuspended in 15 mL of HiP-TAB media (yeast nitrogen base without amino acids and without ammonium sulfate, 1.7 g/L; yeast extract, 3.0 g/L; potassium phosphate monobasic, 10.0 g/L; potassium phosphate dibasic, 10.0 g/L). The cultures were transferred to fresh 250-mL glass bottom-baffled flasks and 2% v/v oleic acid was added. Cultures were incubated at 30°C with shaking at 300 rpm to start fatty acid production. Samples of culture broth were taken every 24 hours for gas chromatographic (GC) analysis to determine the diacid composition of the culture broth. The results of analyses of the transformant strains were compared to those of control strain sAA5082. Strain sAA5082 was generated by transforming strain sAA3656 with antibiotic-free DNA cassettes amplified by PCR with primers oAA2206 and oAA2209 from plasmids pAA1164, pAA1712, pAA573, pAA1672 and pAA1763. Respectively, these plasmids contain DNA encoding Pox5p(F98G),

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1675 of 3077

P450A17, Zwf1p (a *Candida* strain ATCC 20336 glucose-6-phosphate dehydrogenase), and Idp2p and Idp1p$^{\Delta mts}$ (both a *Candida* strain ATCC 20336 isocitrate dehydrogenase). Nucleic acids encoding Pox5p(F98G) and P450A17 are under the control of the *HDE1* promoter and *POX4* terminator. Nucleic acids encoding Idp2p and Idp1p$^{\Delta mts}$ are under the control of the promoter and terminator of *Candida* strain ATCC 20336 *SPS19* gene (encoding peroxisomal 2,4-dienoyl-CoA reductase). DNA encoding Zwf1p is under the control of the *Candida* strain ATCC 20336 *POX4* promoter and terminator. Strain sAA5082 does not contain genetic modifications for rerouting acetyl-CoA (i.e., DNA encoding the recycling loop enzymes Crc1p, Tes3p$^{\Delta pts}$, Cat2p$^{\Delta mts\Delta pts}$ and Acc1(S1153A)p); however, it does contain modified genes for the expression of enzymes producing cytoplasmic NADPH.

The dicarboxylic acid content of the culture broth of shake flask fermentations comparing transformant and control strains are provided in Table 13. As shown in the Table, the transformant strains containing DNA encoding Tes3p$^{\Delta pts}$, Cat2p$^{\Delta mts\Delta pts}$, Acc1(S1153A)p and Crc1p (under the control of the G6PI promoter), generally exhibit higher adipic acid (C6 diacid) production than a control strain that does not contain these transforming nucleic acids.

Table 13: Diacid content of culture broth after 48-hour shake flask growth of strains

Strain	g/L C8 diacid	g/L C6 diacid (yield as %Ymax*)
sAA5082	0.2	15.6 (83)
sAA6317	0.4	15.6 (84)
sAA6424	0.2	16.6 (90)
sAA6425	0.6	17.2 (92)
sAA6426	0.3	16.8 (93)
sAA6427	1.3	16.5 (93)
sAA6428	1.3	15.6 (95)

*Ymax is maximum theoretical yield. It is the amount of product that can be produced for a given biochemical pathway given a certain amount of consumed feedstock (i.e., grams adipic acid/grams oleic acid). The empirical yield is calculated by: mass of product produced divided by the mass of substrate consumed.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1676 of 3077

Diacid production by strains sAA875, sAA6317 and sAA6425 from oleic acid in shake flasks

Transformant strains sAA6317 and sAA6425 were also grown in additional shake flask fermentations of oleic acid (using the same growth and fermentation methods as described in the preceding example for shake flask fermentation and production of diacids by strains sAA6317 and sAA6424-sAA6428) and analyzed for production of fatty acids. Samples of culture broth were taken every 24 hours for gas chromatographic (GC) analysis to determine the diacid composition of the culture broth. The results of analyses of the transformant strains were compared to those of control strain sAA875 (*ura3/ura3 pox4a::ura3/pox4b::ura3 POX5/POX5 faa1::* P_{URA3}*/faa1::*P_{URA3} *fat1-Δ1::*P_{URA3}*/fat1-Δ2::URA3*). Strain sAA875 does not contain genetic modifications for rerouting acetyl-CoA (i.e., DNA encoding the recycling loop enzymes Crc1p, $Tes3p^{\Delta pts}$, $Cat2p^{\Delta mts\Delta pts}$ and Acc1(S1153A)p).

The dicarboxylic acid content of the culture broth of shake flask fermentations comparing transformant and control strains are provided in Table 14. As shown in the Table, the transformant strains containing DNA encoding $Tes3p^{\Delta pts}$, $Cat2p^{\Delta mts\Delta pts}$, Acc1(S1153A)p and Crc1p (under the control of the G6PI promoter), generally exhibit higher adipic acid (C6 diacid) production than a control strain that does not contain these transforming nucleic acids.

Table 14: Diacid content of culture broth after 48-hour shake flask growth of strains

Strain	g/L C8 diacid	g/L C6 diacid (yield as %Ymax*)
sAA875	6.4	8.5 (79)
sAA6317	0.7	18.7 (93)
sAA6425	1.3	17.5 (92)

*Ymax is maximum theoretical yield. It is the amount of product that can be produced for a given biochemical pathway given a certain amount of consumed feedstock (i.e., grams adipic acid/grams oleic acid). The empirical yield is calculated by: mass of product produced divided by the mass of substrate consumed.

Diacid production by strains sAA6317, sAA6425, and sAA6428 from palm fatty acid distillate (PFAD) in a fermentor

Fermentation medium of composition 0.810 g/L phosphoric acid, 1.22 g/L potassium phosphate dibasic, 1.536 g/L magnesium sulfate heptahydrate, 0.058 g/L calcium sulfate dihydrate, 90

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1677 of 3077

mg/L citric acid anhydrous, 3.0 mg/L biotin, 54 g/L glucose, 1 mL of iron solution (1000X stock solution: 60 g/L iron(II)sulfate heptahydrate, 10% sulfuric acid) and 1 mL trace metals mix (1,000X stock solution: 2.5 g/L citric acid, 0.165 g/L cupric sulfate pehtahydrate, 1.209 g/L manganese sulfate monohydrate, 0.54 g/L sodium molybdate, 1.08 g/L zinc sulfate heptahydrate) was filter sterilized and transferred to a sterile fermentation vessel (New Brunswick Bioflo 310 system using 2.5 L vessels). Growth of all *Candida* strains was initiated with a 5% inoculum (initial OD_{600nm}=1.0) and growth conditions of 35°C, 1000 rpm, 1 vvm, pH 5.8 and initial volume of 1.0 L. The pH was maintained with a 7% solution of NH_4OH. Growth continued for approximately 15 hours before exhaustion of the initial carbon source. When the initial carbon source was exhausted as noted by a marked increase in DO%, the temperature control was changed to 30°C, the pH control was changed to 4.5, and the conversion phase was initiated by turning on a continuous feed of ethylated PFAD at a rate of 1.04 g/L-h. PFAD is a fatty acid-containing residue resulting from the process of refining palm oil. Fermentation conditions were maintained at 30°C, 1000 rpm, 1 vvm, and pH 4.5 for 12 hours at which point the continuous feed rate of ethylated PFAD was changed to 2.25 g/L-h. The fermentation was carried out for a total of 130 hours. Samples were collected for gas chromatographic analysis every 24 hours after initiating the conversion phase.

The dicarboxylic acid content of the final fermentation broth is provided in Table 15. As shown in the Table, the transformant strains sAA6317, sAA6425, and sAA6428 containing DNA encoding $Tes3p^{\Delta pts}$, $Cat2p^{\Delta mts\Delta pts}$, Acc1(S1153A)p and Crc1p (under the control of the G6PI promoter), generally exhibit higher adipic acid (C6 diacid) titer and yield than a control strain (sAA5082) that does not contain these transforming nucleic acids.

Table 15: Diacid content of fermentation broth after 130-hour fermentation of strains

Strain	g/L C8 diacid	g/L C6 diacid (yield as %Ymax*)
sAA5082	0.9	59.0 (81.5)
sAA6317	2.7	60.8 (90.0)
sAA6425	3.1	63.9 (92.4)
sAA6428	3.9	69.3 (96.2)

*Ymax is maximum theoretical yield. It is the amount of product that can be produced for a given biochemical pathway given a certain amount of consumed feedstock (i.e., grams adipic acid/grams oleic acid). The empirical yield is calculated by: mass of product produced divided by the mass of substrate consumed.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1678 of 3077

EXAMPLE 22: Generation of organisms expressing modified ACL1 and ACL2 genes

Heterolgous nucleic acid constructs containing DNA encoding *Yarrowia lipolytica* CLIB122 ATP citrate lyase subunits (Acl1p and Acl2p) that was codon-optimized for expression in *Candida* strain ATCC 20336 were used to transform yeast host cells. Plasmids pAA1980 and pAA1981 contained the Acl1p and Acl2p subunit-encoding nucleic acids, respectively, operably linked to an *HDE1* promoter and *PEX11* terminator. The heterologous nucleic acids were used to transform strain sAA886 lacking functional *POX4*, *FAA1* and *FAT1* genes.

For transformation of host cells with DNA contained within pAA1980 and pAA1981, the plasmids were digested with BSTZ17I and the linear DNA was introduced into the cells for integration at the *ura3* alleles. Transformation reactions were plated on SC-URA to select for transformants that had integrated the DNA cassettes containing the *URA3* selectable marker. Individual colonies were restreaked onto YPD plates for isolation prior to further use. Four isolates were verified by PCR to have integrated both pAA1980 and pAA1981 and were saved as strains sAA5244, sAA5245, sAA5246, and sAA5247.

EXAMPLE 23: Analysis of organisms in which ATP citrate lyase expression has been modified

Strains transformed with heterologous nucleic acids encoding Acl1p and Acl2p were analyzed for dicarboxylic acid production during growth in shake flasks. Glass flasks (250 mL) containing 25 mL of SP92 media were inoculated with a 5 mL YPD overnight culture to an initial OD_{600nm} of 0.2. After 24 h incubation at 30°C with shaking at 250 rpm (2" throw incubator), the cells were centrifuged and the cell pellet resuspended in 12.5 mL of HiP-TAB media (yeast nitrogen base without amino acids and without ammonium sulfate, 1.7 g/L; yeast extract, 3.0 g/L; potassium phosphate monobasic, 10.0 g/L; potassium phosphate dibasic, 10.0 g/L). The cultures were transferred to fresh 250-mL glass bottom-baffled flasks and 2% v/v oleic acid was added. Cultures were incubated at 30°C with shaking at 300 rpm to start adipic acid production. Samples of culture broth were taken every 24 hours for gas chromatographic (GC) analysis. Table 16 shows the diacid profile in the culture broth of the shake flasks 48 hours after addition of the oleic acid.

The results of analyses of the transformed strains containing heterologous DNA encoding Acl1p and Acl2p were compared to those of the *Candida* strain sAA875 that was not transformed with

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1679 of 3077

the heterologous DNA and does not express an endogenous ATP citrate lyase. As shown in Table 16, the transformants containing heterologous ACL enzyme display an increased titer of adipic acid (C6 diacid) compared to the control strain. Additionally, the transformants selectively produce C6 diacid with very little C8 diacid produced.

Table 16: Diacid content of culture broth after 48-hour shake flask growth of strains transformed with DNA encoding ATP citrate lyase

Strain name	Heterologous ATP-Citrate Lyase	g/L C8 diacid	g/L C6 diacid
sAA875	None	3.51	7.60
sAA5244	*Y. lipolytica* ACL1 & ACL2	0.01	8.76
sAA5245	*Y. lipolytica* ACL1 & ACL2	0.01	8.31
sAA5246	*Y. lipolytica* ACL1 & ACL2	0.01	8.43
sAA5247	*Y. lipolytica* ACL1 & ACL2	0.01	7.87

EXAMPLE 24: Analysis of organisms in which thioesterase expression has been modified

Plasmid pAA1609 harboring the nucleic acid encoding the Tes3p$^{\Delta pts}$ protein under the control of the *HDE1* gene promoter and *POX4* gene terminator was transformed into strain sAA2800 (*ura3/ura3 pox4a::PEX11-*P_{URA3}*/pox4b::PEX11-*P_{URA3} *faa1::CPRB-*P_{URA3}*/faa1::CPRB-*P_{URA3} *fat1-Δ1::*P_{URA3}*/fat1-Δ2::*P_{URA3} *pox5Δ::*T_{URA3}*/pox5Δ::*T_{URA3}). Two transformants with verified integration of plasmid pAA1609 were compared to the Ura+ parent of strain sAA2800 (strain sAA2748). The control strain (sAA2748) and the two transformants have both alleles of *POX4* and *POX5* disrupted and are completely beta-oxidation blocked. They are still able to grow on glucose but cannot metabolize fatty acids. Cultures (5 ml) of each strain were grown overnight in YPD medium and used to inoculate 250-mL glass flasks containing 50mL of SP92 media to an initial OD600nm of 0.4 (SP92 media: 6.7 g/L Difco yeast nitrogen base, 3.0 g/L Difco yeast extract, 3.0 g/L ammonium sulfate, 1.0 g/L potassium phosphate monobasic, 1.0 g/L potassium phosphate dibasic, 75 g/L glucose). After 24 h incubation at 30°C with shaking at 250 rpm (2" throw incubator) samples of culture broth were taken for gas chromatographic (GC) analysis to determine the diacid composition of the culture broth.

As shown in Table 17, transformants with the nucleic acid encoding the Tes3p$^{\Delta pts}$ protein localized to the cytoplasm displayed increased concentrations of long chain (C16 and C18)

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1680 of 3077

diacids compared to the control strain that does not contain the Tes3p$^{\Delta pts}$-encoding nucleic acid. These long chain diacids may arrive from the activity of the cytoplasmic Tes3p$^{\Delta pts}$ protein converting long chain acyl-CoA products of fatty acid synthesis into free fatty acids. The free fatty acids can then be converted to diacids by the omega oxidation pathway.

5

Table 17: Diacid content of culture broth after 24-hour shake flask growth of strains

Strain	C16 Diacid (g/L)	C18 Diacid (g/L)	Total long chain Diacid (g/L)
sAA2748	0.31	0.37	0.68
sAA2800 + pAA1609 clone 1	0.88	1.41	2.28
sAA2800 + pAA1609 clone 2	0.93	1.51	2.44

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1681 of 3077

EXAMPLE 25: Amino acid sequences referenced herein

Table 18

Gene (Organism)	Amino acid sequence	SEQ ID NO
URA3 (*Candida viswanathii* ATCC 20336)	MVSTKTYTERASAHPSKVAQRLFRLMESKKTNLCASIDVTTTAEFLSLIDKLGPHICLVKTHIDIISDFSYEGTIEPLLVL AERHGFLIFEDRKFADIGNTVMLQYTSGVYRIAAWSDITNAHGVTGKGVVEGLKRGAEGVEKERGVLMLAELSSKGS LAHGEYTRETIEIAKSDREFVIGFIAQRDMGGREEGFDWIIMTPGVGLDDKGDALGQQYRTVDEVVLTGTDVIIVGRG LFGKGRDPEVEGKRYRDAGWKAYLKRTGQLE	1
CAT2 (*Candida viswanathii* ATCC 20336)	MFNFKLSQQVLKNSTKSIMPILKKPFSTSHAKGDLFKYQSQLPKLPVPTLEETASKYLKTVEPFLNQEQLESTKAKVA EFVRPGGAGEALQARLNNFAADKDNWLAEFWDDYAYMSYRDPVVPVYSYFFSHKDVKNIIGQDQLLKATLIAYYTIE FQEKVLDESLDPEVIKGNPFCMNAFKYMFNNSRVPAEGSDITQHYNGEENQFFVVIYKNNFYKVPTHKNGQRLTKG EIYSYLQEIKNDATPKGLGLGALTSLNRDEWLSAYNNLLKSPINEASLGSIFASSFVIALDSNNPVTIEEKSKNCWHGD GQNRFFDKPLEFFVSANGNSGFLGEHSRMDATPTVQLNNTIYKQILETNPNDLIVEIGSSAPRFGNAEILPFDINPTTR ANIKDAIAKFDATIAAHDEEIFQHYGYGKGLIKKFKVSPDAYVQLLMQLAYFKYTGKIRPTYESAATRKFLKGRTETGR TVSNESKKFVETWSDPNASSADKVATFQAAAKQHVAYLSAAADGKGVDRHLFGLKQMIQPGEPIPEIFTDPIFSYSQ TWYISSSQVPSEFFQSWGWSQVIDDGFGLAYLINNDWIHVHISCKRGNGLQSDHLKWYLVDSANEMKDVLTKGLLT DAKPKL	2
CAT2$^{\Delta nts}$ contained in pAA1610	MPILKKPFSTSHAKGDLFKYQSQLPKLPVPTLEETASKYLKTVEPFLNQEQLESTKAKVAEFVRPGGAGEALQARLN NFAADKDNWLAEFWDDYAYMSYRDPVVPVYSYFFSHKDVKNIIGQDQLLKATLIAYYTIEFQEKVLDESLDPEVIKGN PFCMNAFKYMFNNSRVPAEGSDITQHYNGEENQFFVVIYKNNFYKVPTHKNGQRLTKGEIYSYLQEIKNDATPKGLG LGALTSLNRDEWLSAYNNLLKSPINEASLGSIFASSFVIALDSNNPVTIEEKSKNCWHGDGQNRFFDKPLEFFVSANG NSGFLGEHSRMDATPTVQLNNTIYKQILETNPNDLIVEIGSSAPRFGNAEILPFDINPTTRANIKDAIAKFDATIAAHDEE IFQHYGYGKGLIKKFKVSPDAYVQLLMQLAYFKYTGKIRPTYESAATRKFLKGRTETGRTVSNESKKFVETWSDPNA SSADKVATFQAAAKQHVAYLSAAADGKGVDRHLFGLKQMIQPGEPIPEIFTDPIFSYSQTWYISSSQVPSEFFQSWG WSQVIDDGFGLAYLINNDWIHVHISCKRGNGLQSDHLKWYLVDSANEMKDVLTKGLLTDAK	3
CAT2$^{\Delta nts\Delta pts}$ contained in pAA1667	MPILKKPFSTSHAKGDLFKYQSQLPKLPVPTLEETASKYLKTVEPFLNQEQLESTKAKVAEFVRPGGAGEALQARLN NFAADKDNWLAEFWDDYAYMSYRDPVVPVYSYFFSHKDVKNIIGQDQLLKATLIAYYTIEFQEKVLDESLDPEVIKGN PFCMNAFKYMFNNSRVPAEGSDITQHYNGEENQFFVVIYKNNFYKVPTHKNGQRLTKGEIYSYLQEIKNDATPKGLG LGALTSLNRDEWLSAYNNLLKSPINEASLGSIFASSFVIALDSNNPVTIEEKSKNCWHGDGQNRFFDKPLEFFVSANG NSGFLGEHSRMDATPTVQLNNTIYKQILETNPNDLIVEIGSSAPRFGNAEILPFDINPTTRANIKDAIAKFDATIAAHDEE IFQHYGYGKGLIKKFKVSPDAYVQLLMQLAYFKYTGKIRPTYESAATRKFLKGRTETGRTVSNESKKFVETWSDPNA SSADKVATFQAAAKQHVAYLSAAADGKGVDRHLFGLKQMIQPGEPIPEIFTDPIFSYSQTWYISSSQVPSEFFQSWG WSQVIDDGFGLAYLINNDWIHVHISCKRGNGLQSDHLKWYLVDSANEMKDVLTKGLLTDAK	4

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1682 of 3077

Table 18

Gene (Organism)	Amino acid sequence	SEQ ID NO
CAT2$^{\Delta pts}$ contained in pAA1876	Mfnfklsqqvlknstksimpilkkpfststshakgdlfkyqsqlpklpvptleetaskylktvepflnqeqlestkakvaefvrpggagealqarlnnfaadkdnwlaefwd dyaymsyrdpvvpyvsyffshkdvkniiggqdqllkatliayytiefqekvldesldpevikgnpfcmnafkymfnnsrvpaegsditqhyngeenqffvviyknnfyk vpthknggrltkgeiysylqeiknddatpkglglgaltslnrdewlsaynnllkspineaslgsifassfvialdsnnpvtieeksknccwhgdgqnrffdlkpleffvsangns gflgehsrmdatptvqlnntiykqiletnpndliveigssaprfgnaeilpfdinpttranikdaiakfdatiaahdeeifqhygygkglikkfkvspdayvqllmqlayfkyt gkirptyesaatrkflkgrtetgrtvsneskkfvetwsdpnassadkvatfqaaakqhvaylsaaadgkgvdrhlfglkqmiqpgepipeiftdpifsysqtwyisssqv pseffqswgwsqviddgfglaylinndwihvhisckrgnglqsdhlkwylvdsanemkdvltkglltdak	5
YAT1 (*Candida viswanathii* ATCC20336)	Mstyqfqetlekllpipdlnqtcanylnvlrplqteqehiktktavenflkngtgqyldaalreyaqtrpsyieqfwydaylnydspvvlnlnpffilleddpftnesssvnpqv kratslvmssllkfiqalknetlsvdtlkggkpolcmyqytklfgasripsedgcvmqsdpasnhivvmsksqlywfdvldsknnliliseaelnvnfqsiihdslrtpsdeia kssfgvlttenrriwanvrhnlmsttnnkvnhevlsiidsalfvlcfddivindlselsknmlcglsildngiqvgtctnrwydklqiivtknakaginfehtgvdghtvlnfvsd iytdsilsfansinsnapslwnttttnykeldgedlitvprkleweltpdlslalrfgetrlsdlinqnefrhlefknygstqikkmkfspdafvqmafqatyyalygkvectyep amtkqfyhgrteairtvsqesnlfvrkffdstvsiqkkleyltqactkhsqqtrmssaggqvdrhlyalfciwkrylhdaedddddaskdqpsstelsrddtiiaqtngdst ddddsttvvghganhhhnlksadllktipeifadngwdklnntiistsncgnpslrlfgfgpvsangfgigyilkddsisicasskhrqtqrflvtlnsylleiqniwkqaqk meklkaelakavedakpkgqksgeaegiiepkpnnlstllggygyfdmgdediksrgqspeppflhragsgtfsireigkklrlsey	6
COX4 mitochondrial targeting sequence (*Candida viswanathii* ATCC20336)	Mlsrttlrvarqqtrllstsrilfnsktdq	7
CIT1 mitochondrial targeting sequence (*Candida viswanathii* ATCC20336)	Msalrsfqrssnvakstlknsvrty	8

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1683 of 3077

Table 18

Gene (Organism)	Amino acid sequence	SEQ ID NO
CAT2 mitochondrial targeting sequence (*Candida viswanathii* ATCC20336)	Mfnfklsqqvlknstksimpilkkpfstshakg	9
YAT1$^{+COX4mts}$ contained in pAA1967	Mlsrttlrvarqqtrllstsrilfnsktdqstyqfqetlekipidlnqtcanylnvlrplqteqehiktktavenflkngtgqyldaalreyaqtrpsyieqfwydaylnydspvvl nlnpfflleddpftnesssvnpqvkratslvmsslkfiqalknetlsvdtlkggkplcmyqytklfgasripsedgcvmqsdpasnhivvmsksqlywfdvldskknlills eaelnvnfqsiihdslrtpsdeiakssfgvlttenrriwanvrhnlmsttnnkvnhevlsiidsalfvlcfddivindlselsknmlcglsildngiqvgtctnrwydklqiivtk nakaginfehtgvdghtvlrfvsdiytdsilsfansinsnapslwntttnykeldgeditvprkleweltpdlslalrfgetrlsdlinqnefrhlefknygstqikkmkfspda fvqmafqatyyalygkvectyepamtkqfyhgrteairtvsqesnlfvrkffdstvsiqkkleyltqactknsqqtrmssagggvdrhlyalfciwkrylhdaeddddda skdqpsstelsrddtiiaqtngdstddddsttvvghganhhhnlksadllktipeifadngwdklnntiistsncgnpslrlfgfpvsangfgigyilkddsisicasskhr qtqrflvtlnsylleiqniwkqaqkmeklkaelakavedakpkgqksgeaegiiepkpnnlstllggygyfdmgdediksrgqspeppflhragsgfsireigkklrlse y	10
YAT1$^{+CIT1mts}$ contained in pAA1968	Msalrsfqrssnvakstlknsvrtystyqfqetlekipipdlnqtcanylnvlrplqteqehiktktavenflkngtgqyldaalreyaqtrpsyieqfwydaylnydspvvl nlnpfflleddpftnesssvnpqvkratslvmsslkfiqalknetlsvdtlkggkplcmyqytklfgasripsedgcvmqsdpasnhivvmsksqlywfdvldskknlills eaelnvnfqsiihdslrtpsdeiakssfgvlttenrriwanvrhnlmsttnnkvnhevlsiidsalfvlcfddivindlselsknmlcglsildngiqvgtctnrwydklqiivtk nakaginfehtgvdghtvlrfvsdiytdsilsfansinsnapslwntttnykeldgeditvprkleweltpdlslalrfgetrlsdlinqnefrhlefknygstqikkmkfspda fvqmafqatyyalygkvectyepamtkqfyhgrteairtvsqesnlfvrkffdstvsiqkkleyltqactknsqqtrmssagggvdrhlyalfciwkrylhdaeddddda skdqpsstelsrddtiiaqtngdstddddsttvvghganhhhnlksadllktipeifadngwdklnntiistsncgnpslrlfgfpvsangfgigyilkddsisicasskhr qtqrflvtlnsylleiqniwkqaqkmeklkaelakavedakpkgqksgeaegiiepkpnnlstllggygyfdmgdediksrgqspeppflhragsgfsireigkklrlse y	11

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1684 of 3077

Table 18

Gene (Organism)	Amino acid sequence	SEQ ID NO
YAT1$^{+CAT2mts}$ contained in pAA1969	Mfnfklsqqvlknstksimpilkkpfststhakgstyqfqetleklpipdlnqtcanylnvlrplqteqehiktktavenflkngtgqyldaalreyaqtrpsyieqfwydaylnydspvvlnlnpffllddpftnesssvnpqvkratslvmsslkfiqalknetlsvdtlkggkplcmcyqytklfgasripsedgcvmqsdpasnhivvmsksqlywfdvldsknnlilseaelnvnfqsiihdslrtpsdeiakssfgvlttenrriwanvrhnlmsttnnkvnhevlsiidsalfvlcfddivindlselskmlcglsildngiqvgtctnrwydklqiivtknakaginfehtgvdghtvlrfvsdiytdsilsfansinsnapslwntttnykeldgedlitvprklewelttpdlslairfgetrlsdlinqnefrhlefknygstqikkmkfspdafvqmafqatyyalygkvectyepeddddddaskdqpsstelsrddtiiaqtngdstddddsttvvghganhhhnlksadllktipeifadngwdklnntiistsncgnpslrlfgfpvsangfgigyilkddsisicasskhrqtqrflvtlnsylleiqniwkqaqkmeklkaelakavedakpkgqksgeaegiiepkpnnlstllggyyfdmgdediksrgqspeppflhragsgfsireigkklrlseypkl	12
YAT1^{+pts}	Mstyqfqetlekipipdlnqtcanylnvlrplqteqehiktktavenflkngtgqyldaalreyaqtrpsyieqfwydaylnydspvvlnlnpffllddpftnesssvnpqvkratslvmsslkfiqalknetlsvdtlkggkplcmcyqytklfgasripsedgcvmqsdpasnhivvmsksqlywfdvldsknnlilseaelnvnfqsiihdslrtpsdeiakssfgvlttenrriwanvrhnlmsttnnkvnhevlsiidsalfvlcfddivindlselskmlcglsildngiqvgtctnrwydklqiivtknakaginfehtgvdghtvlrfvsdiytdsilsfansinsnapslwntttnykeldgedlitvprklewelttpdlslairfgetrlsdlinqnefrhlefknygstqikkmkfspdafvqmafqatyyalygkvectyepeddddddaskdqpsstelsrddtiiaqtngdstddddsttvvghganhhhnlksadllktipeifadngwdklnntiistsncgnpslrlfgfpvsangfgigyilkddsisicasskhrqtqrflvtlnsylleiqniwkqaqkmeklkaelakavedakpkgqksgeaegiiepkpnnlstllggyyfdmgdediksrgqspeppflhragsgfsireigkklrlseypkl	13
CRC1 (*Candida viswanathii* ATCC20336)	mddvdsaladnvksfaaggfggicavltghpfdlvkvrlqtglykssvqcvketiakdgllfglyrgvlppllgvtpmfavsfwgydvgkklvssftgksvdkfeikdistagfisaipttlvaapfervkvmmqiqegaksksmgavvaemyrtggirsifkgtvatlardgpgsalyfatyewvkkeltapgedlslfaittaggfagiamwlgvfpidtikstqqssnvkvsivqatkniyakggikaffpgvgpalarafpanaatflgvelarkfldkvi	14
ACH1 (*Candida viswanathii* ATCC20336)	Msailkqrvryapylkklrtaeqcvdlfknggylgwsgftgvgapkavpaalvdhveknklqggmgfhlfvgasagpeesrwaennmllsrsphqvgkpiaaaindgrtqffdkhlsmfpqdltygfytknkpngsnldytiieataitedgaiipgpavgaspemlsvsdniiievntktpsfegihdidlpvnppfrqpyphttadyrigrtaipvdpekviaivettegdkvppntpsdaqsqaiaghlieflenevkqgrlpenlhplqsgignianavveglaqsnfknltwwtevlqdsfldffesgsldfatatsirlteegfkkfyenwdvyskklclrsqvvsnspeiirrlgvlafntpvevdiyahanstnvmgsrmlnglggsgdflmaklsvmhtpsarptkvdptgvscivpmathvdqtehdldvvvteggladlrglspkeraqviinncahpdyqaqlqdyfdravfyatkkkmlhephmlneafamhlnlqengtmklnk	15

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1685 of 3077

Table 18

Gene (Organism)	Amino acid sequence	SEQ ID NO
ACH1$^{\Delta mis+pts}$ contained in pAA1846	Mapylkklrtaeqcvdlfknggylgwsgftgvgapkavpaalvdhvekn klqgqmgfhlfvgasagpeesrwaennmllsrsphqvgkpiaaaindgrtqffdkhl smfpqdltygfytknkpngsnldytiieataitedgaiipgpavgaspemlsvsdniiievntktpsfegihdidlpvnppfrqpyphttadyrigrtaipvdpekviaivet tegdkvppntpsdaqsqaiaghlieflenevkqgrlpenlhplqsgignianavveglaqsnfknltwtevlqdsfldffesgsldfatatsirlteegfkkfyenwdvys kklclrsqvvsnspeiirrlgvlafntpvevdiyahanstnvmgsrmlnglggsgdflmaklsvmhtpsarptkvdptgvscivpmathvdqtehdldvvteqgladl rglspkeraqviinncahpdyqaqlqdyfdravfyatkkkmlhephmlneafamhlnlqengtmklnkgrrakl	16
ACC1 contained in pAA245 including 5' intron sequence (*Candida viswanathii* ATCC20336)	mrcqvsqpsrftnllvhrlprtllnypvvntlfiprrhysInfsfknllkkmtdlspsptdslnytqlhsslpshflggnsvltaepsavtdfvkthqghtvitkvliannngigavk eirsvrkwayetfgderaiqfvamatpedmeanaeyirmadqfvevpggtnnnnyanvdliveiaertdvhavwagwghasenpllperlaaspkkivfigpppgs amrslgdkisstivaqhakvpcipwsgtgveevhvdpetklvsvddhvyakgcctspedglekakrigfpvmvkaseggggkgirkvdhekdfislynqaaneip gspifimklagdarhlevqlfadqygtnisifgrdcsvqrrhqkiieeapvtiankdtfvemekaavrlgklvgyvsagtveylysyaedkfyfleInprlqvehpttemvs gvnlpaaqliqiamglpmhriridirllygvdphsateidfefkpsnsilqrkpapkghctacritsedpgegfkpsggtlheInfrsssnvwgyfsvanqssihsfadsqf ghifafgenrqasrkhmivalkelsirgdfrttveyilklletpdfadnitttgwldeiitkkltaerpdpivavvcgavtkahiqaeedkkeyieslekggvpnkslIktifpvef iyegerykflatkssedkytlflngsrcvigarslsdggllcaldgkshsvyvvkeeaaatrlsvdgktcllevendptqlrtpspgklvkylvesgehvdaggqsyaevev mkmcmpliaqengtvqllkqpgstlnagdilialalddpskvkhakpyegtlpemgdptvtgskpahlfqhydtilknilagydnqvilnstlknmmeilknkelpyse wrlqisalhsrippkldealtsliertesrgaefparqilklvnktlgepgnellgdvvaplvsianryqnglveheydyfaslvneycnvehffsgenvreedvilrlrdenk sdlkkvisiclshsrvsaknnlilaileayepllqsnsstavairdslkkivqldsracakvglkarelliqcslpsikersdqlehilrsavvetsygevfakhrepkleiiqevv eskhvvfdvlsqflvhqdcwvaiaaaevyvrrsyraydlgkidyhihdrlpivewkfklaqiagsrynavqsasvgddsttmkhaasvsdlsfvvdskseststrtgvlv parhlddvdeilsaaleyfqpsdalsfqakgerpellnlvinivitdidgysdedeclkriheilneyeddlvfagvrnvtfvfahqvgsypkyyftfgpvyeenkvirhiepal afqlelgrlanfdikpiftnnmihvyeaigknapsdkrfftrgiirggvlkdeisIteyliaesnrlisdildtlevidtsnsdlnhifinfsnvfnvqpadveaafasflerfgrrlwr lrvtgaeirivctdpqgnsfplraiinnvsgyvvkselylevknpkgdwvfksighpgsmhlqpistpypvkeslqpkryrahnmgttfvydfpelfrqatisqwkkhgk kapkdvftslelitdendalvaverdpgankigmvgfkvtaktpeyprgrsfiivandithkigsfgpdedeyfnkctdlarklgvpriylsansgarigvaeeliplyqva wneegnpdkgfrylylnpdakealekdgkgdtivterivedgqerfvikaiigaenglgveclkgsgliagatsrayrdlfitiltvtcrsvgigaylvirlggraiqiegqpiiltg apainkllgrevyssnlqlggtqimynngvshltasddlagvekimewlsyvpakrgmpvpilesedtwdrdidyyppkqeafdinwmiegkqvegeefesglfdk gsfqetlsgwakgvvvgrarlggipigvigvetrtienmipadpanpssteaIiqeaggvwypnsafktaqaindfnngeqlplmilanwrgfsggqrdmynevlky gsfivdalvdfkqpiftyippngelrggswvvvdptinsdmmemyadvdsragvlepegmvgikyrrdkllatmqrldptyaqlkeklndsslspeehaqvstkivkr ekallpiyaqisvqfadlhdrsgrmmakgvirkeikwvdarrffwrlrrrlneeyvlkligeqvknanklekvarlkswmptvdyddddqavstwieenhaklqkrveel rqeknksdivkllqedpsnaasvmrdfvdrlsdeekekflksln	17
ACC1 contained in pAA1910 (*Candida viswanathii* ATCC20336)	mtdlspstdslnytqlhssIpshflggnsvltaepsavtdfvkthqghtvitkvliannngigavkeirsvrkwayetfgderaiqfvamatpedmeanaeyirmadqf vevpggtnnnnyanvdliveiaertdvhavwagwghasenpllperlaaspkkivfigpppgsamrslgdkisstivaqhakvpcipwsgtgveevhvdpetklvsv ddhvyakgcctspedglekakrigfpvmvkaseggggkgirkvdhekdfislynqaaneipgspifimklagdarhlevqlfadqygtnisIfgrdcsvqrrhqkiiee apvtiankdtfvemekaavrlgklvgyvsagtveylysyaedkfyfleInprlqvehpttemvsgvnlpaaqliqiamglpmhriridirllygvdphsateidfefkspnsl itqrkpapkghctacritsedpgegfkpsggtlheInfrsssnvwgyfsvanqssihsfadsqfghifafgenrqasrkhmivalkelsirgdfrttveyliklletpdfadnt ittgwldeiitkkltaerpdpivavvcgavtkahiqaeedkkeyieslekgqvpnksIIktifpvefiyegeryktatkssedkytlflngsrcvigarsldggllcaldgks hsvywkeeaaatrlsvdgktcllevendptqlrtpspgklvkylvesgehvdaggsyaevevmkmcmpliaqengtvqllkqpgstlnagdilialalddpskvkha	18

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1686 of 3077

Table 18

Gene (Organism)	Amino acid sequence	SEQ ID NO
	kpyegtlpemgdptvtgskpahlfqhydtilknilagydnqvilnstlknmmeilknkelpysewrlqisalhsrippkldealtsliertesrgaefparqilklvnktlgep gnellgdvvaplvsiannryqnglveheydyfaslvneycnvehffsgenvreedvilrlrdenksdlkkvisiclshsrvsaknnlilaileayepllqsnsstavairdslk kivqldsracakvglkarelliqcslpsikersdqlehilrsavvetsygevfakhrepkleiiqevveskhvvfdvlsqflvhqdcwvaiaaaevyvrrsyraydlgkidy hihdrlpivewkfklaqiagsrynavqsasvgddsttmkhaasvsdlsfvvdskseststrgtvlvparhlddvdeilsaaleyfqpsdalsfqakgerpellnvlnivitdi dgysdedeclkriheilneyeddlvfagvrrvtfvfahqvgsypkyytftgpvyeenkvirhiepalafqlelgrlanfdikpiftnnrnihvyeaigknapsdkrfftrgiirg gvlkdeisiteyliaesnrlisdildtlevidtsnsdlnhifinfsnvfnvqpadveaafasflerfgrrlwrlrvtgaeirivctdpqgnsfplraiinnvsgyvvkselylevknp kgdwvfksighpgsmhlqpistpypvkeslqpkryrahnmgttfvydfpelfrqatisqwkkhgkkapkdvftslelitdendalvaverdpgankigmvgfkvtakt peyprgrsfiivandithkigsfgpdedeyfnkctdlarklgvpriylsansgarigvaeeliplyqvawneegnpdkgfrylylnpdakealekdgkgdtivterivedg qerhvikaiigaenglgveclkgsgliagatsraryrdiftitlvtcrsvgigaylvrlgqraiqieggpiiltgapainkllgrevyssnlqlggtqimynngvshltasddlagv ekimewlsyvpakrgmpvpilesedtwdrdidyyppkqeafdirwmiegkqvegeefesglfdkgsfqetlsgwakgvvgrarlggipigvgvetrtienmipad panpsstealiqeaggvwypnsafktaqaindfnngeqlplmilanwrgfsgggrdmynevlkygsfivdalvdfkqpiftyipppngelrggswvvvdptinsdmm emyadvdsragvlepegmvgikyrrdkllatmqrldptyaqlkeklndsslspeehaqvstkivkrekallpiyaqisvqfadlhdrsgrmmakgvirkeikwvdarr fffwrlrrrlneeyvlkligeqvknanklekvarlkswmptvdyddddqavstwieenhaklqkrveelrqeknksdivkllqedpsnaasvmrdfvdrlsdeekeklk sln	19
ACC1 contained in pAA1907 (S1153A)	mtdlspstdslnytqlhsslpshflggnsvltaepsavtdfvkthqghtvitkvliannggigavkeirsvrkwayetfgderaiqfvamatpedmeanaeyirmadqf vevpggtnnnnyanvdliveiaertdvhavwagwghasenpllperlaaspkkivfigpppgsamrslgdkisstivaqhakvpcipwsgtgveevhvdpetklvsv ddhvyakgcctspedglekakrigtfpvmvkasegggggkgirkvdhekdfislynqaaneipgspifimklagdarhlevqlfadqygtnislfgrdcsvqrrhqkiiee apvtiankdtfvemekaavrlgklvgyvsagtveylysyaedkfyflelnprlqvehpttemvsgvnlpaaqlqiamglpmhrirdirllygvdphsateidfefkspnsl itqrkpapkghctacritsedpgegfkpsggtlhelnfrsssnvwgyfsvanqssihsfadsqfghifafgenrqasrkhmivalkelsirgdfrttveyliklletpdfadnt ittgwldelitkkltaerpdpivavvcgavtkahiqaeedkkeyieslekgqvpnksllktifpvefiyegeryktfatkssedkytlfingsrcvigarslsdggllcaldgks hsvywkeeaaatrlsvdgktcllevendptqlrtpspgklvkylvesgehvdagqsyaevevmkmcmpliaqengtvqllkqpgstlnagdlialalddpskvkha kpyegtlpemgdptvtgskpahlfqhydtilknilagydnqvilnstlknmmeilknkelpysewrlqisalhsrippkldealtsliertesrgaefparqilklvnktlgep gnellgdvvaplvsiannryqnglveheydyfaslvneycnvehffsgenvreedvilrlrdenksdlkkvisiclshsrvsaknnlilaileayepllqsnsstavairdslk kivqldsracakvglkarelliqcslpsikersdqlehilrsavvetsygevfakhrepkleiiqevveskhvvfdvlsqflvhqdcwvaiaaaevyvrrsyraydlgkidy hihdrlpivewkfklaqiagsrynavqsasvgddsttmkhaaavsdlsfvvdskseststrgtvlvparhlddvdeilsaaleyfqpsdalsfqakgerpellnvlnivitdi dgysdedeclkriheilneyeddlvfagvrrvtfvfahqvgsypkyytftgpvyeenkvirhiepalafqlelgrlanfdikpiftnnrnihvyeaigknapsdkrfftrgiirg gvlkdeisiteyliaesnrlisdildtlevidtsnsdlnhifintsnvfnvqpadveaafasflerfgrrlwrlrvtgaeirivctdpqgnsfplraiinnvsgyvvkselylevknp kgdwvfksighpgsmhlqpistpypvkeslqpkryrahnmgttfvydfpelfrqatisqwkkhgkkapkdvftslelitdendalvaverdpgankigmvgfkvtakt peyprgrsfiivandithkigsfgpdedeyfnkctdlarklgvpriylsansgarigvaeeliplyqvawneegnpdkgfrylylnpdakealekdgkgdtivterivedg qerhvikaiigaenglgveclkgsgliagatsraryrdiftitlvtcrsvgigaylvrlgqraiqieggpiiltgapainkllgrevyssnlqlggtqimynngvshltasddlagv ekimewlsyvpakrgmpvpilesedtwdrdidyyppkqeafdirwmiegkqvegeefesglfdkgsfqetlsgwakgvvgrarlggipigvgvetrtienmipad panpsstealiqeaggvwypnsafktaqaindfnngeqlplmilanwrgfsgggrdmynevlkygsfivdalvdfkqpiftyipppngelrggswvvvdptinsdmm emyadvdsragvlepegmvgikyrrdkllatmqrldptyaqlkeklndsslspeehaqvstkivkrekallpiyaqisvqfadlhdrsgrmmakgvirkeikwvdarr fffwrlrrrlneeyvlkligeqvknanklekvarlkswmptvdyddddqavstwieenhaklqkrveelrqeknksdivkllqedpsnaasvmrdfvdrlsdeekeklk	

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1687 of 3077

Table 18

Gene (Organism)	Amino acid sequence	SEQ ID NO
	sln	
ACS1 (*Candida viswanathii* ATCC20336)	Mpestqqshisldhekmqapptgfterstakpnladfdtykklykqsvenpnefteqankndwfkpfdlarfpvdpkddfkngdlpawfinggqlnasynavdrw aiknpdkpaiiyeggdepdsgriitygellkevskvlaqtlklgvkkgdsvavylpmipeaivtllaivrigavhsvvfagfssaslrdrildadsrivitadeskrggktietkki vddalkecpkvrnvivfkrtgnshvpfsagrdlwwheemakyggyfppvpvnsedpiflytysgstgkpkgvqhntagyllgailttkytfdvheedvlftagdvgwitg htycvygpllcgattvvfegtpaypnysryweivdqykvnqfyvaptalrllkragtkyvekydlsslrvlgsvgepiaaevwhwyndnigrgkahivdtywqtesgsh lltplagvtptkpgsaslpffgvvpkildpttgeelegndvegvlaiksawpsitrgiyndynrfietylapyhdhyfsgdgaardndgfywilgrvddvvnvsghrlstaei eaaliehplvaesavvgyaddltgqavaayvslkkdkvigddmeavkkelillvrkeigpfaapkllllvddlpktrsgkimrrilrkvlageedqlgdistlsnpgvvqqii eivhackk	20
ACS2 (*Candida viswanathii* ATCC20336)	Mttdsnthkvvheangvklrqtpkeffqrqpnkghiqdteeykkllyeqsikdpqgffgplakellswdsdfntvksgtlkngdaawfiggklnasyncvarhalanpn kpaiiyeadeerdsyiltygdllrevskvagvlhswgirkgdtvavylpmnaqaliamlaiarlgaahsvifagftssgsikdrvndasckalitcdegrrggrttnikklcd ealkqcptvkhvlvhrrtgnpeiktlegrdyywdeetaktsgyfppvpvdsedpiflytysgstgtpkgvvhttagyllgaalstkyifdvhpedvfftagdvgwitghtyal ygplllgvpsvvfegtpaypdygrfwqivekhkathfyvaptalrllrkageqevgkydlssirtlgsvgepispdiwewynefvgkdqchisdtywqtesgshliapla gaiankpgsasypffgieaalidpvtgveikgndvegvlvvkdhwpsmartvfnnhvkymdtymnpypgyyftgdgaardndgyywirgrvddvvnvsghrlst aeiesaliedprvgesavvginddltgqaviayvalksgdvadedalrkelillvrkeigpfaapksvilvqdlpktrsgkimrrilrkvssneadqlgdittlqnpgsveg misafgaqfgrk	21
TES1-1 (*Candida viswanathii* ATCC20336)	Mlttlsgpnplpdfeealrvikvddthyvgahslrlpvkgggrgvyggghmiaqsllvgiestrddktnkvfipdsyhsyfigagnakipmnytveklyddenvskrfiiaeq kgthrltclvtlrrpgtkpfhdsdnldisipvpkiqlkhpdpdklhqvqhtdfirnafgkelmdyrecpeenelyaaerwltvftgirnqpkpgasletvveelpdaqggqm htveksilrpkdsqsfkdpiynfvgladlsdsafittmarilhipwapsieiddtydpardatyimrstlnaahifhynamsldhhiyfhnedytsddgsgfdickdwlaft yqmkrlsnnrtlvrgfffnekhkciatvvqegltimqngvgrtadksrl	22
TES2-1 (*Candida viswanathii* ATCC20336)	Mienisgngnypqnhevdlekefgvekiginlyrgkspipkpdrrsrgayggylaggallvamkstppeyrphsfhsyfikavndketlewrveetsngrnyanrslq afqagnlvytanvsltkknsakkaeeatgvkpfefqgkphewfekhkrddlplatpsssliiyhkffpevvsleaskeeeskpaadrelswyfkwginneeghhqpl vnlnsdyqyvgmaaltdavylnrllrilrvedadhtqlvhyfsvsldhtmyfhdddfdvtkwmgfflkvtrfshnralcqgevyndkgvhvctivqeglmmlngleqga kl	23
TES2-2 (*Candida viswanathii* ATCC20336)	Miesisengnypqnhevdlekefgvekiginlyrgkspipkpdrrsrgayggylaggallvamkstppdyrphsfhsyfikavndkvtlewrveetsngrnyanralq afqagdlvytanvsltkknsakktteeatgvkpfefqgkphewfdkhkiddlplatpsssliiyhkffpevvsleaskeeeskpaadrelswyfkwginneeghhqplv nlnsdyqyvgmaaltdavylnrllrilrvqdadhtqlvhyfsvsldhtmyfhdddfdvtkwmgfflkvtrfshnralcqgevyndkgvhvctivqeglmmlngleegakl	24

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1688 of 3077

Table 18

Gene (Organism)	Amino acid sequence	SEQ ID NO
TES3-1 (*Candida viswanathii* ATCC20336)	Mshptpeevygvtkvaenkyvgnrplnkptpktrgvyggnfcaqailvaiesapegftphsihsnfirggdpevpvewevevisngksfanrivkgvqhgivvyvat vsltnknsttrnesftydtppdetvktygnaeldtyyqgwlylevknypkqlhshqisysvkwgpendawkdasqtyqfvglaaisdvldlggilrnldihlstpkfnvsl dhsvyfhgadfdvtrwstttirmtklahgraliegemysdqgrhiasivqerlyiaespkl	25
TES4-1 (*Candida viswanathii* ATCC20336)	Mtsiapatdipavdistiydvkqidannyrgvrplqkqarefrgvfggnlvaqsvvvalrsvpagfhpnsvhayyvravtdetpieweveetttgrtfanrsirglqnkkv vftasisltrknsnaeviqktghpslqfqrdtepyyeqmrarpgeckvlyinanthltvrqfpevssrdvfsflvrfgndgrerivgmtpeyqyaaalaalsdwvrlrfyfdn mgvdvqtsfdvsldhsiyfhddgfdateylvfsvkvsrishsrvlydggifndrgahvasirqerlyvvsnkpkf	26
TES5-1 (*Candida viswanathii* ATCC20336)	Mptfnykdgetidvqkefgvvetapnkyvgvkplvkpmphvkgvfggnlaggallvamksvgpdfsphslhsyfiragsdqtpvewtvqaisdgnsfcnrflkgvq ngqviyianvsltkrnsaadamkkyeeyhaqirqkgkdgdadeeddddddednapakpfgfqtpshkwikdrdldklplpvsdmesnlllyyklppefvslksste eeslpvserrmgalakwgienegqfmqpltnldksfqyvglanitdglylgtlnrilriddltlderatnyfsvsldhviyfhdddfdvtkwmgtfrcsryshnrvifegeiys dkgvqvasiiqeglvrfkdgylknakl	27
TES6-1 (*Candida viswanathii* ATCC20336)	Msttgiynakpaaailetktkvklvytegpklvyeglhpvetiketvrgtyggdfiaqgvnaawesignrmdfqphslhayyiksgsqesvlrwevfkvsdsrsfanr mltayqthtnqlvftmqisftkdnneeikraeyeklasggkirsipfvikkapnekyfklrdrinelhyseytngiiarnaspldfheyatemnydtlgnkeygtfmkildn yslgkdyskqsflglsflsdgwwvfcftralglplgsyekeffrvsldhtvyfhdanfdssewlfldfrfvnmknnrfltvlnyytlqgkliatvlqevytslhqgiidksqeiavk sgnkkqvttpkl	28
TES7-1 (*Candida viswanathii* ATCC20336)	Mdklqaevyeaeppvakleaktaaklissdgakltyegvypvelvrkglrgtyggdfiagginvawesignktdfqphslhayfvkagsdssvlrwevlkvsdsrnfa nrlmlayqthtnelvftmqisftkdnneeikraeykqllqsggkirsipfaikkppnekyfklkdkvddlpyfehtngnmataippdfleyatemnhdtvgnkefgifmk vlddyslgknyergsflglaflsdawlssftpalglplgtlerkffrvsldhtmyfhdanfdssewifvdfrfvnlnnnrllgvvnfytlqgklvatviqeaymflhqaiidksq eiaeksghkkqvitpkl	29
TES7-2 (*Candida viswanathii* ATCC20336)	Mdklqaevydaeppvakleaktaaklisadgakliyegvypvelvrkglrgtyggdfiagginvawesignktdfqphslhayfvkagsdqsvlrwevlkvsdsrsfa nrlmlayqthtnqlvftmqisftkdnneeikraeykqllesggkirsipfaikkppnekyfklkdkvddlpyfehtngnmataippdfleyatemnhdtvgnkefgifmk vlddysldknyergsflglaflsdawlssftpalglplgtferkffrvsldhtmyfhdanfdssewifvdfrfvnlnnnrllgvvnfytlqgklvatviqeaymflhqgiidqsq eiaeksghkkqvitpkl	30

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1689 of 3077

Table 18

Gene (Organism)	Amino acid sequence	SEQ ID NO
TES8-1 (*Candida viswanathii* ATCC20336)	Merlqaevydakptvakletkttvklvaqdgkrlvyeaiypvehikegipgayggdtlaqgmnaawdslgdkkdfqphsvhsyfvkpatnksilrwevikvsdgrsf anrfvsgyqthnnalvftmqisftkyndevvkiaeynkllesggkirsipfvikkapnekyfkfkdnlgdlryiehtngniatamsedlfeyatemnhdtlgnqefgifmk vldnyslgsdytkqsylglaflsdaiwmsvcsralglpfgsyhrqffgvsmdhsmyfhdanfdstewifldfrfvnlkndrllgvanfytlegklistliqeayiflhpgiiaks qeiavksgnkrqvkmpkl	31
TES3-1Δpts Contained in pAA1609	Mshptpeevygvtkvaenkyvgnrplnkptpktrgvyggnfcaqailvaiesapegftphsihsnfirggdpevpvevevevisngksfanrivkgvqhgivvyvat vsltnknsttrnesftydtppdetvktygnaeldtyyqgwlylevknypkqlhshqisysvkwgpendawkdasqtyqfvglaaisdvidlggilrnldihlstpkfnvsl dhsvyfhgadfdvtrwsttiirmtklahgraliegemysdqgrhiasivqerlyiaes	32
PEX11 (*Candida viswanathii* ATCC20336)	rnvadslvyhptvsklvkfldttpkrekvfrllsylsrflgyyayrkgyskетialfanlkgnfffirkamrflkpinhlqlaskaydnklldpvlqittiirnlayagyttidgviffkllgl idakkfpnlatyasrfwligliaglinsliriiyslkdyehqegdketdakaintklyaakrklvwdlldtfialnsidilhftegdvgfagtitsllgledlwkat	33
CPRB (*Candida viswanathii* ATCC20336)	MALDKLDLYVIITLVVAVAAYFAKNQFLDQPQDTGFLNTDSGSNSRDVLSTLKKNNKNTLLLFGSQTGTAEDYANKLS RELHSRFGLKTMVADFADYDWDNFGDITEDILVFFIVATYGEGEPTDNADEFHTWLTEEADTLSTLRYTVFGLGNSTY EFFNAIGRKFDRLLSEKGGDRFAEYAEGDDGTGTLDEDFMAWKDNVFDALKNDLNFEEKELKYEPNVKLTERDDLS AADSQVSLGEPNKKYINSEGIDLTKGPFDHTHPYLARITETRELFSSKERHCIHVEFDISESNLKYTTGDHLAIWPSNS DENIKQFAKCFGLEDKLDTVIELKALDSTYTIPFPTPITYGAVIRHHLEISGPVSRQFFLSIAGFAPDEETKKTFTRLGGD KQEFATKVTRRKFNIADALLYSSNNTPWSDVPFEFLIENIQHLTPRYYSISSSSLSEKQLINVTAVVEAEEEADGRPVT GVVTNLLKNIEIAQNKTGEKPLVHYDLSGPRGKFNKFKLPVHVRRSNFKLPKNSTTPVILIGPGTGVAPLRGFVRERV QQVKNGVNVGKTLLFYGCRNSNEDFLYKQEWAEYASVLGENFEMFNAFSRQDPSKKVYVQDKILENSQLVHELLTE GAIIYVCGDASRMARDVQTTISKIVAKSREISEDKAAELVKSWKVQNRYQEDVW	34
FAA1 (*Candida viswanathii* ATCC20336)	MGAPLTVAVGEAKPGETAPRRKAAQKMASVERPTDSKATTLPDFIEECFARNGTRDAMAWRDLVEIHVETKQVTKII DGEQKKVDKDWIYYEMGPYNYISYPKLLTLVKNYSKGLLELGLAPDQESKLMIFASTSHKWMQTFLASSFQGIPVVT AYDTLGESGLTHSLVQTESDAVFTDNQLLSSLIRPLEKATSVKYVIHGEKIDPNDKRQGGKIYQDAEKAKEKILQIRPDI KFISFDEVVALGEQSSKELHFPKPEDPICIMYTSGSTGAPKGVVITNANIVAAVGGISTNATRDLVRTVDRVIAFLPLAHI FELAFELVTFWWGAPLGYANVKTLTEASCRNCQPDLIEFKPTIMVGVAAVWESVRKGVLSKLKQASPIQQKIFWAAF NAKSTLNRYGLPGGGLFDAVFKKVKAATGGQLRYVLNGGSPISVDAQVFISTLLAPMLLGYGLTETCANTTIVEHTRF QIGTLGTLVGSVTAKLVDVADAGYYAKNNQGEIWLKGGPVVKEYYKNEEETKAAFTEDGWFKTGDIGEWTADGGLN IIDRKKNLVKTLNGEYIALEKLESIYRSNHLILNLCVYADQTKVKPIAIVLPIEANLKSMLKDEKIIPDADSQELSSLVHNKK VAQAVLRHLLQTGKQQGLKGIELLQNVVLLDDEWTPQNGFVTSAQKLQRKKILESCKKEVEEAYKSS	35

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1690 of 3077

Table 18

Gene (Organism)	Amino acid sequence	SEQ ID NO
POX5 (*Candida viswanathii* ATCC20336)	MPTELQKERELTKFNPKELNYFLEGSQERSEIISNMVEQMQKDPILKVDASYYNLTKDQQREVTAKKIARLSRYFEHE YPDQQAQRLSILGVFDPQVFTRIGVNLGLFVSCVRGNGTNSQFFYWTINKGIDKLRGIYGCFGMTELAHGSNVQGIE TTATFDEDTDEFVINTPHIGATKWMIGGAAHSATHCSVYARLKVKGKDYGVKTFVVPLRDSNHDLEPGVTVGDIGAK MGRDGIDNGWIQFSNVRIPRFFMLQKYCKVSRLGEVTMPPSEQLSYSALIGGRVTMMMDSYRMTSRFITIALRYAIH RRQFKKKDTDTIETKLIDYPLHQKRLFPFLAAAYLFSQGALYLEQTMNATNDKLDEAVSAGEKEAIDAAIVESKKLFVA SGCLKSTCTWLTAEAIDEARQACGGHGYSSYNGFGKAYSDWVQCTWEGDNNILAMNVAKPMVRDLLKEPEQKG LVLSSVADLDDPAKLVKAFDHALSGLARDIGAVAEDKGFDITGPSLVLVSKLNAHRFLIDGFFKRITPEWSEVLRPLGF LYADWILTNFGATFLQYGIITPDVSRKISSEHFPALCAKVRPNVVGLTDGFNLTDMMTNAAIGRYDGNVYEHYFETVK ALNPPENTKAPYSKALEDMLNRPDLEVRERGEKSEEAAEILSS	36
POX5(F98G) (*Candida viswanathii* ATCC20336 mutant)	MPTELQKERELTKFNPKELNYFLEGSQERSEIISNMVEQMQKDPILKVDASYYNLTKDQQREVTAKKIARLSRYFEHE YPDQQAQRLSILGVFDPQVGTRIGVNLGLFVSCVRGNGTNSQFFYWTINKGIDKLRGIYGCFGMTELAHGSNVQGIE TTATFDEDTDEFVINTPHIGATKWMIGGAAHSATHCSVYARLKVKGKDYGVKTFVVPLRDSNHDLEPGVTVGDIGAK MGRDGIDNGWIQFSNVRIPRFFMLQKYCKVSRLGEVTMPPSEQLSYSALIGGRVTMMMDSYRMTSRFITIALRYAIH RRQFKKKDTDTIETKLIDYPLHQKRLFPFLAAAYLFSQGALYLEQTMNATNDKLDEAVSAGEKEAIDAAIVESKKLFVA SGCLKSTCTWLTAEAIDEARQACGGHGYSSYNGFGKAYSDWVQCTWEGDNNILAMNVAKPMVRDLLKEPEQKGL VLSSVADLDDPAKLVKAFDHALSGLARDIGAVAEDKGFDITGPSLVLVSKLNAHRFLIDGFFKRITPEWSEVLRPLGFL YADWILTNFGATFLQYGIITPDVSRKISSEHFPALCAKVRPNVVGLTDGFNLTDMMTNAAIGRYDGNVYEHYFETVKA LNPPENTKAPYSKALEDMLNRPDLEVRERGEKSEEAAEILSS	37
POX5(W429F) (*Candida viswanathii* ATCC20336 mutant)	MPTELQKERELTKFNPKELNYFLEGSQERSEIISNMVEQMQKDPILKVDASYYNLTKDQQREVTAKKIARLSRYFEHE YPDQQAQRLSILGVFDPQVFTRIGVNLGLFVSCVRGNGTNSQFFYWTINKGIDKLRGIYGCFGMTELAHGSNVQGIE TTATFDEDTDEFVINTPHIGATKWMIGGAAHSATHCSVYARLKVKGKDYGVKTFVVPLRDSNHDLEPGVTVGDIGAK MGRDGIDNGWIQFSNVRIPRFFMLQKYCKVSRLGEVTMPPSEQLSYSALIGGRVTMMMDSYRMTSRFITIALRYAIH RRQFKKKDTDTIETKLIDYPLHQKRLFPFLAAAYLFSQGALYLEQTMNATNDKLDEAVSAGEKEAIDAAIVESKKLFVA SGCLKSTCTWLTAEAIDEARQACGGHGYSSYNGFGKAYSDWVQCTWEGDNNILAMNVAKPMVRDLLKEPEQKG LVLSSVADLDDPAKLVKAFDHALSGLARDIGAVAEDKGFDITGPSLVLVSKLNAHRFLIDGFFKRITPEWSEVLRPLGF LYADWILTNFGATFLQYGIITPDVSRKISSEHFPALCAKVRPNVVGLTDGFNLTDMMTNAAIGRYDGNVYEHYFETVK ALNPPENTKAPYSKALEDMLNRPDLEVRERGEKSEEAAEILSS	38
POX4 (*Candida viswanathii* ATCC20336)	MTFTKKNVSVSQGPDPRSSIQKERDSSKWNPQQMNYFLEGSVERSELMKALAQQMERDPILFTDGSYYDLTKDQQ RELTAVKINRIARYREQESIDTFNKRLSLIGIFDPQVGTRIGVNLGLFLSCIRGNGTTSQLNYWANEKETADVKGIYGCF GMTELAHGSNVAGLETTATFDKESDEFVINTPHIGATKWMIGGAAHSATHCSVYARLIVDGQDYGVKTFVVPLRDSN HDLMPGVTVGDIGAKMGRDGIDNGWIQFSNVRIPRFFMLQKFCKVSAEGEVTLPPLEQLSYSALLGGRVMMVLDSY RMLARMSTIALRYAIGRRQFKGDNVDPKDPNALETQLIDYPLHQKRLFPYLAAAYVISAGALKVEDTIHNTLAELDAAV EKNDTKAIFKSIDDMKSLFVDSGSLKSTATWLGAEAIDQCRQACGGHGYSSYNGFGKAYNDWVVQCTWEGDNNVL	39

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1691 of 3077

Table 18

Gene (Organism)	Amino acid sequence	SEQ ID NO
	AMSVGKPIVKQVISIEDAGKTVRGSTAFLNQLKDYTGSNSSKVVLNTVADLDDIKTVIKAIEVAIIRLSQEAASIVKKESF DYVGAELVQLSKLKAHHYLLTEYIRRIDTFDQKDLVPYLITLGKLYAATIVLDRFAGVFLTFNVASTEAITALASVQIPKL CAEVRPNVVAYTDSFQQSDMIVNSAIGRYDGDIYENYFDLVKLQNPPSKTKAPYSDALEAMLNRPTLDERERFFEKSD ETAAILSK	
PXA1 (*Candida viswanathii* ATCC20336)	Mvniskltgynkqdimvvllqefvktykdnkiklnyssrpvilfstlvatagigvfftlrsivtkyneyllnkrlrrpslirqssnilkngsreifiqkgngkvtriiipkanndqya adkylykdfarneqilqqkgrffnsrfinqltiiwkilipkfycqntsllsqcfflifftwlslliakldgqivknliaadgrkfardliyfilfiafpasytnaaikylelrlalgfrtnltryih dmyldktmsyykvglngadiqnidqyitedvtkfcmslcslfssmgkpfidliffsvylrdnlgtgaiigifanyfataimlkkatprfgklaakrthlegvyfnqqlnimtns eeigfykgskiekskiaenfdklmghvsreinlsssyaaledyvlkytwsawgyifsglpvfldvlfpkedpssghiadidddhahghghtgeetssttenmktfvtn krlllsladagsrlmvslkevttltgitnrvfmmltqlhrvhdpkfdygdkyglpdingtyqlnydglrlehvpitvptaegsystplipdltfdikgknllfvgpngsgktsvarvl aglwplyaglvskpsdlfnpqksyfttgslrdqvvypnrsenttndqlfhillhcvhldhivkrygllnqnldfaktlsggekqrlsfarvlfnrpsivilddstsalspdmeelm yqvlqdhkinyvtlsnrpslskfhdkvfei	40
PXA2 (*Candida viswanathii* ATCC20336)	Mtvenaklqknslavsllkvyksnrsllltntsyiiiiiaaftgatntgrgtssrssakvetdeeqsvkkkhpklsresfhrlrkailptffdrtivyffanltllvvralltlrvatldgqlv galvsrirvfakyllywmllgipaaltnallnwtksnlsksirmnlnnnimeeylpdnldpnyysiihltdnkirdpnqrittdtsrlsdalaslpghilkptldiilcaqqlsksg vgngegtlalgilahfstmiirffsppfaklaaeranlegglrsahskivanseeiaflgghdreldhidhcyytlerfskgeywkraiheitqtfivkyfwgvaglvlcsapvf iakylgepedknvagnfitnrrllmsasdsldrliysrryllqvvghatrvsdfldtlheveekkkritsnvqfnndeitfdhvrlmtptevtlipdlnfsikpgdhlilvgpngsg kssllrmlgglwpvrfgtiripntenmfylpqkaylvegsfreqiiyphnvtqqkktdqqlkeilkvlkledysgqldevkkwseelsigaqqrlamarlyyhepkfavlde ctsavspdmeqlmyqhaqglgitllsvahrpalwhfhkyllefdgkgsyyfgtldekhkmkleeeerlkkenekksvakk	41
ACL1 (*Yarrowia lipolytica* CLIB122)	rnsanenisrfdapvgkehpayelfhnhtrsfvyglqpracqgmldfdfickrenpsvagviypfgggfvtkmywgtketllpvyqqvekaaakhpevdvvvnfass rsvysstmelleypqfttiaiiaegvperrareilhkaqkkgvtiigpatvggikpgcfkvgntggmmdnivasklyrpgsvayvskgggmsnelnniishttdgvyegi aiggdrypgttfidhilryeadpkckivllgevggveeyrvieavkngqikkpivawaigtcasmfktevqfghagsmansdletakaknaamksagfyvpdtfed rnpevlaelyekrnvakgelsrisepevpkipidyswaqelglirkpaafistisddrggellyagrnpisevfkedigiggvrnslwfrrrlpdyaskflemvlrnltadhgp avsgamntiittragkdlisslvagltigtrfggaldgaatefttaydkglsprqfvdtrnrkqnklipgighrvksrnnpdfrvelvkdfvkknfpstqlldyalaveevttskk dnlilnvdgaiavsfvdlrnscgaftveetedylknglvlngtlfvlgrsigliahnldqkrlktglyrhpwdditylvgqeaiqkkrveisaggdvskaktrs	42
ACL2 (*Yarrowia lipolytica* CLIB122)	Msaksiheadgkallahfskapvwaeqqpintfemgtpklaslfedgvapeqifaaaektypwllesgakfvakpdqlikrrgkaglivlnksweeckpwiaeraa kpinvegidgvlrtflvepfvphdqkheyyinihsvregdwilfyheggvdvgdvdakaakilipvdieneypsnatltkellahvpedqhqtlldfinrlyavyvdlqftyl einplvviptaggvevhyldlagkldqtaefecgpkwaaarspaalgqvvtidagstkvsidagpamvfpapfgrelskeeayiaeldsktgaslkltvlnakgriwtlv agggasvvyadaiasagfadelanygeysgapnetqtyeyaktvldlmtrgdahpegkvlfigggianftqvgstfkgiirafrdyqssllnnhkvkiyvrrggpnwqe glrliksagdelnlpmeiygpdmhvsgivplallgkrpknvkpfgtgpsteastplgv	43
FAT1 (*Candida viswanathii* ATCC20336)	MSGLEIAAAAILGSQLLEAKYLIADDVSLAKTVAVNALPYLWKASRGKASYWYFFEQSVFKNPNNKALAFPRPRKNAP TPKTDAEGFQIYDDQFDLEEYTYKELYDMVLKYSYILKNEYGVTANDTIGVSCMNKPLFIVLWLALWNIGALPAFLNFN TKDKPLIHCLKIVNASQVFVDPDCDSPIRDTEAQIREELPHVQINYIDEFALFDRLRLKSTPKHRAEDKTRRPTDTDSS ACALIYTSGTTGLPKAGIMSWRKAFMASVFFGHIMKIDSKSNVLTAMPLYHSTAAMLGLCPTLIVGGCVSVSQKFSAT	44

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1692 of 3077

Table 18

Gene (Organism)	Amino acid sequence	SEQ ID NO
	SFWTQARLCGATHVQYVGEVCRYLLNSKPHPDQDRHNVRIAYGNGLRPDIWSEFKRRFHIEGIGEFYAATESPIATT NLQYGEYGVGACRKYGSLISLLLSTQQKLAKMDPEDESEIYKDPKTGFCTEAAYNEPGELLMRILNPNDVQKSFQGY YGNKSATNSKILTNVFKKGDAWYRSGDLLKMDEDKLLYFVDRLGDTFRWKSENVSATEVENELMGSKALKQSVVVG VKVPNHEGRACFAVCEAKDELSHEEILKLIHSHVTKSLPVYAQPAFIKIGTIEASHNHKVPKNQKLPKGEDGKDL IYWLNGDKYQELTEDDWSLICTGKAKL	
CYP52A17 (*Candida viswanathii* ATCC20336) FROM pAA1712	Mieqlleywyvvvpvlyiikqllaytktrvlmkklgaapvtnklydnafgivngwkalqfkkegraqeyndykfdhsknpsvgtyvsilfgtrivvtkdpenikailatqfgd fslgkrhtlfkpllgdgiftldgegwkhsramlrpqfareqvahvtslephfqllkkhiilkhkgeyfdiqelfffrftvdsateflfgesvhslkdesiginqddidfagrkdfaesf nkaqeylairtlvqtfywlvnnkefrdctksvhkftnyyvqkaldaspeelekqsgyvflyelvkqtrdpnvlrdqslnillagrdttaglllsfavfelarhpeiwaklreeieq qfglgedsrveeitfeslkrceylkaflnetlriypsvprnfiatknttlprgggsdgtspiliqkgeavsyginsthldpvyygpdaaefrperwfepstkklgwaylpfng gpriclgqqfalteagyvlvrlvqefshvrsdpdevyppkrltnltmclqdgaivkfd	45
ADH1-2 alcohol dehydrogenase (*Candida viswanathii* ATCC20336)	MHALFSKSVFLKYVSSPTTSAIPHSLEFIVSRSSYLRRRIPPYLPRCSHFPSFYYSSSSVYTKKSFHTMSANIPKTQKA VVFEKNGGELKYKDIPVPTPKANELLINVKYSGVCHTDLHAWKGDWPLDTKLPLVGGHEGAGVVVGMGENVKGWKI GDFAGIKWLNGSCMSCEFCQQGAEPNCGEADLSGYTHDGSFEQYATADAVQAARIPAGTDLAEVAPILCAGVTVYK ALKTADLAAGQWVAISGAGGGLGSLAVQYAVAMGLRVVAIDGGDEKGDFVKSLGAEAYIDFLKEKGIVAAVKKATDG GPHGAINVSVSEKAIDQSVEYVRPLGKVVLVGLPAGSKVTAGVFEAVVKSIEIKGSYVGNRKDTAEAVDFFSRGLIKC PIKIVGLSELPQVFKLMEEGKILGRYVLDTSK	46
FAO1 fatty alcohol oxidase (*Candida viswanathii* ATCC20336)	MAPFLPDQVDYKHVDTLMLLCDGIIHETTVDQIKDVIAPDFPADKYEEYVRTFTKPSETPGFRETVYNTVNANTTDAIH QFIILTNVLASRVLAPALTNSLTPIKDMSLEDREKLLASWRDSPIAAKRKLFRLVSTLTLVTFTRLANELHLKAIHYPGRE DREKAYETQEIDPFKYQFLEKPKFYGAELYLPDIDVIIIGSGAGAGVVAHTLANDGFKSLVLEKGKYFSNSELNFDDKD GVQELYQSGGTLTTVNQQLFVLAGSTFGGGTTVNWSACLKTPFKVRKEWYDEFGVDFAADEAYDKAQDYVWQQM GASTEGITHSLANEIIIEGGKKLGYKAKVLDQNSGGHPQHRCGFCYLGCKHGIKQGSVNNWFRDAAAHGSQFMQQV RVLQILNKKGIAYGILCEDVVTGAKFTITGPKKFVVAAGALNTPSVLVNSGFKNKNIGKNLTLHPVSVVFGDFGKDVQA DHFHNSIMTALCSEAADLDGKGHGCRIETILNAPFIQASFLPWRGSNEARRDLLRYNNMVAMLLLSRDTTSGSVSAH PTKPEALVVEYDVNKFDRNSILQALLVTADLLYIQGAKRILSPQAWWPIFESDKPKDKRSIKDEDYVEWRAKVAKIPFD TYGSPYGSAHQMSSCRMSGKGPKYGAVDTDGRLFECSNVYVADASLLPTASGANPMVTTMTLARHVALGLADSLK TKAKL	47
HPD1 3-hydroxy-propionate-dehydrogenase (*Candida*	Mlrssvrtfstqsrvlanygfvglglmgqhmarhvynqlqpadklyvhdvnpqhttqfvtdvttqkpqnatqltplsslkefttepesqldfivtmvpegkhvkavvsel vdhynasgkydpskkltfvdsstidiptsrevhqlvadklqgatfidapvsggvagarngtlsfmvsrdtkedvdpnlvtllnymgsnifpcggthgtglaaklannylla itniavadsfqlansfglnlqnyaklvststgkswasvdncpipgvypeknltcdngykggfvtklrkdvvlatesakannqflmlgevgrywydkacedekyanrdl svlfeflgdlkk	48

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1693 of 3077

Table 18

Gene (Organism)	Amino acid sequence	SEQ ID NO
viswanathii ATCC20336)		
ALD6 malonate semialdehyde dehydrogenase (*Candida viswanathii* ATCC20336)	Mlsrvlfktkprvptksitamairnksivtlsstttsytpdhttpstepyitpsfvnnefiksdsntwfdvhdpatnyvvskvpqstpeeleeaiasahaafpkwrdtsiikrq giafkfvqllrenmdriasvivleggktfvdaggdvtrglqvaeeaacnitndlkgeslevstdmetkmireplgvvgsicpfnfpamvplwslplvlvtgntavikpserv pgasmiicelaakagvppgvlnivhgkhdtvnkliedprikalttfvggdkagkyiyekgsslgkrvqanlgaknhlvvlpdahkqsfvnavngaafgaaggrcmais vlvtvgktkewvqdvikdakllntgsgfdpksdlgpvinpesltraeeiiadsvangavleldgrgyrpedarfakgnflgptiltnvkpglraydeeifapvlsvvnvdtid eaielinnnkygngvslftssggsaqyftkridvggvginvpipvplpmfsftgsrgsflgdlnfygkagitftkpktitsawktnliddeilkpstsmpvqq	49
ECI1 enoyl-CoA isomerase (*Candida viswanathii* ATCC20336)	MSDEESDILYEVRDRTAIITLNIPKRLNALNGAQYLKLGKFLERANNEEDTVLTLIQASGRFFSAGANFADNDMAKVEM SKLFSHEYWLERFVARNIWLTNLFNDHKKILAAAVNGPVIGLSTGLLLLVDLVYVHDLNKFYLLAPFANLGLVAEGASS ATLFNRLGWSKASEALLLAKPIGGQDCYNAGFINKHYDGKFSSTEEFNEHVYKELTEAFENLHDDSILQNKQLLKLSR DQAIN	50
ECI2 enoyl-CoA isomerase (*Candida viswanathii* ATCC20336)	MSDDLITYEVKDRAAVITLNNPKKLNALSIPQYDTICKLLERANAEEDTVITLLQSTGRVFSAGANADSIVGQDAELET WLNMSVAKQTFLVQTFLAHKKILAVALNGPVIGLSAAFVALCDLVYVHNAAKTFFLTPFANIGILAEGGTSATLPMRVG WSRAAEALLLSKRISGDDLQRAGFFNKDYKGQFKSAEEFNEVVLKELLDATENLHEDSIIQNKELLKAIFKPKISEVNS QEVSRGVYKWTSGVPMDRFKKLLNGELKHKL	51
DCR1 dienoyl-CoA reductase (*Candida viswanathii* ATCC20336)	MPNTLDHNYLKKSVWKSDIFAGKVIFITGGAGTICRVQAEAMVLLGANAAIIGRNVEKTEEAAKEIASLRPGAKVIGIGA VDVRKIQTIKDAVDRTVAELGRIDYVIAGAAGNFLCDFNHLSANAFKSVIDIDLLGSFNTVKVTFDQLRKNKGAVLFVS ATLHYTGVPMQSHVGAAKSGVDALSNALAVELGPLGIRFNCIAPGAIAGTEGMSRLAPPTDTPLETKIPLQRQGTTED IADATVFLFSPASSYITGDVLVVDGAMWQTGGILNDFYPDIIIHQNADPEGKL	52
DCR2 dienoyl-CoA reductase (*Candida viswanathii* ATCC20336)	MPNTLTEAYKQESSWKPDLFKGKVVFITGGAGSICRVQAEAMVLLGANAAIIGRNVEKTESAAKEIASLRSGAKVLGI GGIDVRKVDSLKSAVDKTVAELGRIDFVIAGAAGNFLCDFNHLSSNAFKSIIDIDLLGSFNTVKVTFDQLRKNKGAILFV SATLHYYGVPFQIGVGAAKAGVDALSNALAVELGPLGIRSNCIAPGPIDGTEGVERLVRASKAEAAKKVPLQRLGTTQ DIADGTVYLFSPAASFVTGDVLVVDGASWQISSGVGATNYPVSILNAIDAPKGGKL	53

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1694 of 3077

Table 18

Gene (Organism)	Amino acid sequence	SEQ ID NO
MCR malonyl-CoA reductase (*Sulfolobus islandicus*)	Mrrtlkaailgstglvgieyvrmlanhpyikvgylagkgsvgkpyeevvrwqtvgqvpkeiadmevkptdpklmddvdiifsplpqgaagpveeefakhgfpvisn spdhrfdpdvpllipeinphtislidkqrerhdwrgfivttplctaggaaiplapiymnfridssllttiqslsgagypgipsldvvdnvlplgdnydnktvkeisrilsetkrmv nddndlslgatthristihghyeviyvtfredvsvekiretldsfrgepqklklptapdkpilltnqdarpqvyfdrwagdppgmsvvvgrlsqinrrtirlvsvvhntvrgaa gggilaaellvekgyidkr	322
2PS 2-pyrone synthase (*Gerbera hybrida*)	Mgsyssddvevireagragglatilaigtatppncvaqadyadyyfrvtksehmvdlkekfkricektaikkrylaltedylqenptmcefmapslnarqdlvvtgvp mlgkeaavkaidewglpkskithlifctagvdmpgadyqlvkllglspsvkrymlyqqgcaaggtvlrlakdlaennkgsrvlivcseitailffgpnenhldslvaqalf gdgaaalivgsphlaverpifeivstdqtlpdtekamklhlreggltfqlhrdvplmvaknienaaekalsplgitdwnsvfwmvhpggraildqverklnlkedklra srhvlseygnlisacvlfildevrkrsmaegkstlgegldcgvlfgfgpgmtvetvvlrsvrvtaavangn	324

EXAMPLE 26: Nucleic acid sequences of genes referenced herein

Table 19

Gene (Organism)	Nucleic acid sequence of the gene protein-encoding sequence (from START codon to STOP codon except where otherwise noted)	SEQ ID NO
URA3 (*Candida viswanathii* ATCC 20336)	Atggttagcacaaaaacatacacagaaagggcatcagcacaccctccaaggttgcccaacgtttattccgcttaatggagtccaaaaagaccaacctctgcgcc tcgatcgacgtgaccacaaccgccgagttccttcgctcatcgacaagctcggtcccacatctgtctcgtgaagacgcacatcgatatcatctcagacttcagctac gagggcacgattgagccgttgcttgtgcttgcagagcgcccacgggttcttgatattcgaggacaggaagtttgctgatatcggaaacaccgtgatgttgcagtacacc tcgggggtataccggatcgcggcgtggagtgacatcacgaacgcgcacgcggagtgactgggaagggcgtcgttgaagggttgaaacgcggtgcggagggggta gaaaaggaaagggcgtgttgatgttggcggagttgtcgagtaaaggctcgttggcgcatggtgaatatacccgtgagacgatcgagattgcgaagagtgatcgg gagttcgtgattgggttcatcgcgcagcgggacatgggggtagagaagaaggtttgattggatcatcatgacgcctggtgtgggttggatgataaaggcgatgc gttgggccagcagtatatgggactgttgatgaggtggttctgactggtaccgatgtgattattgtcgggagagggttgtttggaaaaggaagaccctgaggtggagg gaaagagatacagggatgctggatggaaggcatacttgaagagaactggtcagttagaataa	54
P_{URA3} (*URA3* promoter from *Candida viswanathii* ATCC 20336)	cgacgggtacaacgagaattgtattgaattgatcaagaacatgatcttggttacagaacatcaagttcttggaccagactgagaatgcacagatatacaaggcgt catgtgataaaatggatgagatttatccacaattgaagaaagagtttatggaaagtggtcaaccagaagctaaacaggaagaagcaaacgaagaggtgaaaca agaagaagaaggtaaataagtatttttgtattatataacaaacaaagtaaggaatacagatttatacaataaattgccatactagtcacgtgagatatctcatccattcc ccaactcccaagaaaaaaaaaaagtgaaaaaaaaaatcaaacccaaagatcaacctcccatcatcatcgtcatcaaaccccagctcaattcgca	55

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1695 of 3077

Table 19

Gene (Organism)	Nucleic acid sequence of the gene protein-encoding sequence (from START codon to STOP codon except where otherwise noted)	SEQ ID NO
T_{URA3} (*URA3* terminator from *Candida viswanathii* ATCC 20336)	Taaatattgtaataaataggtctatatacatacactaagcttctaggacgtcattgtagtcttcgaagttgtctgctagtttagttctcatgatttcgaaaaccaataacgca atggatgtagcagggatggtggttagtgcgttcctgacaaacccagagtacgccgcctcaaaccacgtcacattcgcccttgcttcatcgcatcacttgcttgaaagg tatccacgtacgagttgtaatacaccttgaagaacggcttcgtct	56
P_{URA3}-*URA3*-T_{URA3}-P_{URA3} (sequence of each element from *Candida viswanathii* ATCC 20336)	ttaaaggtatctacggttgtttcggtatgaccccgggggatctgacggtacaacgagaattgtattgaattgatcaagaacatgatcttggttacagaacatcaagt tcttggaccagactgagaatgcacagatatacaaggcgtcatgtgataaaatggatgagatttatccacaattgaagaaagagtttatggaaagtggtcaaccaga agctaaacaggaagaagcaaacgaagaggtgaaacaagaagaaggtaaataagtatttttgtattatataacaaacaaagtaaggaatacagatttataca ataaaattgccatactagtcacgtgagatatctcatccattcccaactcccaagaaaaaaaaaaaaagtgaaaaaaaaaaatcaaacccaaagatcaacctcccatc atcatcgtcatcaaaccccccagctcaattcgcaattgcgacgtcaatggttagcacaaaaaacatacacagaaaaggcatcagcacacccctccaaggttgcccaacgtttattccgctt aatggagtccaaaaagaccaacctctgcgcctcgatcgacgtgaccacaaccgccgagttccttgctcatcgacaagctcggtcccacatctgtctcgtgaag acgcacatcgatatcatctcagacttcagctacgagggcacgattgagccgttgcttgtgctgcagagcgccacgggttcttgatattcgaggacaggaagttttgctg atatcggaaacaccgtgatgttgcagtacacctcgggggtataccggatcgcggcgtggagtgacatcacgaacgcgcacggagtgactgggaagggcgtcgtt gaagggttgaaacgcggtgcggagggggtagaaaaggaaagggcgtgttgatgttgcggagttgtcgttggcgcatggtgaatatacccg tgagacgatcgagattgcgaagagtgatcgggagttcgtgattgggttcatcgcgcagcgggacatgggggtagagaagaaggtttgattggatcatcatgacg cctggtgtgggttggatgataaaggcgatgcgttgggccagcagtataggactgttgatgaggtggttctgactggtaccgattgtgattattgtcgggagagggttgtt tggaaaaggaagagaccctgaggtggagggaaagagatacagggatgctggatggaaggcatacttgaagagaagaactggtcagttagaataaatattgtaataa ataggtctatatacatacactaagcttctaggacgtcattgtagtcttcgaagttgtctgctagtttagttctcatgatttcgaaaaccaataacgcaatggatgtagcagg gatggtggttagtgcgttcctgacaaacccagagtacgccgcctcaaaccacgtcacattcgcccttgcttcatcgcatcacttgcttgaaggtatccacgtacgagt tgtaatacaccttgaagaacggcttcgtctacggtcgacgacgggtacaacgagaattgtattgaattgatcaagaacatgatcttggttacagaacatcaagttctt ggaccagactgagaatgcacagatatacaaggcgtcatgtgataaaatggatgagatttatccacaattgaagaaagagtttatggaaagtggtcaaccagaagc taaacaggaagaagcaaacgaagaggtgaaacaagaagaaggtaaataagtatttttgtattatataacaaacaaagtaaggaatacagatttatacaata aattgccatactagtcacgtgagatatctcatccattcccaactcccaagaaaaaaaaaaaaagtgaaaaaaaaaaatcaaacccaaagatcaacctcccatcatc atcgtcatcaaaccccccagctcaattcgcagagctcggtaccc	57
T_{URA3}-P_{URA3}-*URA3*-T_{URA3} (sequence of each element from *Candida viswanathii* ATCC 20336)	taaatattgtaataaataggtctatatacatacactaagcttctaggacgtcattgtagtcttcgaagttgtctgctagtttagttctcatgatttcgaaaaccaataacgcaa tggatgtagcagggatggtggttagtgcgttcctgacaaacccagagtacgccgcctcaaaccacgtcacattcgcccttgcttcatcgcatcacttgcttgaaggt atccacgtacgagttgtaatacaccttgaagaacggcttcgtctcgacgggtacaacgagaattgtattgaattgatcaagaacatgatcttggttacagaacatca agttcttggaccagactgagaatgcacagatatacaaggcgtcatgtgataaaatggatgagatttatccacaattgaagaaagagtttatggaaagtggtcaacca gaagctaaacaggaagaagcaaacgaagaggtgaaacaagaagaaggtaaataagtatttttgtattatataacaaacaaagtaaggaatacagatttatata caataaattgccatactagtcacgtgagatatctcatccattcccaactcccaagaaaaaaaaaaaaagtgaaaaaaaaaaatcaaacccaaagatcaacctcccc atcatcatcgtcatcaaaccccccagctcaattcgcaattgcgacgtcaatggttagcacaaaaacatacacagaaaaggcatcagcacacccctccaaggttgcccaacgtttattcc gcttaatggagtccaaaaagaccaacctctgcgcctcgatcgacgtgaccacaaccgccgagttccttgctcatcgacaagctcggtcccacatctgtctcgtg aagacgcacatcgatatcatctcagacttcagctacgagggcacgattgagccgttgcttgtgctgcagagcgccacgggttcttgatattcgaggacaggaagttt	58

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1696 of 3077

Table 19

Gene (Organism)	Nucleic acid sequence of the gene protein-encoding sequence (from START codon to STOP codon except where otherwise noted)	SEQ ID NO
	gctgatatcggaaacaccgtgatgttgcagtacacctcgggggtataccggatcgcggcgtggagtgacatcacgaacgcgcacggagtgactgggaagggcgt cgttgaagggttgaaacgcggtgcggagggggtagaaaaggaaagggcgtgttgatgttggcggagttgtcgagtaaaggctcgttggcgcatggtgaatatac ccgtgagacgatcgagaattgcgaagagtgatcgggagttcgtgattgggttcatcgcgcagcgggacatgggggtagagaagaaggtttgattggatcatcatg acgcctggtgtgggtttggatgataaaggcgatgcgttgggccagcagtataggactgttgatgaggtggttctgactggtaccgatgtgattattgtcgggagaggggt tgtttggaaaaggagaagaccctgaggtggagggaaagagatacagggatgctggatggaaggcatacttgaagagaaactggtcagttagaataaatattgtaat aaataggtctatatacatacactaagcttcatggacgtcattgtagtcttcgaagttgtctgctagtttagttctcatgatttcgaaaaccaataacgcaatggatgtagca gggatggtggttagtgcgttcctgacaaacccagagtacgccgcctcaaaccacgtcacattccgccctttgcttcatccgcatcacttgcttgaaggtatccacgtacg agttgtaatacaca	
CAT2 (*Candida viswanathii* ATCC 20336)	atgtttaacttaagttgtcgcaacaagtattaaagaattccaccaaatccattatgccaatttttgaaaaaaaccattctccaccagccacgcaaagggtgacttgttcaa ataccagtcacaattacccaagttgcctgttcctacttggaagaaaccgcatccaagtacctcaagaccgttgagccattctgaaccaagagcaattggaatccac caaggccaaagtcgctgagtttgttagaccaggtggtgccggtgaagccttgcaagccagattgaacaacttgccgccgacaaggacaactggttggctgaatttt gggacgactatgcatacatgtcttatagagatccgttgttccatatgtttcttacttttcagtcacaaggatgtcaagaacatcattgggcaagaccaattgttgaaggc cactttgattgcttactacactattgagttccaagaaaaggtttggacgaaagttggacccagaagtcatcaagggtaacccattctgtatgaacgccttcaagtaca tgttcaacaactcgagagttccagctgaaggctccgacatcacccaacactacaacggtgaagaaaaccaattttcgttgtcatctacaagaacaacttctacaag gttccaacccacaagaacggccaaagattgaccaagggtgaaatctacagctacttgcaagaaatcaagaacgatgccactccaaagggtctcggtttgggtgct ttgacctcattgaacagagacgaatggttgagtgcctacaacaactgttgaagtcccccaatcaacgaagcttccttgggatccatcttgcttccagcttttgtcattgcctt ggactccaacaacccagtcaccattgaagaaaaatccaagaactgctggcacggggacggtcaaaacagattctttgacaagcctttggaattcttcgtcagtgct aacggtaacctctggtttccttggtgaacactccagaatggacgctaccccaaccgtgcaattgaacaacaccatctacaagcaaatcttggaaaccaatccaaacg acttgattgttgaaattggttcttctgctccaagattcggcaatgctgaaatcttgcctttcgacatcaacccaacccagaggccaacatcaaagacgctattgccaag tttgacgccaccattgctgcccacgacgaagaaatctccaacactacggttacggtaaggattgatcaagaagttcaaggtctccccagatgcctacgtgcaattg ttgatgcaattggcatacttcaagtacaccggcaagatcagaccaacttatgaatccgccgccaccagaaagttcttgaaggtagaaccgaaaccggtagaact gtttccaacgaatccaagaagttgttgagacctggtccgatccaaacgctagcagccgccgacaaggttgccacttccaagctgccgctaagcaacacgttgctta cttgtctgctgccgccgatggtaagggtgtcgaccgtcacttgtttggtttgaagcagatgattcaaccaggcgaaccaatcctcgaaatcttcactgaccccaatctttag ctattctcaaacctggtacatttcttcttcccaagtcccatctgaattcttccaatcttgggttggtcgcaagtcattgacgacggtttcggtttggcttactgatgatgaacaac gactggatccacgttcacatttctgtaagagaggtaacggcttgcaatccgaccacttgaaatggtactggttgatagtgctaacgaaatgaaggatgtcttgactaa gggattattgactgatgctaagcctaagttgtaa	59

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1697 of 3077

Table 19

Gene (Organism)	Nucleic acid sequence of the gene protein-encoding sequence (from START codon to STOP codon except where otherwise noted)	SEQ ID NO
Engineered DNA fragment *CAT2*$^{\Delta mis}$	atgccaatttgaaaaaaccattctccaccagccacgcaaagggtgacttgttcaaataccagtcacaattacccaagttgcctgttcctacttggaagaaaccgcat ccaagtacctcaagaccgttgagccattcttgaaccaagagcaattggaatccacccaaggccaaagtcgctgagtttgttagaccaggtggtgccggtgaagccctt gcaagccagattgaacaacatttgccgccgacaaggacaactggttggctgaatttgggacgactatgcatacatgtctatagagatcctgttgttccatatgttcttac ttttttcagtcacaaggatgtcaagaacatcattggccaagaccaattgttgaaggccacttgattgcttactacactattgagttccaagaaaaaggttttggacgaaagt ttggacccagaagtcatcaagggtaacccattctgtatgaacgccttcaagtacatgttcaacaactcgagagttccagctgaaggctccgacatcacccaacacta caacggtgaagaaaaccaatttttcgttgtcatctacaagaacaacttctacaaggttccaacccacaagaacggccaaagattgaccaaggtgaaatctacag ctacttgcaagaaatcaagaacgatgccactccaaagggtctcggtttgggtgctttgacctcattgaacagagacgaatggttgagtgcctacaacaacttgttgaa gtcccaatcaacgaagcttccttgggatccatctttgcttcaagcttgcttgcattgcctggactcccaacaacccagtcaccattgaagaaaaatccaagaactgctggc acggggacggtcaaaacagattctttgacaagccttggaattcttcgtcagtgctaacggtaactctggtttccttggtgaacactccagaatggacgctacccccaac cgtgcaattgaacaacaccatctacaagcaaatctggaaaccaatccaaacgacttgattgttgaaattggttctctgctccaagattcggcaatgctgaaatcttgc ctttcgacatcaacccaacccaccagagccaacatcaaagacgctattgccaagtttgacgcccaccattgctgcccacgacgaagaaatcttccaacactacggtta cggtaagggattgatcaagaagttcaaggtctccccagatgcctacgtgcaattgttgatgcaattggcatacttcaagtacaccggcaagatcagaccaacttatga atccgccgccaccagaaaagttcttgaagggtagaaccgaaaaccggtagaactgtttccaacgaatccaagaagttgttgagacctggtccgatccaaacgctag cagcgccgacaaggttgccacttccaagctgccgctaagcaacacgttgcttacttgtctgctgccgcgatggtgtcgaccgtcacttgtttggtttgaagc agatgattcaaccaggccgaaccaatccctgaaatcttcactgacccaatctttagctattctcaaaacctggtacatttcttcttcccaagtcccatctgaattcttccaatctt gggttggtcgcaagtcattgacgacggtttcggtttggcttacttgatcaacaacgactggatccacgttcacatttcttgtaagaggtaacggcttgcaatccgac cacttgaaatggtacttggttgatagtgctaacgaaatgaaggatgtcttgactaagggattattgactgatgctaagcctaagttgtaa	60

Table 19

Gene (Organism)	Nucleic acid sequence of the gene protein-encoding sequence (from START codon to STOP codon except where otherwise noted)	SEQ ID NO
Engineered DNA fragment *CAT2*$^{\Delta mls\ \Delta pts}$	atgccaatttgaaaaaaccattctccaccagccacgcaaagggtgacttgttcaaataccagtcacaattacccaagttgcctgttcctacttggaagaaaccgcat ccaagtacctcaagaccgttgagccattcttgaaccaagagcaattggaatccacccaaggccaaagtcgctgagtttgttagaccaggtggtgccggtgaagcctt gcaagccagattgaacaacttgccgccgacaaggacaactggttggctgaatttgggacgactatgcatacatgtctatagagatcctgttgttccatatgttcttac ttttttcagtcacaaggatgtcaagaacatcattggccaagaccaattgttgaaggccacttgattgcttactacactattgagttccaagaaaaaggttttggacgaaagt ttggacccagaagtcatcaagggtaacccattctgtatgaacgccttcaagtacatgttcaacaactcgagagttccagctgaaggctccgacatcacccaacacta caacggtgaagaaaaccaatttttcgttgtcatctacaagaacaacttctacaaaggttccaacccacaagaacggccaaagattgaccaaggtgaaatctacag ctacttgcaagaaatcaagaacgatgccactccaaagggtctcggtttgggtgctttgacctcattgaacagagacgaatggttgagtgcctacacaacaacttgttgaa gtcccccaatcaacgaagcttccttgggatccatctttgcttccagctttgtcattgccttggactccaacaacccagtcaccattgaagaaaaatccaagaactgctggc acgggggacggtcaaaacagattctttgacaagccttttggaattcttctgtcagtgctaacggtaactctggtttccttggtgaacactccagaatggacgctacccccaac cgtgcaattgaacaacaccatctacaagcaaatctttggaaaccaatccaaacgacttgattgttgaaattggttctctgctccaagattcggcaatgctgaaatcttgc ctttcgacatcaacccaaccacccagagccaacatcaaagacgctattgccaagtttgacgcccaccattgctgcccacgacgaagaaatcttccaacactacggtta cggtaagggattgatcaagaagttcaaggtctccccagatgcctacgtgcaattgttgatgcaattggcatacttcaagtacaccggcaagatcagaccaacttatga atccgccgccaccagaaaagttcttgaagggtagaaccgaaaaccggtagaactgtttccaacgaatccaagaagttgttgagacctggtccgatccaaacgctag cagcgccgacaaggttgccacttccaagctgccgctaagcaacacgttgcttacttgtctgctgccgcgatggtaagggtgtcgaccgtcacttgtttggtttgaagc agatgattcaaccaggcgaaccaatccctgaaatcttcactgacccaatctttagctattctcaaaacctggtacatttcttcttcccaagtcccatctgaattcttccaatctt gggttggtcgcaagtcattgacgacggtttcggtttggcttacttgatcaacaacgactggatccacgttcacatttcttgtaagaggtaacggcttgcaatccgac cacttgaaatggtacttggttgatagtgctaacgaaatgaaggatgtcttgactaagggattattgactgatgctaagtaa	61

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1699 of 3077

Table 19

Gene (Organism)	Nucleic acid sequence of the gene protein-encoding sequence (from START codon to STOP codon except where otherwise noted)	SEQ ID NO
Engineered DNA fragment *CAT2*$^{\Delta pts}$	atgtttaacttaagttgtcgcaacaagtattaaagaattccaccaaatccattatgccaatttttgaaaaaaaccattctccaccagccaccgcaaaggtgacttgttcaa ataccagtcacaattacccaagttgcctgttcctacttttggaagaaaccgcatccaagtacctcaagacgttgagccattcttgaaccaagagcaattggaatccac caaggccaaagtcgctgagtttgttagaccaggtggtgccggtgaagccttgcaagccagattgaacaactttgccgccgacaaggacaactggttggctgaatttt gggacgactatgcatacatgtcttatagagatcctgttgttccatatgtttcttacttttttcagtcacaaggatgtcaagaacatcattggccaagaccaattgttgaaggc cactttgattgcttactacactattgagttccaagaaaaaggttttggacgaaagttgtggacccagaagtcatcaagggtaacccattctgtatgaacgccttcaagtaca tgttcaacaactcgagagttccagctgaaggctccgacatcacccaacactacaacaacggtgaagaaaaccaatttttcgttgtcatctacaagaacaacttctacaag gttccaacccacaagaacggccaaagattgaccaagggtgaaatctacagctacttgcaagaaatcaagaacgatgccactccaaagggtctcggtttgggtgct ttgacctcattgaacagagacgaatggttgagtgcctacaacaacttgttgaagtcccccaatcaacgaagcttcctgggatccatctttgcttccagctttgtcattgcctt ggactccaacaacccagtcaccattgaagaaaaatccaagaaactgctggcacgggggacggtcaaaacagattctttgacaagccttggaattcttcgtcagtgct aacggtaactctggtttccttggtgaacactccagaatggacgctacccccaacgttgcaattgaacacaccatctacaagcaaatcttggaaaccaatccaaacg acttgattgttgaaattggttctctgctccaagattcggcaatgctgaaatctttgccttcgacatcaacccaacaccagagccaacatcaaagacgctattgccaag tttgacgccaccattgctgcccacgacgaagaaatctccaacactacggttacggtaagggattgatcaagaagttcaaggtctccccagatgcctacgtgcaattg ttgatgcaattggcatacttcaagtacacacggcaagatcagaccaacttatgaatcgccgccgccaccagaaagttcttgaagggtagaaccgaaaccggtagaact gtttccaacgaatccaagaagttgttgagacctggtccgatccaaacgctagcagccgccgacaaggttgccacttccaagctgccgctaagcaacacgttgctta cttgtctgctgccgccgatggtaagggtgtcgaccgtcacttgtttggtttgaagcagatgattcaaccaggcgaaccaatccctgaaatcttcactgacccaatctttag ctattctcaaacctggtacatttcttcttcccaagtcccatctgaattcttccaatcttgggttggtcgcaagtcattgacgacggtttcggtttggcttactttgatcaacaac gactggatccacgttcacatttcttgtaagaggtaacggcttgcaatccgaccacttgaaatggtgctactggttgatagtgctaacgaaatgaaggatgtcttgactaa gggattattgactgatgctaagtaa	62

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1700 of 3077

Table 19

Gene (Organism)	Nucleic acid sequence of the gene protein-encoding sequence (from START codon to STOP codon except where otherwise noted)	SEQ ID NO
YAT1 (*Candida viswanathii* ATCC20336)	ATGTCAACTTACCAGTTCCAAGAAACTTTAGAAAAGCTTCCAATCCCAGACTTAAACCAAACATGCGCTAACTAC CTCAACGTCTTGCGTCCGCTCCAAACAGAGCAAGAACACATCAAGACCAAGACTGCCGTTGAGAATTTCCTCAA GAATGGCACCGGCCAGTACCTCGACGCTGCCTTGCGCGAATACGCCCAGACCCGCCCCAGCTACATTGAACA GTTCTGGTACGACGCATACTTAAACTACGACTCCCCCGTTGTTTTAAACTTGAATCCGTTTTTCTTACTCGAAGA CGACCCATTCACCAACGAGTCCAGCTCCGTCAATCCCCAGGTTAAGCGTGCCACCAGCTTGGTGATGTCCTCG TTGAAGTTCATCCAGGCGTTGAAGAACGAGACGTTGAGCGTCGACACCTTGAAAGGAGGCAAGCCGTTGTGCA TGTACCAGTACACCAAATTGTTTGGGGCGCTGCGTATCCCCTCCGAGGACGGGTGTGTGATGCAGAGCGACCC GGCAAGTAACCACATTGTCGTGATGTCTAAGTCCCAGTTGTATTGGTTTGACGTGTTGGATAGCAAGAACAACT TGATTTTGAGCGAAGCGGAGTTGAATGTGAACTTCCAGTCGATCATTCATGACTCTTTGAGAACTCCCTCGGAT GAGATTGCAAAGAGCTCGTTTGGGGTGTTGACGACGGAGAATAGAAGAATTTGGGCTAATGTTAGACATAATTT GATGTCGACTACCAATAATAAGGTTAACCATGAGGTGTTGTCGATTATTGACTCAGCGTTGTTTGTGTTATGCTT TGATGATATCGTGATTAATGACTTGTCGGAGTTGTCCAAGAATATGCTTTGTGGGTTGTCCATTTTGGATAATGG TATCCAAGTGGGTACTTGTACAAACAGATGGTATGACAAGTTGCAGATTATCGTTACCAAGAATGCCAAAGCCG GTATTAATTTCGAACATACTGGTGTTGACGGTCATACGGTGTTGAGATTTGTCAGTGATATCTACACCGACTCCA TCTTGTCCTTTGCCAACTCCATCAATAGCAACGCTCCATCTTTATGGAACACCACTACAAACTACAAAGAACTTG ATGGGGAAGATTTAATCACCGTGCCACGTAAGTTGGAATGGGAGTTGACCCCTGACTTATCCTTGGCTTTGAGA TTTGGAGAAACCAGATTGTCGGACTTGATCAACCAGAACGAGTTCCGCCACTTGGAGTTCAAGAACTATGGGTC TACACAAATCAAGAAGATGAAGTTCTCACCAGATGCGTTTGTTCAAATGGCGTTCCAAGCAACCTACTATGCGC TTTACGGTAAGGTTGAATGTACTTATGAGCCAGCCATGACGAAGCAGTTCTACCACGGAAGAACAGAAGCTATA AGAACCGTCTCGCAAGAATCGAACTTGTTTGTGCGTAAGTTCTTTGACTCCACTGTGTCCATCCAGAAGAAGTT AGAGTACTTGACCCAAGCTTGCACCAAGCACTCCCAGCAAACCAGAATGAGTTCTGCCGGTCAAGGTGTCGAT CGTCACTTGTATGCTTTATTCTGTATCTGGAAGAGATACTTACACGATGCCGAAGACGACGATGACGATGCCTC CAAAGACCAACCATCATCGACGGAATTATCCCGTGATGACACCATCATTGCTCAAACCAACGGTGACAGCACCG ATGATGACGACTCCACCACCGTTGTCGGACACGGCGCCAACCACCATCACAACTTGAAAAGCGCCGATTTGTT GAAAACCATCCCTGAGATATTCGCAGACAATGGCTGGGACAAGTTGAACAACACGATCATATCCACCTCCAACT GTGGTAATCCTTCTTTGAGATTGTTTGGATTTGGTCCTGTTTCAGCTAACGGGTTTGGTATCGGGTATATCTTGA AGGACGACTCGATCTCGATCTGTGCTTCGTCGAAGCATCGTCAGACGCAGAGATTCTTGGTCACTTTGAACTCG TACTTGTTGGAGATCCAGAACATCTGGAAACAAGCGCAAAAGATGGAGAAGTTGAAGGCTGAGTTGGCTAAAG CCGTCGAGGATGCAAAACCTAAGGGACAGAAATCAGGTGAAGCAGAGGGGATCATCGAGCCTAAGCCTAATAA CTTGAGCACTCTTTTGGGAGGTTATGGTTATTTCGATATGGGCGATGAGGACATCAAGAGCAGAGGACAATCTC CAGAACCTCCGTTCTTGCATCGTGCTGGATCCGGTTTCAGCATTCGTGAAATAGGCAAGAAGTTGAGATTAAGC GAGTATTAG	63

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1701 of 3077

Table 19

Gene (Organism)	Nucleic acid sequence of the gene protein-encoding sequence (from START codon to STOP codon except where otherwise noted)	SEQ ID NO
COX4 first 90 nucleic acids encoding mitochondrial targeting sequence (*Candida viswanathii* ATCC20336)	Atgctttccgtaccacttaagagttgccagacacaacaaaccagattattgtctacttccagaatcttgttcaacagcaagaccgaccaa	64
CIT1 first 75 nucleic acids encoding mitochondrial targeting sequence (*Candida viswanathii* ATCC20336)	Atgtctgcattaagatcattccaacgctcttctaacgttgctaagtccacctaaagaactcagtcagaacctat	65
CAT2 first 99 nucleic acids encoding mitochondrial targeting sequence (*Candida viswanathii* ATCC20336)	Atgtttaactttaagttgtcgcaacaagtattaaagaattccaccaaatccattatgccaattttgaaaaaaccattctccaccagccacgcaaagggt	66
Engineered DNA fragment *YAT1*$^{+COX4mts}$	atgctttccgtaccacttaagagttgccagacacaacaaaccagattattgtctacttccagaatcttgttcaacagcaagaccgaccaatcaacttaccagttccaag aaacttagaaaagcttccaatccagacttaaaccaaacatgcgctaactacctcaacgtcttgcgtccgctcaaacagagcaagaacacatcaagaccaag actgccgttgagaatttcctcaagaatggcaccggccagtacctcgacgctgccttgcgcgaatacgcccagaccgccccagctacattgaacagttctggtacg acgcatacttaaactacgactccccgttgtttaaacttgaatccgttttcttactcgaagacgacccattcaccaacgagtccagctccgtcaatcccaggttaagc gtgccaccagcttggtgatgtcctcgttgaagttcatccaggcgttgaagaacgagacgttgagcgtcgacaccttgaaaggaggcaagccgttgtgcatgtaccag tacaccaaattgtttgggcgctgcgtatcccctccgaggacgggtgtgatgcagagcgaccccggcaagtaaccacattgtcgtgatgtctaagtcccagttgtatt	67

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1702 of 3077

Table 19

Gene (Organism)	Nucleic acid sequence of the gene protein-encoding sequence (from START codon to STOP codon except where otherwise noted)	SEQ ID NO
	ggtttgacgtgttggatagcaagaacaacttgatttgagcgaagcggagttgaatgtgaacttccagtcgatcattcatgactctttgagaactccctcggatgagattg	
	caaagagctcgtttgggtgttgacgacggagaatagaagaatttgggctaatgttagacataatttgatgtcgactaccaataataaggttaaccatgaggtgttgtc	
	gattattgactcagcgttgtttgtgttatgctttgatgatatcgtgattaatgacttgtcggagttgtccaagaatatgctttgtgggttgtccatttggataatggtatccaagtg	
	ggtacttgtacaaacagatggtatgacaagttgcagattatcgttaccaagaatgccaaagcggtattaatttcgaacatactggtgttgacggtcatacggtgttga	
	gatttgtcagtgatatctacaccgactccatcttgtcctttgccaactccatcaatagcaacgctccatctttatggaacaccactacaaactacaaagaactgatggg	
	gaagatttaatcaccgtgccacgtaagttggaatgggagttgaccctgacttatcctggctttgagatttggagaaaccagattgtcggacttgatcaaccagaacg	
	agttccgccacttggagttcaagaactatgggtctacacaaatcaagaagatgaagttctcaccagatgcgtttgttcaaatggcgttccaagcaacctactatgcgct	
	ttacggtaaggttgaatgtacttatgagccagccatgacgaagcagttctaccacggaagaacagaagctataagaaccgtctgcgcaagaatcgaactgtttgtgc	
	gtaagttctttgactccactgtgtccatccagaagaagttagagtactttgacccaagcttgcaccaagcactcccagcaaaccagaatgagttctgccggtcaaggtg	
	tcgatcgtcacttgtatgcttattctgtatctggaagagatacttacacgatgccgaagacgacgatgacgatgcctccaaagaccaacaaccatcatcgacggaattatc	
	ccgtgatgacaccatcattgctcaaaccaacggtgacagcacccgatgatgacgactccaccaccgttgtcggacacggcgccaaccaccatcacaacttgaaaa	
	gcgccgatttgttgaaaaccatccctgagatattcgcagacaatggctgggacaagttgaacaacacgatcatatccacctccaactgtggtaatccttctttgagatt	
	gtttggatttggtcctgtttcagctaacgggtttggtatatcttgaaggacgactcgatctcgatctgtgcttcgtcgaagcatcgtcagacgcagagattcttgg	
	tcacttgaactcgtactgttggagatccagaacatctggaaacaagcgcaaaagatggagaagttgaaggctgagttggctaaagccgtcgaggatgcaaaac	
	ctaagggacagaaatcaggtgaagcagagggggatcatcgagcctaagcctaataacttgagcactcttttgggaggtttatggttattcgatatgggcgatgaggac	
	atcaagagcagaggacaatctccagaacctccgttcttgcatcgtgctggatccggtttcagcattcgtgaaataggcaagaagttgagattaagcgagtattag	

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1703 of 3077

Table 19

Gene (Organism)	Nucleic acid sequence of the gene protein-encoding sequence (from START codon to STOP codon except where otherwise noted)	SEQ ID NO
Engineered DNA fragment *YAT1*$^{+CIT1mts}$	atgtctgcattaagatcattccaacgctcttctaacgttgctaagtccacttaaagaactcagtcagaacctattcaacttaccagttccaagaaaacttagaaaagctt ccaatcccagacttaaaccaaacatgcgctaactacctcaacgtcttgcgtccgctccaaacagagcaagaacacatcaagaccaagactgccgttgagaatttc ctcaagaatggcaccggccagtacctcgacgctgccttgcgcgaatacgcccagagaccgcccccagctacattgaacagttctggtacgacgcatacttaaactac gactccccgttgttttaacttgaatccgttttttcttactcgaagacgacccattcaccaacgagtccagctccgtcaatcccccaggttaagcgtgccaccagcttggtg atgtcctcgttgaagttcatccaggcgttgaagaacgagacgttgagcgtcgacaccttgaaaggaggccaagccgttgtgcatgtaccagtacaccaaattgtttgg ggcgctgcgtatcccctccgaggacgggtgtgatgcagagcgacccggcaagtaaccacattgtcgtgatgtctaagtcccagttgtattggtttgacgtgttggat agcaagaacaacttgatttgagcgaagcgggagttgaatgtgaacttccagtcgatcattcatgactctttgagaactccctcggatgagattgccaaagagctcgtttg gggtgttgacgacggagaatagaagaaatttgggctaatgttagacataatttgatgtcgactaccaataataaggttaaccatgaggttgttgtcgattattgactcagcg ttgtttgtgttatgcttgatgatatcgtgattaatgacttgtcggagttgtccaagaatatgctttgtgggttgtccatttttggataatggtatccaagtgggtacttgtacaaac agatggtatgacaagttgcagattatcgttaccaagaatgccaaagccggtattaatttcgaacatactggtgttgacggtcatacggttgagatttgtcagtgatatc tacaccgactccatcttgtccttgccaactccatcaatagcaacgctccatctttatggaacaccactacaaactacaaagaacttgatgtgggaagatttaatcaccg tgccacgtaagttggaatgggagttgaccctgacttatcctggctttgagatttggagaaaccagattgtcggacttgatcaaccagaacgagttccgccacttgga gttcaagaactatgggtctacacaaatcaagaagatgaagttctcaccagatgcgtttgttcaaatggcgttccaagcaacctactatgcgctttacggtaaggttgaa tgtacttatgagccacgccatgacgaagcagttctaccacggaagaacagaagctataagaaccgtctgcaagaatcgaactgtttgtgcgtaagttctttgactcc actgtgtccatccagaagaagaagttagagtactgacccaagcttgcaccaagcactcccagcaaaccagaatgagttctgccggtcaaggtgtcgatcgtcactgtat gctttattctgtatctggaagagatacttacacgatgccgaagacgacgatgacgatgcctccaaagaccaacccatcatcgacggaattatcccgtgatgacaccat cattgctcaaaccaacggtgacagcaccgatgatgacgactccaccaccgttgtcggacacggcgccaaccaccatcacaacttgaaaagcgccgatttgttgaa aaccatccctgagatattcgcagacaatggctgggacaagttgaacaacacgatcatatcccacctccaactgtggtaatccttctttgagattgtttggatttggtcctgtt tcagctaacgggttggtatcgggtatatcttgaaggacgactcgatctcgatctgtgcttcgtcgaagcatcgtcagacgcagagattcttggtcacttttgaactcgtact tgttggagatccagaacatctggaaacaagcgcaaaagatggagaagttgaaggctgagttggctaaagccgtcgaggatgcaaaacctaagggacagaaaat caggtgaagcagagaggggatcatcgagcctaagcctaataacttgagcactctttttgggaggttatggttatttcgatatgggcgatgaggacatcaagagcagagg acaatctccagaacctccgttcttgcatcgtgctggatccggtttcagcattcgtgaaataggcaagaagttgagattaagccgagtattag	68

Table 19

Gene (Organism)	Nucleic acid sequence of the gene protein-encoding sequence (from START codon to STOP codon except where otherwise noted)	SEQ ID NO
Engineered DNA fragment *YAT1*$^{+CAT2mts}$	atgtttaacttaagttgtcgcaacaagtattaaagaattccaccaaatccattatgccaatttttgaaaaaaaccattctccaccagccaccgcaaaggttcaacttacca gttccaagaaaacttagaaaaagcttccaatcccagacttaaaccaaacatcgctaactacctcaacgtcttgcgtccgctccaaacagagcaagaacacatcaa gaccaagactgccgttgagaatttcctcaagaatggcaccggccagtacctcgacgctgccttgcgcgaatacgcccagaccgccgccccagctacattgaacagtt ctggtacgacgcatacttaaactacgactccccgttgttttaaacttgaatccgttttctactgaagacgacccattcaccaacgagtccagctccgtcaatcccca ggttaagcgtgccaccagcttggtgatgtcctcgttgaagttcatccaggcgttgaagaacgagacgttgaggaggcaagccgttgtgca tgtaccagtacaccaaattgtttgggcgctgcgtatccctccgaggacgggtgtgtgatgcagagcgacccggcaagtaaccacattgtcgtgatgtctaagtccc agttgtattggtttgacgtgttggatagcaagaacaactgatttgagcgaagcggagttgaatgtgaacttccagtcgatcattcatgactctttgagaactcctcggga tgagattgcaaagagctcgtttgggtgttgacgacggagaatagaagaaatttgggctaatgttagacataatttgatgtcgactaccaatataaggttaaccatga ggtgttgtcgattattgactcagcgttgtttgtgtttatgcttgaatgatatcgtgaattgactgttgtgggagttgtccaaagaatatgctttgtgggttgtccatttttggataatggt atccaagtgggtactgtacaaacagatggtatgacaagttgcagattatcgttaccaagaatgccaaagccggtattaatttcgaacatactggtgttgacggtcata cggtgttgagatttgtcagtgatatctacaccgactccatctgtccttgccaactccatcaatagcaacgctccatcttatggaacaccactacaaactacaaagaa cttgatgggggaagatttaatcaccgtgccacgtaagttggaatgggagttgaccctgacttatccttggctttgagatttggagaaaccagattgtcgggacttgatcaa ccagaacgagttccgccacttggagttcaagaaactatgggtctacacaaatcaagaagatgaagttctcaccagatgcgttgttgtcaaatggcgttccaagcaacct actatgcgctttacggtaaggttgaatgtactatgagccagccatgacgaagcagttctaccacggaagaacagaagctatagaaccgtctcgcaagaatcgga acttgttgtgcgtaagttcttgactccactgtgtccatccagaagaagttagagtacttgaccccaagcttgcaccaagcactcccagcaaaccagaatgagttctgcc ggtcaaggtgtcgatcgtcacttgtatgctttattctgtatctggaagagatacttacacgatgccgaagacgacgatgacgatgccttccaaagaccaaccaatcatcg acggaattatccgtgatgacaccatcattgctcaaaccaacggtgacagcaccgatgatgacgactccaccacgttgtcggacacgggcgccaaccaccaccatca caacttgaaaagcgccgatttgttgaaaaccatcctgagatattcgcagacaatggctgggacaagttgaacaacacgatcatatccacctccaactgtggtaatc cttctttgagattgtttggatttggtcctgtttcagctaacgggtttgtatcgggtatatcttgaaggacgactcgatctcgatctgtgcttcgtcgaagcatcgtcagacgc agagattcttggtcacttttgaactcgtactttggagatccagaacatctggaaacaagcgcaaaagatggagaagttgaaggctgagttggctaaagccgtcga ggatgcaaaacctaagggacagaaatcaggtgaagcagaggggatcatcgagcctaagcctaataacttgagcactcttttgggaggttatggttatttcgatatgg gcgatgaggacatcaagagcagaggacaatctccagaacctccgttcttgcatcgtgctggatccggtttcagcattcgtgaaataggcaagaagttgagattaag cgagtattag	69

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1705 of 3077

Table 19

Gene (Organism)	Nucleic acid sequence of the gene protein-encoding sequence (from START codon to STOP codon except where otherwise noted)	SEQ ID NO
Engineered DNA fragment *YAT1*$^{+pts}$	ATGTCAACTTACCAGTTCCAAGAAACTTTAGAAAAGCTTCCAATCCCAGACTTAAACCAAACATGCGCTAACTAC CTCAACGTCTTGCGTCCGCTCCAAACAGAGCAAGAACACATCAAGACCAAGACTGCCGTTGAGAATTTCCTCAA GAATGGCACCGGCCAGTACCTCGACGCTGCCTTGCGCGAATACGCCCAGACCCGCCCCAGCTACATTGAACA GTTCTGGTACGACGCATACTTAAACTACGACTCCCCCGTTGTTTTAAACTTGAATCCGTTTTTCTTACTCGAAGA CGACCCATTCACCAACGAGTCCAGCTCCGTCAATCCCCAGGTTAAGCGTGCCACCAGCTTGGTGATGTCCTCG TTGAAGTTCATCCAGGCGTTGAAGAACGAGACGTTGAGCGTCGACACCTTGAAAGGAGGCAAGCCGTTGTGCA TGTACCAGTACACCAAATTGTTTGGGGCGCTGCGTATCCCCTCCGAGGACGGGTGTGTGATGCAGAGCGACCC GGCAAGTAACCACATTGTCGTGATGTCTAAGTCCCAGTTGTATTGGTTTGACGTGTTGGATAGCAAGAACAACT TGATTTTGAGCGAAGCGGAGTTGAATGTGAACTTCCAGTCGATCATTCATGACTCTTTGAGAACTCCCTCGGAT GAGATTGCAAAGAGCTCGTTTGGGGTGTTGACGACGGAGAATAGAAGAATTTGGGCTAATGTTAGACATAATTT GATGTCGACTACCAATAATAAGGTTAACCATGAGGTGTTGTCGATTATTGACTCAGCGTTGTTTGTGTTATGCTT TGATGATATCGTGATTAATGACTTGTCGGAGTTGTCCAAGAATATGCTTTGTGGGTTGTCCATTTTGGATAATGG TATCCAAGTGGGTACTTGTACAAACAGATGGTATGACAAGTTGCAGATTATCGTTACCAAGAATGCCAAAGCCG GTATTAATTTCGAACATACTGGTGTTGACGGTCATACGGTGTTGAGATTTGTCAGTGATATCTACACCGACTCCA TCTTGTCCTTTGCCAACTCCATCAATAGCAACGCTCCATCTTTATGGAACACCACTACAAACTACAAAGAACTTG ATGGGGAAGATTTAATCACCGTGCCACGTAAGTTGGAATGGGAGTTGACCCCTGACTTATCCTTGGCTTTGAGA TTTGGAGAAACCAGATTGTCGGACTTGATCAACCAGAACGAGTTCCGCCACTTGGAGTTCAAGAACTATGGGTC TACACAAATCAAGAAGATGAAGTTCTCACCAGATGCGTTTGTTCAAATGGCGTTCCAAGCAACCTACTATGCGC TTTACGGTAAGGTTGAATGTACTTATGAGCCAGCCATGACGAAGCAGTTCTACCACGGAAGAACAGAAGCTATA AGAACCGTCTCGCAAGAATCGAACTTGTTTGTGCGTAAGTTCTTTGACTCCACTGTGTCCATCCAGAAGAAGTT AGAGTACTTGACCCAAGCTTGCACCAAGCACTCCCAGCAAACCAGAATGAGTTCTGCCGGTCAAGGTGTCGAT CGTCACTTGTATGCTTTATTCTGTATCTGGAAGAGATACTTACACGATGCCGAAGACGACGATGACGATGCCTC CAAAGACCAACCATCATCGACGGAATTATCCCGTGATGACACCATCATTGCTCAAACCAACGGTGACAGCACCG ATGATGACGACTCCACCACCGTTGTCGGACACGGCGCCAACCACCATCACAACTTGAAAAGCGCCGATTTGTT GAAAACCATCCCTGAGATATTCGCAGACAATGGCTGGGACAAGTTGAACAACACGATCATATCCACCTCCAACT GTGGTAATCCTTCTTTGAGATTGTTTGGATTTGGTCCTGTTTCAGCTAACGGGTTTGGTATCGGGTATATCTTGA AGGACGACTCGATCTCGATCTGTGCTTCGTCGAAGCATCGTCAGACGCAGAGATTCTTGGTCACTTTGAACTCG TACTTGTTGGAGATCCAGAACATCTGGAAACAAGCGCAAAAGATGGAGAAGTTGAAGGCTGAGTTGGCTAAAG CCGTCGAGGATGCAAAACCTAAGGGACAGAAATCAGGTGAAGCAGAGGGGATCATCGAGCCTAAGCCTAATAA CTTGAGCACTCTTTTGGGAGGTTATGGTTATTTCGATATGGGCGATGAGGACATCAAGAGCAGAGGACAATCTC CAGAACCTCCGTTCTTGCATCGTGCTGGATCCGGTTTCAGCATTCGTGAAATAGGCAAGAAGTTGAGATTAAGC GAGTATCCTAAGTTGTAG	70

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1706 of 3077

Table 19

Gene (Organism)	Nucleic acid sequence of the gene protein-encoding sequence (from START codon to STOP codon except where otherwise noted)	SEQ ID NO
CRC1 (*Candida viswanathii* ATCC20336)	atggacgacgttgattctgctttagccgataatgttaaatccttcgctgccggtggttttggtggtatttgtgccgttgaccggtcatccattcgacttggttcaaggtcaga ttgcaaactgggttgtacaaatcgtcagtgcaatgtgtcaaagaaacaatagccaaagacgggttgtttgggttatacagaggtgtcttgcctccattgttgggtgtcac cccaatgtttgctgtttcctctgggggttacgacgttggtaagaaattggtttcttccttcactggcaaatcagtcgacaagtttgaaatcaaagacatttcgactgccggttt catcagtgccatcccaacccacccttggttgccgctccattcgaaagagtcaaggtcatgatgcagatccaagaaggtgccaagtccaagtcccatgggtgccgttgttg ctgaaatgtacagaacccggtggtatcagatccatcttcaagggtactgtggctacttttggccagagatggtccagttctgccttgtattccgctacttatgaatgggtca agaaggaattgactgccctggtgaagacttgtcctgtttgctattaccactgctggtggtttccgctgtattgccatggtcgtttcccaattgacaccatcaagt ctacccagcagtcttcaacgtcaaggtccattgtgcaggccactaagaacatctacgctaaaggtggtatcaaggctttctccctggttgttggtcctgccttggcc agagctttccagccaacgctgctacctttctaggggttgaattggccagaaagttcttggacaaagttatataa	71
ACH1 (*Candida viswanathii* ATCC20336)	atgtctgccatttttgaaacaaagagttagatatgctccatacctcaaaaaattgagaacagcagaacaatgtgtgacttgttcaagaacggccaatacttgggttggt cgggtttcacggtgtcggtgccccaaaggctgtcccagccgctttggttgaccacgttgaaaagaacaaattgcagggccaaatgggcttccacttgtttgtcggtgc ttctgccggtccagaagaaagcagatgggctgaaaacaacatgctttatctagatctccacatcaagtcggtaagccaattgctgctgccatcaatgacggcagaa ctcagttcttcgacaagcacttgtccatgttcccacaagatttgacctacggcttctacaccaagaacaagccaaacggctccaacttggattacaccatcattgaagc taccgccataacagaggatggtgccattatcccaggtcctgccgtcggtgcctcccccagaaaatgctttcggtttccgacaacatcatcattgaagtcaacacaaaaa caccatccttcgaaggtatacacgacattgacttgccagtcaacccaccattcagacaacctacccacacaccactgccgactacagaattggcagaaccgcca tcccagtcgacccagaaaaggttatcgccattgtcgaaaccacagaggttgacaaggttccaccaaacacccccaagtgacgctcaatcccaagccattgctggtc acttgattgaattcttggaaaacgaagtcaaacaagggtcgtctcccagaaaacttgcacccattacaatcggtatcggtaacattgccaacgctgtcgttgaaggttt ggcccaatccaacttcaagaacttgactgtctggaccgaagtcttgcaagattccttcttggacttctcgaatccgggaagtttggacttgcaaccgccacctccatca gattgaccgaggaaggtttcaagaagttctacgaaaaactgggacgtctactccaagaagttgtgcttgagatcgcaagttgtctccaactgccagaaatcatcaga agattgggtgtcttggcttcaacacccccagttgaagttgacatctacggcccacgccaactccaccaccaagtcatgggttcgagaatgttgaacggtttgggaggctcgg gtgatttcttgagaaacgccaagttgtccgtcatgcacacccccatctgctgtccaaccaaggttgacccaacccggtgtctctgtattgtccaatggccacccacgtt gatcaaaccgaacacgatctcgatgttgtcgttaccgaacaaggtttggctgatctcagaggtctcagtccaaaggaaagagctcaagttataatcaacaactgtgc ccacccagactaccaggctcaattgcaagactacttgacagagctgtttctacgccaccaagaagaagatgttgcacgaaccacacatgttgaacgaagctttcg ctatgcacttgaacttgcaagaaaatggtaccatgaagttgaacaaatag	72

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1707 of 3077

Table 19

Gene (Organism)	Nucleic acid sequence of the gene protein-encoding sequence (from START codon to STOP codon except where otherwise noted)	SEQ ID NO
Engineered DNA *ACH1*$^{\Delta mts+pts}$	atggctccatacctcaaaaaattgagaacagcagaacaatgtgtcgacttgttcaagaacggccaatacttgggttggtcgggtttcaccggtgtcggtgccccaaa ggctgtcccagcgccgcttggttgaccacgttgaaaagaacaaattgcagggccaaatgggcttccacttgttgtcggtgctctgccggtccagaagaaagcagatg ggctgaaaacaacatgctttatctagatctccacatcaagtcggtaagccaattgctgctgccatcaatgacggcagaactcagttcttcgacaagcactgtccatgt tcccacaagatttgacctacgggcttctacaccaagaacaagccaaacggctccaacttggattacaccatcattgaagctaccgccataacagaggatggtgccat tatcccaggtcctgccgtcggtggtgcctcccagaaaatgctttcggtttccgacaacatcatcattgaagtcaacacacaaaaacaccatcctcgaaggtatacacgacatt gacttgccagtcaacccaccattcagacaacccacacacactgccgactacagaattggcagaacgccatcccagtcgaccagaaaaggttatcgc cattgtcgaaaccacagagggtgacaaggttccacccaaacacccccaagtgacgctcaatcccaagccattgctggtcacttgattgaattcttggaaaacgaagtc aaacaaggtcgtctcccagaaaaacttgcacccattacaatccggtatcggttaacattgccaacgctgtcgttgaaggtttggcccaatccaacttcaagaactgact gtctggaccgaagtcttgcaagattccttcttggaccttcttcgaatccggaagtttggacttgcaaccgccacctccatcagattgaccgaggaggaaggtttcaagaagtt ctacgaaaactgtggacgtctactccaagaagttgtgcttgagatcgcaagttgtctccaactcgccagaaatcatcagagaagattgggtgtcttggcttcaacacccc agttgaagttgacatctacgcccacgccaactccacccaagctcatgggttccagaatgttgaacggtttgggaggctcgggtgatttcttgagaaacgcccaagttgtc cgtcatgcacacccccatctgctcgtccaacccaaggttgacccaacccggtgtctcttgtattgtcccaatggccacccacgttgatcaaacccgaacacggatctcgatgtt gtcgttaccgaacaaggtttggctgatctcagaggtctcagtccaaaggaaagagctcaagttataatcaacaactgtgcccacccagactaccaggctcaattgc aagactacttgacagagctgtttctacgccaccaagaagaagatgttgcacgaaccacacatgttgaacgaagcttcgctatgcacttgaacttgcaagaaaatg gtaccatgaagttgaacaaaggtagacgggccaagttgtag	73
ACC1 (*Candida viswanathii* ATCC20336) contained in pAA1910	atgacagatcttcccaagtccaacagactccttaattacacacagttgcactcatcctgccatcacatttcttaggtgggaactcggtgctcaccgctgagccttct gccgtgacagatttcgtcaaaacacaccaaggtcacactgttatcaccaaagtcttgattgccaacaacggtattggtgccgtcaaagaaataagatccgtcagaa aatgggcctacgaaactttggtgacgaaagagctatacagtttgtcgccatggccactccggaagatatggaggctaacgccgagtacattcgaatggccgacca gtttgtcgaggtcccaggtggtaccaataacaacaactacgcgaatgttgacttgattgtcgaaatcgctgaaagaaccgatgtccacgccgtttgggctggttgggg tcatgcctccgaaaaccctttgttgccagraaggttggcagcttccctaagaagatcgtgtttattggtcctccaggtctgccatgagatctttgggtgacaagatttctt ccaccattgttgcacaacacgccaaagtgccatgtatcccatggtctggtgtcgaagaggtccacgtcgaccagaaaaccaagttggtgtctgttgtgacgac cacgtctacgccaaaggttgtgtacctcgccagaagacggtttggaaaaagccaaacgtatcggattcccagttatggttaaggcatccgaaggtggtggtggta aaggtatcagaaaagtcgaccacgaaaagggacttcatcagtttgtacaaccaggcggctaacgaaatacccaggtcaccaatttcatcatgaagttggccggtg acgccagacacttggaagtgcaattgtttgccgatcagtacggtaccaacatttcgctttttcggtagagattgttctgtgcaaagaagacatcaaaagatcattgaaga agctccagtcacaattgccaacaaagacacttttgttgagatggagaaagctgccgtcagattgggtaagttggttgattacgtgtctgccggtaccgttgaatacctttt actcctacgccgaagacaagttctacttttggaattgaacccaagattgcaagttgaacatccaactaccgaaatggtttccggtgtcaacttaccagccgctcagttg caaattgctatggctcccaatgcacagaatcagagacatcagattgttgtacggtgttgatccacactgccactgagatttcgagttcaagtcccccaaactc attgatcacgcaaagaaagccagctccaaagggtcactgtacgcttgtcgtatcacttctgaagatccaggtgaaggttcaagccaagtggtggtactcttcacg agttgaacttccgttcttcgtccaatgtctgggttacttctcggtcgccaaccaatcttctatccactccttgctgattcccagtttggtcacatttccgctttggtgaaaatc gtcaagcctctagaaagcacatgattgttgccttgaaggaattgagtatcagaggtgacttagaaccactgttgaatactgatcaagttgttggaggactccagatttcg ccgacaaccacatcactaccggttggttgagatgttgatcaccaagaagttgactgccgaaagaccagatcctatcgttgctgttgtctggtggccgtcaccaaag cccacatccaagccgaagaagacaagaaggagtacattgagtctttggaaaaggtcaagttccaaacaagtccttgttgaaaactatcttcccagttgagtttatct acgaaggtgaaagatacaagtttactgccaccaagtcctccgaagacaagtacacttttgttcctccaacggttctagatgtgtcattgtcgctcattgtctgatggtg gcttgttgtgtgctttggacggtaagtccccactctgtctactgtggaaggaagaagcagcggccactagattgtctgttgacggtaagacttgcttgttggaagttgaaaac	74

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1708 of 3077

Table 19

Gene (Organism)	Nucleic acid sequence of the gene protein-encoding sequence (from START codon to STOP codon except where otherwise noted)	SEQ ID NO
	gacccaacccaattgagaactccgtctccaggtaagttggtcaagtacttggttgagagtggtgaacacgttgatgccggccaatcttatgccgaagttgaagtcatg	
	aagatgtgtatgccttgattgcacaagaaaaacggtactgttcaattgctcaaacaaccaggttccactcttaacgctggtgacatcttggcaattcttggcattggacgat	
	ccatctaaagttaaacacgccaagccatatgaaggcacttgccagagatgggtgatccaactgttaccggttccaaacccagctcacttgttccaacattacgcacac	
	catcttgaagaacatcttggctggttacgataaccaagtcatttttgaactccacttgaagaacatgatggagatcttgaagaacaaggagttgccttattctgaatggga	
	gattgcaaatctccgccttgcattcaagaatcccaccaaagttggatgaggctttgacgtcctttgattgaaagaaccgaaagcagaggcccgaattcccagctcgt	
	cagatttgaagctcgtcaacaagactcttggtgaacccaggcaacgaattgttgggcgatgttgttgctcctctgtctccattgccaacccgctacccagaacggcttggtt	
	gaacacgagtacgactacttttgcttcattggttaacgagtactgcaatgttgaacacttctttagtggtgaaaacgtgagagaagaagaatgttatcttgagattgagaga	
	cgagaacaagtctgatttgaagaaggttatcagcatttgcttgtcccactccccgtgtcagtgctaagaacaacttgatttttggccatcttggaagcttatgaaccattgttg	
	caatcaacttcttcaactgccgttgccattagagaagatagtcgcagttggattctcgtgcttgtgccaaggttgtttgaaagctagagaagaacttttgattca	
	atgttcttgccatccatcaaggaaaagatctgaccaattggagaacacatttgagaagtgcagtcgttgagactcttatggtgaagtttcgccaagcacagagaaccta	
	aattggaaatcatccaagaagttgtcgaatccaagcacgttgttttcgatgtcttgtcgcaatttttggtccaccaagactgctggttgccattgctgctgccgaagtctat	
	gttagacgttcctacagagcttatgatttgggtaagatcgattacacattcatgacagattgccaattgttgaatggaagttcaagttgcctgcaaatcgcaggttccaga	
	tacaacgccgtccaatctgccagtgttggtgacgactcgaccactatgaagcatgctgcatctgttctgacttgttgttgttgattccaagagcgcaatccacttcca	
	gaactggtgttttggttccagctagacatttggacgatgttgatgagattctttctgctgcattggagtacttccaacccatctgatgcactctctttccaagctaagggagaa	
	agaccagagttgttgaatgttttgaacattgtcatcaccgacattgacggttactctgacgaagatgaatgcttgaagagaattcatgaaatcttgaacgagtacgaag	
	acgatttggtctttgctggtgttgctgctgttactttgtttgcccaccaagttggttcttatccaaagtactacaccttcactggtcagtgtatgaagaaaacaaggttatc	
	agacacatcgaaccagctttggcttccaattggaattgggaagattagccaacttttgacatcaagccaatttcaccaataacagaaaacattcatgtttacgaaagctat	
	tggtaagaatgctccttcggataagagattcttcactagaggtattattagagttggtgtcctcaaagatgaaatcagtcttactgagtacttgattgctgaatcgaacag	
	attgatcagtgnatacttggataccttggaagttattgacactccaactccgatttgaaccacatttcatcaacttctccaacgtttcaacgtccaaccagctgatgttga	
	agctgctttgcttcattcttggaaagattttggtagaagattgtggagattgagagttgtctgtaccgacccacaggcgcaactcattccc	
	attgcgtgccattatcaataacgtttcaggttatgttgtcaagtcggaattgtacttggaagtgaagaaccctaagggtgattgggtcttcaaatccattggccacccctgg	
	ctcaatgcacttgcaaccaatctcgactccatacccagtcaaggaatccttgcagccaaaacgttacagagctcacacaacatgggaaccacttttgtttacgatttccca	
	gagttgttccgtcaagccaccatttccaatggaagaagcacggcaagaaggctcctaaagatgtcttcactctttggagttgatcaccgatgaaaacgatgctttgg	
	ttgccgttgaaagagatccaggtgccaacaagattggtatggtttcaaggtcactgccaagaccctgaatacccacgcggacgttcattcatcattgttgccaa	
	tgatatcacccacaagattggttcctttggtccagatgaagatgaatacttcaacaagtgtaccgacttggccagaaagttgggtgttccaagaatttaccttctgccaa	
	ctccggtgccagaattggtgttgctgaagagttgattccatgtaccaagttgcttggaacgaagaaggtaacccagatataaggtttcagatactgtactgaaccca	
	gacgccaaagaagcttggaaaaagacggcaagggtgacactattgttactgaacgtattgtcgaagatggtcaagaacgtcacgttatcaaggccattattggtgc	
	tgagaacggcttgggtgttgaatgtttgaaaggttccggtttgattgctggtgccacttcaagagcctacagagacatcttcaccattaccttggtcactgtagatctgttg	
	gtattggtgcctatttggtcagattgggtcaaagagctatccaaattgaaggtcaaccaatcatttgactggtgcaccagctatcacaagtgttgggtagagaagagttt	
	actcgtcgaacttgcaattgggtggtacccagatcatgtacaacaatggtgtttcccactaactgccagtgacgatttggctggttgttgagaagatcatggaatggttgt	
	cctacgttccagctaagcgtggtatgccagtacaatcttggaaagtgaagatacctgggacagagacattgactactacccaccaaagcaagaagctttcgacat	
	cagatggatgatgcgaaggtaagcaagttgaaggtgaagagtttgaatctggttgtttgacaaaggttcattccaggaaacttatcaggatgggctaaaggtgttgtc	
	gttggtagagctcgtctcggtggtatcccaattggtgttgagaccagaactattgaaaacatgatcccagctgacccagcaacccaagttcactgaa	
	gccttgatccaagaagccggtcaagtctggtatccaaactctcgttcaagaccgcacaagccattaacgactcaacaacggtgaacaattgccattgatgatcttg	
	gccaactggagaggtttctcggtggtcagagagatatgtacaacgaggtcttgaagtacggttccttcattgttgacgcttagttgatttcaagcagccaatcttcactt	

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1709 of 3077

Table 19

Gene (Organism)	Nucleic acid sequence of the gene protein-encoding sequence (from START codon to STOP codon except where otherwise noted)	SEQ ID NO
	acatcccaccaaatggtgaattgagaggtggctcttgggtcgttgttgatccaaccatcaactccgacatgatggaaatgtatgccgacgttgactccagagctggtgt tttggaaccagaaggtatggttggtatcaaatacagacgggacaagttgttggctaccatgcaaagattggatccaactatgcccaattgaaggagaagttgaacg actcgagcttgtcgccagaagaacatgcccaagtcagcaccaagattgtcaagcgtgaaaaggcattgttgccaatctatgcccaaatttctgtccagtttgccgactt gcacgacagatccgggacgtatgatggctaaaggtgtcattagaaaagaaatcaagtgggttgacgccagacgtttcttcttctggagattgagaagaagaattgaac gaagagtacgttttgaagttgattggtgaacaggtcaagaatgccaacaagttggaaaaaggttgcccaggttgaagagttggatgccaactgttgactacgacgatga ccaagctgtcagtacttggattgaagagaaccacgccaaattgcaaaagagagttgaagaatttgagacaggagagaagaacaagtccgacattgtcaaattgttgc aagaagacccatcaaacgctgcctctgttatgagggatttcgttgatagattgtccgatgaagaaaaggaaaagttcctaaatcattgaactag	
ACC1 (*Candida viswanathii* ATCC20336) contained in pAA245 (includes 5' intron sequence)	atgagatgccaagtatctcaaccatcacgatttactaacttgcttgtacatagactcccacgaacactacttaattatccagttgtaaataccctatttattctagacgtca ttattccctaatttttcattcaagaacctactaaagaaaatgacagatctttcccaagtccaacagactccctaatttacacacagttgcactcatcctgccatcacatt tcttaggtgggaactcggtgctcaccgctgagccttctgccgtgacagatttcgtcaaaacacaccaggtcacactgttatcaccaaagtcttgattgccaacaacg gtattggtgccgtcaaagaaataagatccgtcagaaaaatgggcctacgaaacttttggtgacgaaagagctatacagtttgtcgccatggccactcccgaagatatg gaggctaacgccgagtacattcgaatggccgaccagtttgtcgaggtccccaggtggtaccaataacaacaactacgcggaatgttgacttgattgtcgaaatcgctga aagaaccgatgtccacgccgtttgggctggttgggctcatgcctccgaaaaccccttttgttgccagaaaggttggcagcttcccctaagaagatcgtgtttattggtcctc cagggtctgccatgagatctttgggtgacaagatttcttccaccattgttgcacaacacgccaaagtgccatgtatcccatggtctggtactggtgtcgaagaggtcca cgtcgacccagaaaccaagttggtgtctgttgacgaccacgtctacgccaaaggttgctgtacctcgccagaagacggtttggaaaaagccaaacgtatcggattc ccagtttatggttaaggcatccgaaggtggtggtgtaaggtatcagaaaaggtcgaccacgcgaaaaggacttcatcagtttgtacaaccaggcggctaacgaaata ccagggtcaccaatttttcatcatgaagttggccggtgacgccagacacttggaagtgcaattgtttgccgatcagtacggtaccaacatttcgctttcggtagagattgt tctgtgcaaagaagacatcaaaagatcattgaagaagctccagtcacaattgccaacaaagacactttgttgagatggagaaagctgccgtcagattgggtaagtt ggttggttacgtgtctgccggtaccgttgaatacctttactctacgccgaagacaagttctacttttttggaattgaacccaagattgcaagttgaacatccaactaccga aatggtttcggtgtcaacttaccagccgctcagttgcaaattgctatgggtctcccaatgcacagaatcagagacatcagattgttgtacggtgttgatccacactctg ccactgagattgatttcgagttcaagtcccaaactcattgatcacgcaaagaaagccagctccaaagggtcactgtaccgcttgtcgtatcacttctgaagatccag gtgaagggttcaagccaagtggtggtactcttcacgagttgaacttccgttcttcgtccaatgtctgggggttacttctcggtcgccaaccaatcttctatccactccttgctg attcccagtttggtcacatttccgccttggtgaaaatcgtcaagcctctagaaagcacatgattgttgccttgaaggaattgagtatcagaggtgactttagaaccactgt tgaatacttgatcaagttgttggagactccagatttcgccgacaacaccatcactaccggttggttgatgagttgatcaccaagaagttgactgccgaaagaccaga tcctatcgttgctgttgtctgtggtgccgtcaccaaagccccacatccaagccgaagaagacaagaaggagtacattgagtctttggaaaagggtcaagttccaaaca agtcctgttgaaaactatcttccagttgagtttatctacgaaggtgaaagatacaagttttactgccaccaagtcctccgaagacaagtacacttgttcctcaacggtt ctagatgtgtcattggtgctcgctcattgtctgatggtggcttgttgtgtcgttgaacggtaagtcccactctgtctactggaaggaagaagcagcggccactagattgtc tgttgacggtaagacttgcttgttggaagttgaaaacgacccaaccaacccaattgagaactccgtctccaggttaagttggtcaagtacttggttgagagtggtgaacacgtt gatgccggccaatcttatgccgaagttgaagtcatgaagatgtgtatgcctttgattgcacaagaaaacggtactgttcaattgctcaaacaacccaggttccactcttaa cgctggtgacatcttggcaatcattggcatttgaacgatccatctaaagttaaacacgccaagccaatgaaggcacttgccagagatgggtgatccaactgttaccgg ttccaaaccagctcacttgttccaacattacgacaccatctgaagaacatcttggctggttacgataaccaagtcatttgaactccacttgaagaacatgatggaga tcttgaagaacaaggagttgccttattctgaatggagattgcaaatctccgccttgcattcaagaatcccaccaaagttggatgaggcttgacgtccttgattgaaaga accgaaagcagaggcgccgaatttccagctcgtcagatttgaagctcgtcaacaagactctggtgaaccaggcaacgaattgttgggcgatgttgttgtcctctt gtctccattgccaaccgctaccagaacggcttggttgaacacgagtacgactacttgcttcattggttaacgagtactgcaatgttgaacacttctttagtggtgaaaac	75

Table 19

Gene (Organism)	Nucleic acid sequence of the gene protein-encoding sequence (from START codon to STOP codon except where otherwise noted)	SEQ ID NO
	gtgagagaagaagatgttatcttgagattgagagacgagaacaagtctgatttgaagaaggttatcagcatttgcttgtcccactccgtgtcagtgctaagaacaac	
	ttgatttggccatcttggaagctatgaaccattgttgcaatccaactcttcaactgccgttgccattagagattctttgaagaagatagtccagttggattctcgtgcttgtg	
	ccaaggttggtttgaaagctagagaagaacttttgattcaatgttctttgccatccatcaagggaaagatctgaccaattggaacacattttgagaagtgcagtcgttgagactt	
	cttatggtgaagtttccgccaagcacagagaacctaaattggaaatcatccaagaagttgtcgaatccaagcacgttgtttcgatgtcttgtcgcaatttttggtccacca	
	agactgctgggttgccattgctgctgccgaagttctatgattgggtaagatcgattaccacattcatgacagattgccaattgttgaa	
	tggaagttcaagttggctcaaatcgcaggttccagatacaacgccgtccaatctgccagttggtgacgactgaccactatgaagcatgctgcatctgttctgactt	
	gtcgtttgttgttgattccaagagcgaatccacttccagcaactttttggttccagctagacattggacgatgttgatgagattctttctgctgcattggagtacttccaa	
	ccatctgatgcactctctttccaagctaagggagaaagaccagaagttgttgaatgtttgaacattgtcatcaccgacattgacggttactctgacgaagatgaatgctt	
	gaagagaattcatgaaatcttgaacgagtacgaagaagacgatttggtctttgctggtgttcgtcgtgttactttttgtttccgccaccaagttggttcttatccaaagtactacac	
	cttcactggtccagtgtatgaagaaaacaaggttatcagacacatcgaacccagcttggctttccaattgggaagattagccaacttttgacatcaagccaatt	
	ttcaccaataacagaaaacattcatgtttacgaagctatttggtaagaatgctccttcggataagagattcttcactagaggttattatgaggtggtgtcctcaaagatga	
	aatcagtcttactgagtactgattgctgaatcgaacagattgatcagtgatatatcttggaagttattgacacttccaactccgatttgaaccacatttttcatca	
	acttctccaacgtttcaacgtccaaccagctgatgttgaagctgctttttgcttcattcttggaaagatttggtagaagattgtggagattgcaagttgtgctgaaatc	
	agaattgtctgtaccgacccacagggcaactcattccaattgcgtgccattatcaataacgtttcaggttatgttgtcaagtcgggaattgtacttggaagtgaagaaccct	
	aagggtgattgggtcttcaaatccattggccacctggctcaatgcacttgcaaccaatctcgactccatacccagtcaagggaatccttgcagccaaaacgttacag	
	agctcacaacatgggaaccactttgtttacgatttcccagagttgttccgtcaagccacccatttcccaatggaagaagcacggcaagaaggctcctaaagatgtcttc	
	acttctttggagttgatcaccgatgaaaacgatgcttggttgccgttgaaagagatccaggttgccaacacaagattggttgtttcaaggtcactgccaagaccc	
	ctgaatacccacgcggacgttcattcatcattgttgccaatgatatcacccacaagattggttccttggtccagatgaagatgaatactccaacaagtgtacccgacttg	
	gccagaaagttgggtgttccaagaatttaccttctgccaactccggtgccagaattggttgctgaagagttgattccattgtaccaagttgcttggaacgaagaaggg	
	taacccagataaaggtttcagatacttgtacttgaacccagacgccaaagaagcttggaaaaagacggcaaggggtgacactattgttactgaacgtattgtcgaag	
	atggtcaagaaacgtcacgttatcaaggccattattggtgctgagaaacggcttggtgttgaatgtttgaaaggttccggtttgattgctggtgccacttcaagagcctaca	
	gagacatcttcaccattacctggtcactgtagatctgttggtattggtgcctatttggtcagattgggtcaaagagctatccaaatcatttggact	
	ggtgcaccagctatcaacaagttgttgggtagagaagtttactcgtcgaacttgcaattgggtggtgctacccagatcatgtacacaacaatggtgtttcccacttaactgccag	
	tgacgatttggctggtgttgagaagatcatggaatggttgtcctacgttccagctaagcgtggtatgccagtaccaatctggaaagtgaagatacctgggacagaga	
	cattgactactacccaccaaagcaagaagcttcgacatcagatggatgatcgaaggtaagcaagttgaaggtgaagagtttgaatctggtttgtttgacaaaggttc	
	attccaggaaacttatcaggatgggctaaggtgttgtctggtagatcgtcgtctcggtggtatcccaattggtgtcatttggtgttgaggaccagaactattgaaaacat	
	gatcccagctgaccccaacccaagttccactgaagccttgatccaagaagccggtcaagtctggtatccaaactctgcgttcaagaccgcacaagccattaa	
	cgacttcaacaacggtgaacaattgccattgatgatcttggccaactggagaggtttctctggtggtcagagagatatgtacaacgaggtcttgaagtacggttccttc	
	attgttgacgcttagttgatttcaagcagccaatctcacttacatcccaccaaatggtgaattgagaggtggctctgggtcgttgttgatccaaccatcaactccgaca	
	tgatggaaatgtatgccgacgttgactccagagctggtgtttggaaccagaaggtatgttggtatcaaatacagacgggacaagttgttggctaccatgcaaagat	
	tggatccaacttatgcccaattgaaggagaagttgaacgactcgagcttgtcgccagaagaacatgcccaagtcagcaccaagattgtcaagcgtgaaaaggcat	
	tgttgccaatctatgcccaaatttctgtccagtttgccgacttgcacgacagatccggacgtatgatggctaaaggtgtcattagaaaagaaatcaagtgggttgacgc	
	cagacgtttcttcttctggagattgagaagaagattgacgtttgaagttgattggtgaacaggtcaagaatgccaacaagttggaaaaggttgccag	
	gttgaagagttggatgccaactgttgactacgacgatgaccaagctgtcagtacttggattgaagaaccacgccaaattgcaaaagagagttgaagaattgag	
	acaggagaagaacaagtccgacattgtcaaattgttgcaagaagacccatcaaacgctgcctctgttatgagggatttcgttgatagattgtccgatgaagaaaagg	

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1711 of 3077

Table 19

Gene (Organism)	Nucleic acid sequence of the gene protein-encoding sequence (from START codon to STOP codon except where otherwise noted)	SEQ ID NO
	aaaagttccttaaatcattgaactag	
ACS1 (*Candida viswanathii* ATCC20336)	atgccagaatcaactcaacaatctcacatttcttggaccatgaaaagatgcagcaacctccaactggcttcaccgaacgctccaccgcaaagccaaacttggccg actttgacacctacaagaagttgtacaagcagtcggtagaaaacccaaacgagtttttcacagaacaagcaaacaaaaacttggactggttcaaaccttcgacttg gccagattcccgtcgatccaaaggacgacttcaagaatggcgacttgcctgcctggttcatcaacggtcagttgaacgcttcttataacgcggttgatagatgggcc atcaagaacccagacaagccagctatcatttatgaaggtgacgaaccagacctgggcagaatcatcacttatggcgaattgttgaaagaagtcagtaagttggcc cagacttgaccaagttgggtgtcaaaaaaggtgactctgttgctgtctactgccaatgatccctgaagctatcgtcaccttgttggccattgtcagaatcggtgccgtg cactcggttgtcttgctggtttctcctccgcttccttgagagacagaatctggatgctgactccagaatcgtcactcactgccgacggagtccaagagaggttggtaagac cattgaaaccaagaagattgtcgatgacgccttgaaggaatgtcctaaggttagaaacgttattgtttcaagagaaccggcaactcccacgttcctttctctgctggta gagacttgtggtggcacgaagaaatggccaaataacggtccatactcccaccagtgccagtcaactccgaagatccattgttcttgttgtacacttctggttctactggt aagccaaagggtgtccaacacaaccaccgccggttacttgttgggcgctatcttgaccaccaagtacactttgatgtccacgaagaagacgtcttgttcaccgctggt gacgtcggttggatcacgggtcacacctactgtgtctacggccattgttgtggtgccaccaccgttgtctttgaaggtaccccagcttatccaaactactccagata ctgggagattgtcgaccaatacaaggtcaaccaattctacgttgccccaaccgctttgagattgttgaagagaggccggtaccaagtacgttgaaaaatacgacttgtc ctccttgagagtcttgggctctgtcggtgaaccaattgcggccgaggtctggcactggtacaatgacaacattggtagaggcaaggcccacattgtcgatacctactg gcaaaccgaatctggttccacttgttgaccccattggccggtgtcactccaaccaagccaggttccgcttccttgccattctttggtgttgttccaaagatcttggaccca accaccggtgaagaattagaaggtaacgatgttgaaggtgtccttgccattaaatccgcttggccatccattactagaggtatttacaacgattacaatagatttatcga aacctacttggcccatcaccacgaccactactttccggtgatggtgctgctagagataacgacggcttctactggatcttgggtagagtcgatgatgtcgtcaatgtttc cggtcacagattgtcgactgctgaaattgaagctgccttgattgaggcacccattggttgctgaatccgctgttgttggttatgctgacgacttgactggtcaagccgtcgct gcctatgtgtccttgaagaaagataaagttattggtgacgacatggaggccagtcaagaaggagttgatcttgactgtcagaaaaagaaattggtccatttgctgcacca aagttgatcttgttggttgacgacttgccaaagaccagatccggtaagatcatgagaagaatcttgagctggtgaagaagatcaattgggtgatatct ccaccttgtccaacccaggcgttgtccaacaaatcattgaaattgtccatgcctgcaagaaataa	76

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1712 of 3077

Table 19

Gene (Organism)	Nucleic acid sequence of the gene protein-encoding sequence (from START codon to STOP codon except where otherwise noted)	SEQ ID NO
ACS2 (*Candida viswanathii* ATCC20336)	atgactacagactcaaacacccacaaggttgttcacgaagccaatggcgtcaagttgagacaaacccccaaaggaattcttccaaaggcaaccaaacaaggga cacatccaagacacggaagagtacaagaaattgtatgagcaatccataaaagaccccaaggtttcttggcccattggctaaagaattgttgtcctgggacagcg atttccacaccgtcaagtcaggtaccttaagaacggtgacgctgccttggcggtaagttgaacgcttcttacaactgttgccagacacgctttggccaac ccaaacaagccagccatcatctacgaagcagatgaagaaaggggattcttacatcttgacctacggtgatttgttgagagaagtttccaaagtcgctggtgtcttgcact cctgggcatcagaaaggtgacacgttgctgtttacttaccaatgaatgcccaagccctcattgccattggccattgctagattgggtgctgcccactccgtcatc tttgctgggttctcctccggttctatcaaagacagagtcaacgacgccagctgtaaagcgcttattacctgtgatgaaggtagaagaggtggcagaaccaccaacat caagaagttgtgcgacgaagctttgaagcaatgcccaacggtcaagcacgtcttggtccacagaagaacggcaacccagaaatcaaattgaccgaaggcag agattactactgggacgaagaaaccgccaagtttcccggatacttcccaccccgtgccagttgactctgaagacccattgttcttgttgtacacttctggttccactggtact ccaaagggtgttgtccacaccaccgccggttacttgttgggtgctgctctttccaccaaatacatctttgacgtccacccagaagatgtctcttcactgctggtgatgtcg gttggatcactggtcacacctacgctttatatggtccattgttgttggtgttcctctgttgtttttcgaaggtactccagcttacccagactacggaagattctggcaaattgtt gaaaaacacaaggctacccacttctacgttgctccaactgcttttgagattgttgcgtaaagctggtgaacaagaagttggcaagtacgactgagttcattgagaaacc ttgggttctgttggtgaaccaatctccccagatatttgggaatggtacaatgaatttgtcggtaaggaccaatgtcatatctcggatacctactggcaaaccgaatccgg ttctcacttgattgctccattggctggtgccattgccaacaagccaggttctgctcttacccattcttttggtattgaagctgctttgatcgacccagtcactggtgttgaaatc aagggcaacgatgttgaaggtgtgttggttgtcaaggaccactggccatccatggcaagaactgttttcaacaaccacgtcaagtacatggatacctacatgaaccc atacccaggctactatttcactggtgatggtgctgccagagataacgacggttactactggatcagaggtagagttgatgatgtcgtcaatgtttctggccacagattgt ccaccgccgaaattgaatctgcattgattgaagacccaagagttggtgaatcgcgtgttgttggtatcaacgatgaccttactggtcaagctgttattgcttacgttgcctt gaaatccggcgacgttgctgatgaagacgcttgagaaaggaattgatctgttggttagaaaagaaattggtccatttgcggctccaaaatctgtcatcttggttcaag acttgcctaagaccagatctggtaagatcatgagaagaatcttgagaaaagtcagttctaacgaagccgatcaattgggtgacatcaccacttgcaaaacccagg ttctgttgaaggtatgatttctgccttcggtgctcaattcggtagaaagtaa	77

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1713 of 3077

Table 19

Gene (Organism)	Nucleic acid sequence of the gene protein-encoding sequence (from START codon to STOP codon except where otherwise noted)	SEQ ID NO
TES1-1 (*Candida viswanathii* ATCC20336)	atgcttacactaacttctggtcccaaccccttgcctgatttcgaagaagccttgagagtcatcaaagtggacgatacacactatgttggtgcccattcgttgagattgcct gtgaaaggaggacgaggcgtttatggtggccatatgattgcccagtctttgctagtgggtattgaatcaacaagagatgacaagactaataaagtgttcattcctgact cgtaccacctgtactttattggagccggaaatgccaaaatcccatgaactacacggtagagaaactatacgacgatgaaaacgtcagcaagcgtttcatcattgc ggaacaaaagggaagacatcgtttgacttgttggttacattgaggagaccccggtaccaagccttccacgacaggcgacaacttggacatcagcatcccgtgccc aagatccagttaaaacaccctgatcccgacaagttgcaccaagttcagcacacggacttattaggaatgcattcggtaaagagcttatggattacagaggagtgccc tgaagagaacgaacttatgccgcagaaagatggttgactgttcaccggttattcgaaaccagcctaaaccgggcgccagtttggaaaacggtcgtcgaagaattg ccagacgcccagggccagatgcacacagtggagaagagtatttaagacccaaagattcccagtcgttcaaggaccctatctacaacttgtgggtttagccgacttt gtctgattcggcattcctcaccaccatgcgagaatcttgcacatcccatgggcgccttccatcgaaatcgacgcacatacgatccagcaagaagatgcaacgtat atcatgcgttcaacttgaatgcagcccatatcttccattacaatgcgatgtcattggatcaccatatttatttccacaatgaagattacacttcgacgacggttcaggttt cgacatttgcaaagattggttggctttcacttatcagatgaaaagactctctaataaccgtaccctcgtcgtgggttcttgttccaacgaaaaacacaaatgtatcgcca cagtggtgcaagaagggttgactattatgcaaaatggagttggtcgaactgctgataaatcacgtctatag	78
TES2-1 (*Candida viswanathii* ATCC20336)	Atgattgagaacatttcaggaaacggaaactacccacagaatcacgaggttgacttggagaaaggaattcggagttgaaaaaataggtatcaacttgtaccgtggc aagagcccaataccgaagccagacagacgttccagaggtgcttacggtggttacttggccggtcaagctttgttagttgccatgaagtctacccccgccagagtaca gaccacacctgttccattcctacttttatcaaagctgtcaatgacaaggagacattggagtggagagttgaagaaacatccaacggcagaaactacgccaacagat ccttgcaagccttccaagccggtaacttggtctacacagccaacgtctcctgaccaagaaaaactctgccaagaaagctgaggaagcaacgggtgtcaagcct ttgagttccagggaaaaccacacgaatggtttgaaaaacacaagagagatgaccttcctcttgctacacctcttcatcccttttgatctaccacaaattcttcccagag gttgtctcgttggaagcttccaaagaagaagaatcaaagccagcagccgacagagaattgagctggtacttcaaatgggtatcaacaacgaagaggccacc accagccattagtcaacttgaactccgattaccaatatgttggaatggctgcgttgaccgatgctgtctacttgaacagattattgagaatttgagagttgaagacgctg accacacccagcttgtccactactttccgtcctccttggaccatactatgtacttccatgacgacgacttgacgtaaccaaatggatgggattcacttcaaggttacca gattcagtcacaaccgtgcattgtgccagggtgaagtttataacgacaagggggtacacgtgtgtacaattgtccaggaagggctcatgatgcttaacggtctgaac aaggtgctaaattataa	79

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1714 of 3077

Table 19

Gene (Organism)	Nucleic acid sequence of the gene protein-encoding sequence (from START codon to STOP codon except where otherwise noted)	SEQ ID NO
TES2-2 (*Candida viswanathii* ATCC20336)	Atgattgagagcatttcagaaaacggaaactacccacagaaccacgaagttgacttggagaaggaattcggagttgaaaaaataggtatcaacttgtaccgtggc aagagcccaataccgaagccagacagacgttccagaggtgcttacggtggttactggctgcaagctttgttagttgccatgaaatctacccgccagattacag accacatctgttccattcctacttatcaaagctgtcaatgacaaggtgacattggagtggagagttgaagaaacatccaacggccgaaactacgccaacagaggcc ttgcaagccttccaagccggtgacttggtctacacagctaacgtctccttgaccaagaaaaactccgccaagaaaactgaggaagcaacgggtgtcaagccctttg agttccagggggaagccacacgaatggtttgataagcacaagatagatgacctccttgctacgccttcttcatccctttgatctaccacaaattcttcccagaggttgt gtcgttggaagcttccaaagaagaggaatcaaagccagcagctgacagagaattgagctggtatttcaaatgggtatcaacaacgaagaggggccaccaccag ccattagtcaacttgaactctgattaccaatatgttggtatggctgcgttgaccgatgccgtctactgaacagattattgagaatcttgagagttcaagacgctgaccac acccaacttgtccactactttccgtcctccttggaccacaccatgtacttccatgacgacgacttttgatgttaccaaatggatgggattcactttcaaggttaccagattca gtcacaacgtgcattgtgccaaggtgaagtttacaacgacaaggggtacacgtgtgtacaattgttcaagaaggactcatgatgcttaacggtcttgaagaaggt gctaaattataa	80
TES3-1 (*Candida viswanathii* ATCC20336)	Atgagtcacccaacgccagaagaagtctacggagtcaccaaagttgccgaaaacaaatacgtgggaacagaccctgaacaagcccacgccgaaaacac gcggagtgtatggcggaaacttctgtgcccaggccatcttggtagccatcgagtcagccccggaagggttcacaccacactccatccacagcaacttcattcgtggc ggcgacccgaagtcccgttgagtgggaggtcgaggtcatctcgaacggcaagtctttcgccaaccgcatagtcaagggcgtgcagcacggcatcgtggtctac gtcgccacggtctcgttgaccaacaagaactcaaccaccccgcaacgagagcttcacgtacgacacgccccggacgagaccgtcaagacgtacggcaacgcc gagctagacacatactaccagggctggctctacttggaggtcaagaactacccaagcagttgcactcgcaccagatctcgtactccgtcaagtgggggccggag aacgacgcgtggaaggacgccagcagacgtaccagtttgttgggctcgctgcgatctccgacgtgcttgacctcggccagatcttgcgcaacttggacatccactt gagcacgcccaagttcaacgtgtcgctcgaccacagtgtgtacttccacggcgcggacttgacgtgaccaggtggtcgacgaccacgatccgcatgacgaagct cgcgcacggccgcgcgttgatagagggcgagatgtactcggaccagggcaggcacatcgccagcatagtgcaagagaggttgtacatagctgagtcgccgaa gctctga	81

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1715 of 3077

Table 19

Gene (Organism)	Nucleic acid sequence of the gene protein-encoding sequence (from START codon to STOP codon except where otherwise noted)	SEQ ID NO
TES4-1 (*Candida viswanathii* ATCC20336)	atgacgctgattgcaccccgccacagacatacccgccgtggacatctccaccatatacgacgtcaaacagatcgacgccaaccgctaccgcggcgttcggccgct ccagaaacaggcacgcgagttccgcgggagtctttggaggcaacttggttgcccagtccgtggtggtggcctgcccgtcgtgcccgcaggattccacccccaactcc gtccacgcctactacgtgccgtgccgtgacccgacgagacgccaattgagtgggaggtcgaagagacaaccacgggccgcaccttggccaacccggctgatacgcg gattgcagaacaagaaggtggtgttcacggcgagcatctcgttgaccaggaagaactccaatgcggaagtgatccaaaagaccgggcacccgtcgttgcagttc cagcgggacaccgagccatactacgagcagatgcgagccgccctggcgagtgcaaggtgctctacattaacgccaacacgcacctcacggtcagacagttcc ctgaagtcagctcccgcgacgtgttctcgttcttggtgcggtttggcaacgacgggcgggagaggatcgtgggggatgactcccgagtaccagtacgccgcgttggct gcgttgtcggactgggtgaggctccgtttctacttgacaatatgggggtcgacgtcagacgtcgttcgacgtgtcgttggaccacagcatctactttccacgacgacg ggttcgacgctaccgagtacttggtttttctccgtcaaggtgtccgggatcagccacagccgggtgctttacgacgggcagatcttcaacgaccgtggagcccacgttg ctagcatcagacaggagcgcttatatgtcgtcagcaacaagccaaagtttag	82
TES5-1 (*Candida viswanathii* ATCC20336)	atgcccacattcaactacaaggacggcgaaacaatagacgtgcagaaaagagtttggcgtggtcgagacagcgcccaacaagtacgtcggcgtgaagccactc gtcaagccgatgcctcatgtaaagggagtctttggcggtaacctgcaggacaagcattgctagtcgccatgaagtcggttggcccagactttagcccacactccttg cattcgtacttatacgccgccggcagcgaccagacacccgtcgagtggactgtgcaggcgatatctgatggtaactcgttttgtaacaggtttatcaagggagtccaa aacggacaggtcatatacattgcgaatgtgtcgttgactaaacgcaactctgctgctgatgcgatgaagaagtacgaggagtaccatgcgcagatccgccagaag ggcaaggacggcgatgctgatgaagaggacgaggatgacgatgacgaggatgataatgccccagctaagccatttgggttccagactccgagccatgtgga tcaaggaccgcgatctcgacaagttgccagtttccgacatggagtcgaatcttttgttgtactacaagttgcctcccgagttcgtgagcttgaagtcttcaacagaggaa gagctgttgccagtctcagagcgcaggatgggtgcttggcgaagtgggtattgagaacgagcaggttcaaccagcctttgactaatctcgacaagtcgttcca gtatgttggtttggccaatatcactgatgggttgtacttgggcacattgaataggatattgcgcattgatgacttgacacttgatgagcgggcgactaactacttttagcgta tcattggaccacgtcatctatttccacgatgatgatttgatgtgaccaagtgggtgatgggcttacgttcaggtgttcgagatattcacataatcgggttatttttgaaggtgag atctacagtgacaagggagtccaggttgcatctattattcaagagggcttagttagatttaaagacggctacttgaaaaacgccaagttgtga	83

420

Table 19

Gene (Organism)	Nucleic acid sequence of the gene protein-encoding sequence (from START codon to STOP codon except where otherwise noted)	SEQ ID NO
TES6-1 (*Candida viswanathii* ATCC20336)	Atgtcaaccaccggaatctacaacgccaaacccgcggctgccatattggaaacaaaaaccaaggtcaagttagtttatactgaaggaccgaaattggtctacgaa ggattgcacccccgttgagaccataaagaaactgttagaggtacttatggtggtgatttttattgctcaagggggtgaacgcagcttgggaatcaattggcaacaggatg gattttcaacctcattcattgcacgcatactatataaaaagcggttcccaggagtcagtattacgctgggaagtcttcaaggtgagtgatagcagaagttttgccaacc ggatgttaactgcataccagacgcacaccaacaaattggtgttcacgatgcagatatcgttcaccaaagacaacaacgaggagatcaagcgtgcggaatacgaa aagttgttggccagcggtggcaaaatcagaagcatcccatttgtcatcaaaaaggctccgaatgaaaagtatttcaagttacgagacagaatcaatgagcttcattat tctgagtatacccaatgggatcattgccatggctcccccttgacttccatgagtatgccactgagatgaactatgacacctgggaaacaaagagtacgggacgttta tgaaaatcttggataactattcctaggcaaggattattctaaacaatcctcttgggttgtcattcttatccgatggtgtttgggtattttgcttactagagcttgggttgcc acttggtagctacgaaaaggagttctttcgtgtcagtttggaccacacggtatatttccacgaacgctaatttcgactccagtgagtggctcttctagatttaggtttgtgaa catgaagaacaacagattcttgactgtgttgaactactatacttttgcaaggcaaacttattgcgactgtgctccaagaagtctacacttctttacatcagggtatcattgac aagtcgcaagaaatagctgtcaaaagcggaaataagaagcaggtcacaactccaaagttgtag	84
TES7-1 (*Candida viswanathii* ATCC20336)	atggacaaactacaagctgaagtatacgaggccgaacctcccgtggccaaattggaagccaaaacagcagccaaattgatatcttctgatgggcaaaattgac ctacgaaggagtgtaccctgtggagctagtcagaaaaggtctcagggaacctacgggggagatttcatagcacagggaatcaatgttgcctgggagtcaattgg caataagactgactttcaaccccattcattacatgcatatttttgtcaaggcaggatccgacctgtctgtcctacgttgggaagtgctcaaggtgagtgatagcagaaattt tgccaaccgtctaatgttagcgtaccagactcacaccaacgaattggtgttcacaatgcagatatcgttcactaaggataacaacgaggagatcaagcgtgccgag tacaagcaattattgcagagtggcggtaaaataagaagtatcccgtttgcaatcaagaaaccacccaatgaaaagtactttaaattgaaagacaaggtggacgat cttccatatttgagcacaccaatgggaatatggcgactgcgattccaccagacttccttgaatacgccacagagatgaatcacgacacggtgggggaacaaagagt ttggtatttcatgaaagtattggatgattactccttgggcaaaaactatgagagacagtcgttcttgggttggccttcttgtctgatgccgtgtggttatcgagtttcacccc agccttagggttgccattgggcactttagaaaggaaattctttcgtgttagcttagaccacaccatgtatttccatgacgccaacttttgattctagtgagtggatctttgtgg atttcaggtttgtgaatttgaacaacaacaggctcttgggtgttgtgaactttacacgttacagggtaagcttgttgctactgtgattcaagaagcgtacatgttttgcacc aggcgatcatagacaaatcacaagagatagctgtgaaaaaagcgggcataagaagcaggtgatcaccccgaaattatag	85

Table 19

Gene (Organism)	Nucleic acid sequence of the gene protein-encoding sequence (from START codon to STOP codon except where otherwise noted)	SEQ ID NO
TES7-2 (*Candida viswanathii* ATCC20336)	atggacaaattacaagctgaagtatacgacgccgaacctcctgtggccaaattggaagccaaaacagcggccaaattgatatctgctgatgggcgaaattgatc tacgaaggagtgtacctgtggagctagtcagaaaaggtcttagggaacctacgggggagatttcatagcacaaggaatcaatgttgcctgggagtcgattggc aataagactgacttcaacccattcattacatgcatatttgtcaaggcagggtccgaccagtctgtttacgttgggaagtgctcaaggtgagtgatagcagaagcttt gccaaccgtctaatgttagcatatcagactcacaccaccaattggtcttcacgatgcagatatcgttcactaaggataacaacgaagaaattaaacgtgctgagta caagcagctattggagagtggcggtaaaataagaagtatccgttgcaatcaagaaaccacccaatgagaagtacttcaaattgaaagacaaagtggacgacc ttccatatttgaacataccaatgggaatatggcaactgcaatacctgatttccttgaatacgccacagagatgaatcacgacacggtgggggaacaaagagttt ggtatttcatgaaagtattagatgattactccttggataaaaactatgagagacagtcgttcttgggttggcctcttgtctgacgccgtgttggttgtcgagtttcacgcca gccttagggttgccatttggcactttgaaaggaaattctttcgtgttagtttagaccacacccatgtatttccatgacgccaacttttgattctagtgagtggatcttgtggattt caggttgtgaatttgaacaacaacaacaggctcttgggtgtggtgaacttttacacttacagggtaagcttgttgctactgtgattcaagaaagcgtacatgttttgcaccagg ggatcatagaccaatcacaagagatagctgtgaaaagagcgggcataagaagcaggtgatcaccccgaaattatag	86
TES8-1 (*Candida viswanathii* ATCC20336)	Atggagagattacaggccgaagtctacgacgctaaaccaacagttgctaaattggaaaccaagactaccgtcaaattagttgctcaggacggaaagagattggtt tatgaagctatttatccagttgaacatatcaaagaagggattccaggagcttatggaggtgataccttggcccaagggatgaatgctgcgtgggactcccttggcgac aagaaagatttccagcctcattcagttcattcgtactttgttaagcctgcaacaaacaaatcaattttgagatgggaagtgatcaaggtcagtgatggcagaagttttgct aatcgtttcgtgtctggctatcaaacccacacaacaatgcgttggtcttcactatgcagatctccttaccaagtacaacgatgaagtggtcaagattgctgaatacaacaa gttgttggaaagtggtggcaaaatcagaagtattccattgtcatcaagaaagctccgaatgagaagtacttcaaatttaaggacaatcttggagatctcagatacatt gagcacaccaatggcaacattgccacagccatgctggaagatttgtttgaatacgccactgagatgaaccatgatactctcggaaatcaagagtttggtatttttcatg aaggtcttggataattactcgttaggtagtgactacactaaacaatctacttgggttggcattcttatcggatgctatttggatgtctgtttgttcaagagccttgggttgc catttggtagctaccatagacaatttttttggagttagcatggaccactcgatgtatttccacgacgctaatttcgactccaccgagtggatctttttagacttaggtttgtcaa tttgaaaaatgacaggctcttgggcgttgctaatttttacacttagagggtaagcttatttccacttgattcaagaagcttacctcttttacacccaggaatcattgccaa gtcccaagagatagctgtcaagagtggaaacaagaggcaagttaaaatgccaaagttgtag	87

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1718 of 3077

Table 19

Gene (Organism)	Nucleic acid sequence of the gene protein-encoding sequence (from START codon to STOP codon except where otherwise noted)	SEQ ID NO
Engineered DNA *TES3-1*$^{\Delta pts}$	atgagtcacccaacgccagcaagaagaagtctacggagtcaccaaagttgccgaaaacaaatacgtgggaacagaccctgaacaagcccacgccgaaaacacg cggagtgtatggcggaaacttctgtgcccaggccatcttggtagccatcgagtcagcccgggaagggttcacaccacactccatccacagcaacttcattcgtggcg gcgacccgaagtccctgttgaagtgggaggtcgaggtcatctcgaacggcaagtctttgccaacacgcatagtcaaggcgtgcagcacggcatcgtggtctacgt cgccacggtctcgttgaccaacaagaagaactcaaccacccgcaacgagagcttcacgtacgacacgccccgggacgagaccgtcaagagctacggcaacgccg agctagacacatactaccaggggctggctctacttggaggtcaagaactacccaagcagttgcactcgcaccagatctcgtactccgtcaagtgggggccggaga acgagcgctggaaggacgccagcagacgtacccagtttgttgggctcgctgcgatctccgacgtgcttgacctcggccagatctgcgccaacttggacatccacttg agcacgcccaagttcaacgtgtcgctcgaccacacgtgtacttcaccaggcgcggacttgacgtgaccaggtcggtcgacgtgacgaaccagatccgcatgacgaagctc gcgcacggccgcgcgttgatagagggcgagatgtactcggaccagggcaggcacatcgccagcatagtgcaagagaggttgtacatagctgagtcgtaa	88
PEX11 (*Candida viswanathii* ATCC20336)	Atggtcgccgattctttagtctaccacccaaccgtctccaaattagtcaagttcttggacacaacccaaagagggaaaaggttcttcagattattgtcctacttgtccag attcttgggctactacgcctacagaaaagggctactccaaggaaaccatcgccctttcgccaacttgaaaggaaaacttcacattcatcagaaaggcccatgagattctt gaagccaataaatcacttgcaattggcctccaaggcatacgacacaacaagttgttggacccagtcttcgcagatcaccaccatcatcagaaaacttggcctacgccggc tacttgaccatcgacggtgtcatattcttcaagttgttgggtctcattgacgccaagaagttccctaacttggctacatacgcctccagattctgttgatcgggttgattgc cggtttgatcaactccttgagaatcatctactcctgaaggactacgagcaccaggagggcgacaaggagaaggagaccgacgctaaggctatccacactaagtt gtacgccgctaagagaaaaattggtctgggacttgttggatactttattgctttgaactccttggacatcttgcatttcaccgagggtgacgtcgggttcgctggtactatca cctcccttcttgggattggaagacttgtggaaggccacttaa	89
CPRB (*Candida viswanathii* ATCC20336)	atggctttagacaagttagatttgtatgtcatcataacattggtggtcgctgtggccgcctatttgctaagaaccagttccttgatcagcccccaggacaccgggttcctca acacggacagcggaagcaactccagagacgtcttgctgacattgaagaagaatataaaaacacgttgttgtttgttggtcccagaccggtacggcagaagatt acgccaacaaattgtcaagagaattgcactccagatttggcttgaaaaccatggttgcagatttcgctgattacgattgggataacttcggagatatcaccgaagatat cttggtgttttcatcgttgccacctacggtgagggtgaacctaccgacaatgccgacgagttccacacctggttgactgaagaagctgcacacttgagtactttgagat ataccgttcgggttgggtaactccacctacgagttcttcaatgctattggtagaaagttgacagattgttgagtgagaaaggtggtgacagatttgctgaatatgctg aaggtgacgacggcactggcacctggacgaagatttcatggcctggaaggataatgtctttgacgccttgaagaatgacttgaacttgaagaaaaggaattgaa gtacgaaccaaacgtgaaattgactgagagagatgactgtctgctgccgactcccaagtttccttgggtgagccaaacaagaagttacatcaactccgagggcatc gacttgaccaagggtccattcgaccacacccacctactgccaggatcaccgagaccagagagttgttcagctccaaggaaagacactgtattcacgttgaat ttgacatttctgaatcgaacttgaaatacaccaccggtgaccatctagccatccgaccgaaaacatcaagcaatttgccaagtgtttcggattgga agataaactcgacactgttattgaatggcattggactccacttacaccattccaactccaattacttacggtgctgtcattagacaccatttagaaatctcc ggtccagtctgagacaattcttttgtcgattgctgggtttgctcctgatgaagaaacaaagaagacttcaccagactggtggtgacaaacaagaattcgccacca aggttaccgcagaaagttcaacattgccgatgccttgttatattcctccaacaacactccatggtccgatgttccttttgagttccttattgaaacatccaacacttgact ccacgttactactccatttcttctcgtcgttgagtgaaaaacaactcatcaatgttactgcagtcgttgaggccgaagaagaagccgatggcagaccagtcactggtg ttgttaccaacttgttgaagaacattgaaattgccgcaaaacaagactggcgaaaagccacttgttcactacgattgagcggcccaagaggcaagttcaacaagttc aagttgccagtgcacgtgagaagatccaacttaagttgccaaagaactccaccaccccagttatcttgattggtccaggtactggtgttgcccattgagaggtttcgt tagagaaagagttcaacaagtcaagaatggtgtcaatgttggcaagacttttgttgtttatggttgcagaaactccaacgaggacttttgtacaagcaagaatgggcc	90

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1719 of 3077

Table 19

Gene (Organism)	Nucleic acid sequence of the gene protein-encoding sequence (from START codon to STOP codon except where otherwise noted)	SEQ ID NO
	gagtacgcttctgtttttgggtgaaaacttgagatgttcaatgccttctagacaagacccatccaagaaggtttacgtccaggataagatttagaaaacagccaact tgtgcacgaattgttgaccgaaggtgccattatctacgtctgtggtgacgccagtagaatggccagagacgtccagaccacgatctccaagattgttgccaaaagca gagaaatcagtgaagacaaggccgctgaattggtcaagtcctggaaagtccaaaatagataccaagaagatgtttggtag	
FAA1 (*Candida viswanathii* ATCC20336)	ATGGGTGCCCCTTTAACAGTCGCCGTTGGCGAAGCAAAACCAGGCGAAACCGCTCCAAGAAGAAAGCCGCTC AAAAAATGGCCTCTGTCGAACGCCCAACAGACTCAAAGGCAACCACTTTGCCAGACTTCATTGAAGAGTGTTTT GCCAGAAACGGCACCAGAGATGCCATGGCCTGGAGAGACTTGGTCGAAATCCACGTCGAAACCAAACAGGTTA CCAAAATCATTGACGGCGAACAGAAAAAGGTCGATAAGGACTGGATCTACTACGAAATGGGTCCTTACAACTAC ATATCCTACCCCAAGTTGTTGACGTTGGTCAAGAACTACTCCAAGGGTTTGTTGGAGTTGGGCTTGGCCCCAGA TCAAGAATCCAAGTTGATGATCTTTGCCAGTACCTCCCACAAGTGGATGCAGACCTTCTTAGCCTCCAGTTTCCA AGGTATCCCCGTTGTCACCGCCTACGACACCTTGGGTGAGTCGGGCTTGACCCACTCCTTGGTGCAAACCGAA TCCGATGCCGTGTTCACCGACAACCAATTGTTGTCCTCCTTGATTCGTCCTTTGGAGAAGGCCACCTCCGTCAA GTATGTCATCCACGGGGAAAAGATTGACCCTAACGACAAGAGACAGGGCGGCAAAATCTACCAGGATGCGGAA AAGGCCAAGGAGAAGATTTTACAAATTAGACCAGATATTAAATTTATTTCTTTCGACGAGGTTGTTGCATTGGGT GAACAATCGTCCAAAGAATTGCATTTCCCAAAACCAGAAGACCCAATCTGTATCATGTACACCTCGGGTTCCAC CGGTGCTCCAAAGGGTGTGGTTATCACCAATGCCAACATTGTTGCCGCCGTGGGTGGTATCTCCACCAATGCT ACTAGAGACTTGGTTAGAACTGTCGACAGAGTGATTGCATTTTTGCCATTGGCCCACATTTTCGAGTTGGCCTTT GAGTTGGTTACCTTCTGGTGGGGGGCTCCATTGGGTTACGCCAATGTCAAGACTTTGACCGAAGCCTCCTGCA GAAACTGTCAGCCAGACTTGATTGAATTCAAACCAACCATCATGGTTGGTGTTGCTGCCGTTTGGGAATCGGTC AGAAAGGGTGTCTTGTCTAAATTGAAACAGGCTTCTCCAATCCAACAAAAGATCTTCTGGGCTGCATTCAATGC CAAGTCTACTTTGAACCGTTATGGCTTGCCAGGCGGTGGGTTGTTTGACGCTGTCTTCAAGAAGGTTAAAGCCG CCACTGGTGGCCAATTGCGTTATGTGTTGAATGGTGGGTCCCCAATCTCTGTTGATGCCCAAGTGTTTATCTCC ACCTTGCTTGCGCCAATGTTGTTGGGTTACGGTTTGACTGAAACCTGTGCCAATACCACCATTGTCGAACACAC GCGCTTCCAGATTGGTACTTTGGGTACCTTGGTTGGATCTGTCACTGCCAAGTTGGTTGATGTTGCTGATGCTG GATACTACGCCAAGAACAACCAGGGTGAAATCTGGTTGAAAGGCGGTCCAGTTGTCAAGGAATACTACAAGAA CGAAGAAGAAACCAAGGCTGCATTCACCGAAGATGGCTGGTTCAAGACTGGTGATATTGGTGAATGGACCGCC GACGGTGGTTTGAACATCATTGACCGTAAGAAGAACTTGGTCAAGACTTTGAATGGTGAATACATTGCTTTGGA GAAATTGGAAAGTATTTACAGATCCAACCACTTGATTTTGAACTTGTGTGTTTACGCTGACCAAACCAAGGTCAA GCCAATTGCTATTGTCTTGCCAATTGAAGCCAACTTGAAGTCTATGTTGAAGGACGAAAAGATTATCCCAGATGC TGATTCACAAGAATTGAGCAGCTTGGTTCACAACAAGAAGGTTGCCCAAGCTGTCTTGAGACACTTGCTCCAAA CCGGTAAACAACAAGGTTTGAAAGGTATTGAATTGTTGCAGAAATGTTGTCTTGTTGGATGACGAGTGGACCCCA CAGAATGGTTTTGTTACTTCTGCCCAAAAGTTGCAGAGAAAGAAGATTTTAGAAAGTTGTAAAAAAGAAGTTGAA GAGGCATACAAGTCGTCTTAG	91

Table 19

Gene (Organism)	Nucleic acid sequence of the gene protein-encoding sequence (from START codon to STOP codon except where otherwise noted)	SEQ ID NO
POX5 (*Candida viswanathii* ATCC20336)	atgcctaccgaacttcaaaaagaaagagaactcaccaagttcaacccaaaggagttgaactacttcttggaaggttcccaagaaagatccgagatcatcagcaa catggtcgaacaaatgcaaaaagacccatcttgaaggtcgacgcttcatactacaacttgaccaaagaccaacaaagagaagtcaccgccaagaaagattgcc agactctccagatacttgagcacgagtaccgagaccaacaggcccagagattgtcgatcctcggtgtctttgacccacaagtcttcaccagaatcggtgtcaacttg ggtttgtttgtttcctgtgtccgtggtaacggtaccaactcccagttcttctactggaccataaataagggtatcgacaagttgagaggtatctatggttgttttggtatgactg agttggcccacggttccaacgtccaaggtattgaaacaccacccgccacttttgacgaagacactgacgagttttgtcatcaacacccccacacattggtgccaccaagtg gtggatcggtggtgctgcgcactccgccacccactgctccgtctacgccagattgaaggtcaaagaaaaggactacggtgtcaagacctttgttgtcccattgagag actccaaccacgacctcgagccaggtgtgactgttggtgacattggtgccaagatgggtagagacggtatcgataacggttggatccagttctccaacgtcagaatc ccaagattctttatgttgcaaaagtactgtaaggtttccgctctgggtgaagtcaccatgccaccatctgaacaattgtcttactcggctttgattggtggtagagtcaccat gatgatggactcctacagaatgaccagtagattcatcaccattgccttgagatacgccatccacagaagaaggacaattcaagaagaaggacaccgatacattgaaac caagttgattgactacccattgcatcaaaagagattgttcccattcttggctgccgctacttgttctcccaaggtgccttgtactacttagaacaaacccatgaacgcaaccaa cgacaagttggacgaagctgtcagtgctggtgaaaaggaagccattgacgctgccattgtcgaatccaagaaattgttcgtcgcttccggttgtttgaagtccacctgt acctggttgactgctgaagccattgacgaagctgtcaagcttgtggtgctcaggttactcgtctacaacggtttcggtaaagcctactccgactgggttgtccaatgt acctgggaaggtgacaacaacatcttggccatgaacgttgccaagccaatggttgagaggactgagacttgttgaaggagccagaacaaaagggattggttctctccagcgttg ccgacttggacgacccagccaagttggttaaggctttcgaccacgcccttccggctggccagagacattggtgctgttgctgaagacaaggtttcgacattaccg gtccaagtttggtttggttccaagttgaacgctcacagattcttgattgacggtttcttcaaggcgtatcacccccagaatggtctgaagtcttgagaccttgggtttcttgtat gccgactggatcttgaccaacttggtgccacctcttcagctacggtatcattacccccagatgtctgcagcagaaagatttcctccgagcacttcccagccttgtgtgccaa ggttagaccaaacgttgttggtttgactgatggtttcaacttgactgacatgatgaccaatgctgctattggtagatatgatggtaacgtctacgaacactacttcgaaact gtcaaggctttgaacccaccagaaaacaccaaggctccatactccaaggcttttggaagacatgttgaacgtcagaccttgaagtcagagagaaaagaggtgaaaa gtccgaagaagctgctgaaatcttgtccagttaa	92
POX4 (*Candida viswanathii* ATCC20336)	atgactttacaaagaaaaaacgttagtgtatcacaaggtcctgacctagatcatccatccaaaaggaaagagacagctccaaatggaaccctcaacaaatgaac tacttcttggaaggctccgtcgaaagaagttgagttgatgaaggcttttggcccaacaaatggaaagagacccaatcttgttcacagacggctcctactacgacttgac caaggaccaacaaagagaattgaccgccgtcaagatcaacagaatccgccagatacagagagaacaagaatccatcgacacttcaacaagagattgtccttgattg gtatctttgacccacaggtcggtaccagaaattggtgtcaacctcggtttgttccttcttgtatcagaggtaaccgttcaacgttgctggtttggaaaccaccgccacatttgacaaggaat aggaaaccgctgacgttaaaggtatctacggttgtttcggtatgaccgaattggcccacggttccaacgttgctggtttggaaaccaccgccacatttgacaaggaat ctgacgagtttgtcatcaacaccccacacattggtgccaccaagtggtggattggtggtgctgtctcactccgccacccactgttctgtctacgccagattgattgttgac ggtcaagattacggtgtcaagacttttgttgtcccattgagagactccaaccacgacctcatgccaggtgtcactgttggtgacattggtgccaagatgggtagagatg gtatcgataacggttggatccaattctccaacgtcagaatccaaggattcttatgttgcaaaagttctgtaaggtttctgctgaaggtgaagtcaccttgccaccttttgga acaattgtcttactccgccttgttgggtggtagagtcatgatggttttggactcctacagaaatgttggctagaatgtccaccattgccttgagatacgccattggtagaaga caattcaagggtgacaatgtcgatccaaaagatccaaacgcttggaaacccaattgatagattacccattgcaccaaaagagattgttcccatacttggctgctgccgcc tacgtcatctccgctggtgccctcaaggttgaagacaccatccataacacctggctgaattggacgctgccgttgaaagaacgacaccaaggctatctttaagtct attgacgacatgaagtcattgtttgttgactctggttccttgaagtccactgccactggttgggtgctgaagccattgaccaatgtagacaagcctgtggttggtcacggtt actcgtcctacaacggcttcggtaaagcctacaacgattgggttgtccaatgtacttgggaaggtgacaacaatgtcttggccatgagtgttggtaagccaattgtcaa gcaagttatcagcattgaagatgccggcaagaccgtcagaggttccaccgcttcttgaacccaattgaaggactacactggttccaacagctccaaggttgttttgaa cactgttgctgacttggacgacatcaagactgtcatcaaggctattgaagttgccatcatcagattgtcccaagaagctgcttctattgtcaagaaggaatctttcgacta	93

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1721 of 3077

Table 19

Gene (Organism)	Nucleic acid sequence of the gene protein-encoding sequence (from START codon to STOP codon except where otherwise noted)	SEQ ID NO
	tgtcggcgctgaattggttcaactctccaagttgaaggctcaccactacttgttgactgaatacatcagaagaattgacaccttgaccaaaaggacttggttccatactt gatcacctcggtaagttgtacgctgccactattgtcttggacagatttgccggtgtcttcttgacttcaagttgcctccaccggaagccatcactgctttggcctctgtgca aattccaaagttgtgtgctgaagtcagaccaaacgttgttgcttacaccgactcctccaacaatcgacgacatgattgtcaattctgctattggtagatacgatggtgacat ctatgagaactactttgactggtcaagttgcagaaccccaccatccaagaccaaggctccttactctgatgctttggaagccatgttgaacagaccaacctttggacga aagagaaagatttgaaaagtctgatgaaaccgctgtatcttgtccaagtaa	
PXA1 (*Candida viswanathii* ATCC20336)	atggtcaacatatcgaaattgacgggttataacaagcaggacatcaggaatgtggtgctattgctacaggagtttgtcaagacctacaaagacaacaagatcaaaac tcaactacctgagtagacctgtcatcttgttcttgagtacctgccggtattgggtgttttcaccttgagaagcatcgtcactaagtacaacgagtaccta ctcaacaagagattgagacgcccaagtcttatcagacaatcctccaatatcttgaagaacggatccgtgagatcttatccagaagggcaacggcaaagtaaca agaatcatcatcccaaaagcaaacaacgaccagtatgccgccgacaagtatttgtataaagatttgcccgcaacgagcaaatattgcaacagcaaaagggaag gctcttcaattccagattcttgaaccagttgaccattatctggaagatcttgatttccaaagttctactgccaaaacacttccttgttgttatcgcagtgcttcttttgatttcag aacatggttgtccttgttgattgccaagctagatggtcagaattgttgatttgctgcagacggtaggaagtttgcccgtgacttgatttacttttttgttgattgccttc cctgcttcgtacaccaacgccgctatcaaatacttggaattggcgttaggattcagaactaatcttaccagatacatccatgacatgtacttggacaaaacca tgtcgtactacaaagtgggattgaacggcgccgatatccaaaacatagaccagtacatcaccgaagatgtcaccaaattctgtatgtcgttgttcgttgtttcctcca tgggtaagccattcattgactgatcttttcagtgtttatttgagagacaatttgggtactggtgccattattggcatttttgctaccgccatcatgttgaaaa aggcaacaccaagattcggtaagttggctgccaaaagaacccacttggaaggtgtttattttcaaccaacagttgaacataatgaccaacagtgaagaggattgggtt ctacaaaggatcgaagattgagaagtccaagcttgcggagaacttgacaagttgatgggtcacgtatcgagagaaatcaatttatcgtccagctatgccgctctag aagactacgtgcttaaatacacgtggctggcctgggttacatctttctggtctacctgtgttttggatgtgctttcctaaagaagacccaagtagtggccatattgctg atatagatgacgatgaccatgcccatggacacgggcacaccgggaagagaccagctcaacaactgaaaacatgaagaccttcgtcaccaacaagcgattatt gttgagtcttgccgatgctggttccagattgatggttagtttgaaagaggtcaccacgttgacaggtataacgaatagagtcttcaacatgttgactcagctccaccgtgt ccatgatcctaaatttgactacggtgacaagtatggtttgcctgatattcacgggacttatcaattgaactacgatggtttgagattggaacatgttccaattactgtgcca actgccgaggggttctactccacaccattgatcccagacctcactttgacatcaaggcaagaatttgttattgttggtccaaacggttcgggcaaaacttctgttgcca gggttcttgcaggtctttggccctgatgccgggttagtgctgaaaccactggatttgttcttaacccacaaaagagttattcaccacccggaagtttgcgtgaccaagtt gtttacctaatagatccgaaaacaccaccaacgatcaaatttccacatcttcactgtgtacacttagaccatattgttaaacggtacggattgaaccagaacttgg atttcgctaaaacattgagtggaggtgagaagcaaagattgagttcgccagagtgttgtttaacagaccaagtattgtcattcttgatgattcgacgtcggcgttgtccc cagatatggaagagttgatgtaccaggtgttgcaagatcacaagatcaattacgtcacacttcaaatcgtccctctctcagtaagttccatgataaagtatttgaaatat aa	94

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1722 of 3077

Table 19

Gene (Organism)	Nucleic acid sequence of the gene protein-encoding sequence (from START codon to STOP codon except where otherwise noted)	SEQ ID NO
PXA2 (*Candida viswanathii* ATCC20336)	atgacagtggagaatgcaaaactacagaagaactcgttggcggttctgctcttgaaggtgtacaaatccaacagatcattattgttaaacacctcatacatcattaa tcattgctgccttcactggcgcaacgaatacccgggcgaggcacctcctccagatcatcggcaaaagtagagaccgatgaagaacaatcggttaaaaagaaaca ccccaagctctctagaggagtccttccatagactaagaaaaagcaatcttgccaacttcttttgataggaactatagtttacttttttcgccaacttgacttgttggtggtcagag cattattgacacttagagttgctaccttgacggtcagcttgtgtgggggcattggtttcaagaagaataaggtgtttgccaagtacttgttgtactggatgcttcttggtatcc ccgctgctttgacaaatgccttgttgaactggaccaaactgagcaagagcattagaatgaactgaatataacatcatggaggaatacttgccagataact tggaccccaaactattattcattgatccatttgactgataacaagattagagacccaaatcagaggaataaccactgatactagtcgtttgagcgatgccttggcaagctt gcccggtcacatattgaagccaacgttggatatcatattgtgtgcgcaacagttaagcaagagcggtgttggtaatgggggaaggtacgttggcattaggtatattggc acacttctcaaccatgatcatccgtttcttctcccgccatttgccaagttggcggctgagagagctaaccttgaaggtcagttgccgttccgcgcattccaagattgttgcc aacagtgaagaaattgctttcttgggtggtcatgaccgtgagttggatcacatcgacaccactgctactatacttttggagagattctcgaaaggcgaatattggaagcgag ccatacacgaaatcacacacaaacgtttattgtgaagtactttgggtttagtgttgttgttcaccggtttttcattgccaaatactttggtgagccggaagat aagaatgttgctggtaatttcatcaccaacagaagattgttgatgagtgcctcggattcctggatcgtttaatctatctagaagatacttgttgcaagttgtcggtcatgct accagagtgtctgacttcttggacacttacatgaagtggaggaagaagaagaagaagaatcacatcgaatgtgcagtttaacaacgacgagattacttttcgatcatgtt agattgatgactccaacggaagtgaccttgatcccagactgaacttttccattaaaccaggtgaccatttgttgattgtgggggccaaacggttcaggtaagtcgtcgtt gttcagaatgttgggtggttgtggcccgttaggtttggtactattagaattccaaacacagagaaacatgttctactgccgcaaaaggcttacctgttgaaggatcattc agaggcaaatcatttatccacacaacgtgactcaacagaagaagactgatcaacaattgaaagagatctgaaggtttgaaattggaagattactcagggcaat tggatgaggttaagaaatggaggcgaagaattgtccattggtgctcaacaaagattggctatggctagattgtactaccacgaacctaagtttgctgtcttggacgaatg tacttcagctgtgtcaccagacatggaacaactcatgtaccaacacgcacaaggtttgggtatcacgcttttgtccgttgcccatagacctgcattgtggcacttccaca aatactgttggaattcgacgggaaggtagttactacttggtacgttggatgaaaagcacaaaatgaagttagaagaagaagaacgactcaagaaggagaatg aaaagaagagtgtcgccaagaagtag	95
ACL1 (*Yarrowia lipolytica* CLIB122) codon optimized for *Candida* strain ATCC20336	atgtccgccaacgaaaacatctccagattcgacgccctgtcggtaaggaacacccagcctacgaattattccataaccacacaagatccttcgtttatggattaca acctagagcctgtcaaggaatgttggacttcgacttcatctgcaagagagaaaacccatctgtcgccggagttatctatccattcggtggtcaattcgttaccaagatgt actgggtaccaaggaaactctttaccgtttaccaacaagttgaaaaggccgctgccaagcacccagaagttgatgttgtcgttaacttggcctctccagatctgttt actcttccaccatggaattgttagaatacccacaattcagaaccatcgccattattgccgaaggagttccagaaagaagggccagagaaatcttacacaaggccc aaaagaagggagtcaccatcattggaccagctaccgttggcggaatcaagccaggatgtttcaaggtgggcaacaccggcgcgaatgatggacaacattgttgcc tctaagttatacagaccaggttctgtggcctacgtttctaagtctggcggcatgtctaacgaattgaacaacattatctcccacaccaccgacgggagtttacgaaggaa ttgctattggaggagacagatacccgtgaactaccttcattgaccatatcttgagatacgaagccgacccaaagtgcaagatcatcgttttacttggagaagtgggag gagtggaagaatacagagttatcgaagctgtgaagaacggtcaaatcaagaagccaatcgttgcttgggccattggaacttgcgcctctatgttcaagactgaagtg caattcggtcacgccggttctatggccaactctgacttggaaactgccaaggctaagaacgccgccatgaagtccgctggtttctacgttccagatacccttcgaagac atgccagaagttcttgccgaattatacgaaaagatggttgccaagggtgaattgtccagaatctccgaacctgaagttccaaagatcccaattgactactcctgggcc caagaacttggacttatcagaaagccagctgctttcatctctactatttcgatgacagaggtcaagaactttgtacgctggtatgccaatttctgaagtgttcaaggaa gacattggaatcggtggagttatgtccttgttgtggttcagaagaagattaccagactacgcctctaagtttcttgaaatggtgttaatgcttactgctgaccacggacca gccgtatctggagccatgaacaccattatcaccaccagagctggaaaggatttaatttcctctttggtggctggattattgaccattggaaccagattcggcggagctct tgacggagctgccaccgaattcaccactgcctacgacaaggattgtctccaagacaattcgtggataccatgagaaagcaaaacaagttgattcctggaatttggt catagagttaagtccagaaacaacccagatttcagagttgaacttgttaaggacttgtgaagaagaacttcccatctacccaattgttagactacgccctgctgttga	96

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1723 of 3077

Table 19

Gene (Organism)	Nucleic acid sequence of the gene protein-encoding sequence (from START codon to STOP codon except where otherwise noted)	SEQ ID NO
	agaagttaccacctctaagaaggacaacttgatttgaacgtggacggagctattgctgtgtccttgttgatttaatgagatcctgtggagccttactgtcgaagaaact gaagactacttaaagaacggagtgttaaacggattgttcgtgttaggaagatctattggattaattgcccaccatttagatcaaaagagattaaagaccggattgtaca gacatccttgggacgatatcacctactggtgggtcaagaagctatccaaaagaagagagttgaaatcagtgccggtgacgtgtctaaggccaagactagatcata g	
ACL2 (*Yarrowia lipolytica* CLIB122) codon optimized for *Candida* strain ATCC20336	atgtcagcgaaatctattcacgaagccgacggtaaggccttgttagcacacttttgtctaaggcgccagtctgggccgaacaacaaccaatcaacacgtttgagat gggtacaccaaagttggcgtccttgacgttcgaagacggtgtcgcccagaacaaatcttcgccgccgctgagaagacctacccatggttgttggaatctggtgcca agtttgtcgccaagccagaccaattatcaagagaagaaggtaaggccggtttgttggtattaaacaagtcgtgggaagaatgtaagccatggatgccgaacgtgc cgccaagccaatcaacgtcgaaggtattgacggcgtcttgagaacgttcttggttgaaccatttgtcccacagaccaaaagcacgaatactacatcaacatccact ctgtcagagaaggtgactggatctattctaccacgaaggcggcgttgacgttggtgacgtcgacgccaaggccgccaagatcttaatcccagtggacattgaaaa cgaataccatctaacgccacgttaaccaaggaattgttggcacacgtcccagaagaccacaacaccaaacttgttagacttcatcaaccgtttatacgccgtttacgt tgatttgcaatttacgtatttggaaatcaacccattggttgtcatcccaaccgcccaaggtgttgaagttcactactacttggatcttgccggtaagttagaccaaaccaccgcaga atttgaatgtggtccaaagtgggctgctgcgcgttctccagccgcttgggtcaagttgttaccattgacgccggttctaccaaggtctctatcgacgccggtccagccat ggttttcccagctcctttcggaagagaattgtctaaggaagaagcgtacattgcggaattagattctaagaccggcgcttccttgaagttgactgtgttaaatgccaagg gtagaatctggaccctgctgctggaggcggccgcctctgttgtttacgccgacgccattgcgtccgccggttttgctgacgaattagccaactacggtgaatactccggt gctccaaacgaaacccaaacctacgaatacgccaaaaccgtattggatttaatgaccgtggtgacgctcacccagaaggtaaggtaattgttcattggtggccggca tcgccaacttcacccaagtgggctctaccttcaagggtatcatcgtgccttccgtgactaccaatcctccttgcacaaccacaaggtcaagatttacgtcagaaagag gtggaccaaactggcaagaaggattgcgtctcatcaagtcggctggtgacgaattgaatttgccaatggaaatttacggtcagacatgcacgtctcgggaattgtg cctctcgcttgctggcaagcgtccaaagaatgttaagccttttggtaccggcccttccactgaagcttctactccttaggcgtgtaa	97

Table 19

Gene (Organism)	Nucleic acid sequence of the gene protein-encoding sequence (from START codon to STOP codon except where otherwise noted)	SEQ ID NO
FAT1 (*Candida viswanathii* ATCC20336)	ATGTCAGGATTAGAAATAGCCGCTGCTGCCATCCTTGGTAGTCAGTTATTGGAAGCCAAATATTTAATTGCCGA CGACGTGCTGTTAGCCAAGACAGTCGCTGTCAATGCCCTCCCATACTTGTGGAAAGCCAGCAGAGGTAAGGCA TCATACTGGTACTTTTTCGAGCAGTCCGTGTTCAAGAACCCAAACAACAAGCGTTGGCGTTCCCAAGACCAAG AAAGAATGCCCCCACCCCCAAGACCGACGCCGAGGGATTCCAGATCTACGACGATCAGTTTGACCTAGAAGAA TACACCTACAAGGAATTGTACGACATGGTTTTGAAGTACTCATACATCTTGAAGAACGAGTACGGCGTCACTGC CAACGACACCATCGGTGTTTCTTGTATGAACAAGCCGCTTTTCATTGTCTTGTGGTTGGCATTGTGGAACATTGG TGCCTTGCCTGCGTTCTTGAACTTCAACACCAAGGACAAGCCATTGATCCACTGTCTTAAGATTGTCAACGCTTC GCAAGTTTTCGTTGACCCGGACTGTGATTCCCCAATCAGAGATACCGAGGCTCAGATCAGAGAGGAATTGCCA CATGTGCAAATAAACTACATTGACGAGTTTGCCTTGTTTGACAGATTGAGACTCAAGTCGACTCCAAAACACAGA GCCGAGGACAAGACCAGAAGACCAACCGATACTGACTCCTCCGCTTGTGCATTGATTTACACCTCGGGTACCA CCGGTTTGCCAAAAGCCGGTATCATGTCCTGGAGAAAAGCCTTCATGGCCTCGGTTTTCTTTGGCCACATCATG AAGATTGACTCGAAATCGAACGTCTTGACCGCCATGCCCTTGTACCACTCCACCGCGGCCATGTTGGGGTTGT GTCCTACTTTGATTGTCGGTGGCTGTGTCTCCGTGTCCCAGAAATTCTCCGCTACTTCGTTCTGGACCCAGGCC AGATTATGTGGTGCCACCCACGTGCAATACGTCGGTGAGGTCTGTCGTTACTTGTTGAACTCCAAGCCTCATCC AGACCAAGACAGACACAATGTCAGAATTGCCTACGGTAACGGGTTGCGTCCAGATATATGGTCTGAGTTCAAGC GCAGATTCCACATTGAAGGTATCGGTGAGTTCTACGCCGCCACCGAGTCCCCTATCGCCACCACCAACTTGCA GTACGGTGAGTACGGTGTCGGCGCCTGTCGTAAGTACGGGTCCCTCATCAGCTTGTTATTGTCTACCCAGCAG AAATTGGCCAAGATGGACCCAGAAGACGAGAGTGAAATCTACAAGGACCCCAAGACCGGGTTCTGTACCGAGG CCGCTTACAACGAGCCAGGTGAGTTGTTGATGAGAATCTTGAACCCTAACGACGTGCAGAAATCCTTCCAGGG TTATTATGGTAACAAGTCCGCCACCAACAGCAAAATCCTCACCAATGTTTTCAAAAAAGGTGACGCGTGGTACA GATCCGGTGACTTGTTGAAGATGGACGAGGACAAATTGTTGTACTTTGTCGACAGATTAGGTGACACTTTCCGT TGGAAGTCCGAAAACGTCTCCGCCACCGAGGTCGAGAACGAATTGATGGGCTCCAAGGCCTTGAAGCAGTCC GTCGTTGTCGGTGTCAAGGTGCCAAACCACGAAGGTAGAGCCTGTTTTGCCGTCTGTGAAGCCAAGGACGAGT TGAGCCATGAAGAAATCTTGAAATTGATTCACTCTCACGTGACCAAGTCTTTGCCTGTGTATGCTCAACCTGCGT TCATCAAGATTGGCACCATTGAGGCTTCGCACAACCACACAAGGTTCCTAAGAACCAATTCAAGAACCAAAAGTTG CCAAAGGGTGAAGACGGCAAGGATTTGATCTACTGGTTGAATGGCGACAAGTACCAGGAGTTGACTGAAGACG ATTGGTCTTTGATTTGTACCGGTAAAGCCAAATTGTAG	98

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1725 of 3077

Table 19

Gene (Organism)	Nucleic acid sequence of the gene protein-encoding sequence (from START codon to STOP codon except where otherwise noted)	SEQ ID NO
CYP52A17 (*Candida viswanathii* ATCC20336) FROM pAA1712	atgattgaacaactcctagaatattggtatgtcgttgtgccagtgttgtacatcatcaaacaacaactccttgcatacacaaagactcgcgtcttgatgaaaaagttgggtgct gctccagtcacaaacaagttgtacgacaacgctttcggtatcgtcaatggatggaaggctctccagttcaagaaagagggcagggctcaagagtacaacgattac aagtttgaccactccaagaacccaagcgtgggcacctacgtcagtattcttttcggcaccaggatcgtcgtgaccaaagatccagagaaatatcaaagctatttggc aacccagtttggtgatttttcttgggcaagaggcacactcttttaagccttgttaggtgatgggatcttcacattggacgggcgaaggctggaagcacagcagagcca tgttgagaccacagtttgccagagaacaagttgctcatgtgacgtcgttggaaccacactccagttgttgaagaagcatattcttaagcacaagggtgaatacttttgat atccaggaattgttcttagatttaccgttgattcggccacggagttcttatttggtgagtccgtgcactccttaaaggacgaatctattggtatcaaccaagacgatatag attttgctggtagaaaggacttgctgagtcgttcaacaaagcccaggaatactggctattagaacctggttgcagacgttctactggttggtcaacaacaaggagttt agagactgtaccaagctggtgcacaagttcaccaactactatgttcagaaaagctttggatgctagcccagaagagcttgaaaagcaaagtgggtatgttgttcttgtac gagcttgtcaagcagacaagagacccaatgttgttgcgtgaccagtctttgaacatcttgttggccggaagagacaccactgctgggttgttgtcgtttgctgtcttgag ttggccagacacccagagatctgggccaagttgagaggaggaaattgaacaacagtttggtctggagaaagactctcgtgttgaagagattaccttgagagcttgaa gagatgtgagtacttgaaagcgttccttaatgaaacctgcgtatttacccaagtgtcccaagaaacttcagaatcgccaccaagaacacgacattgccaagggggc ggtggttcagacggtacctcgccaatcttgatccaaaagggagaagctgtgtgtatggtatcaactctactcatttggaccctgtctattacggccctgatgctgctga gttcagaccagagagatggtttgagccatcaaccaaaaagctcgggctgggcttacttgccattcaacggtggtccaagaatctgtttgggtcagcagtttgccttgacg gaagctggctatgtgttggttagattggtgcaagagttctcccacgttaggctggacccagacgaggtgtaccgccaaagaggttgaccaacttgaccatgtgtttgc aggatggtgctattgtcaagtttgactag	99

Table 19

Gene (Organism)	Nucleic acid sequence of the gene protein-encoding sequence (from START codon to STOP codon except where otherwise noted)	SEQ ID NO
ADH1-2 alcohol dehydrogenase (*Candida viswanathii* ATCC20336)	Atgtctgctaatatcccaaaaactcaaaaagctgtcgtcttcgagaagaacggtggtgaattaaatacaaagacatcccagtgccaacccaaaggccaacga attgctcatcaacgtcaagtactcgggtgtctgtcacactgatttgcacgcctggaagggtgactggccattggaccaaattgccattggttggtggtcacgaaggt gctggtgttgttgtcggcatggtgaaaacgtcaagggctggaaaatcggtgatttcgccggtatcaaatggttgaacggttcttgtatgtcctgtgagttctgtcagcaa ggtgctgaaccaaactgtggtgaagctgacttgtctggttgtcacacccacgatggttctttcgaacaatacgcccactgctgatgctgtgcaagccgccagaatcccagct ggcactgatttggccgaagttgcccccaatctttgtgctggtgtcaccgtctacaaagccttgaagactgccgacttggctgctggtcatatcctcggtgctgct ggtggtggtttgggctccttggctgtccaatacgccgtcgccatgggtttgaagatcgttgccattgacggtggtgacgtgaaaaggggtgacttgtcaagtcctgggtgct gaagcctacattgatttcctcaaggaaaagggcattgttgctgctgtcaagaaggccactgatggcggtccacacgggtgctatcaatgtttccgtttccgaaaaagcc attgaccaatctgtcgagtacgttagaccattgggtaaggttgtttggttgccagctggtctccaaggtcactgctggtgtttttcgaagccgttgtcaagtccattga aatcaaggtttcttacgtcggtaacagaaaaggatactgccgaagccgttgacttttttctccagaggcttgatcaagtgtccaatcaagattgtgggcttgagtgaattgc cacaggtcttcaagttgatggaagaaggtaagatcttggtagatacgtcttggatacctccaaa	100
FAO1 fatty alcohol oxidase (*Candida viswanathii* ATCC20336)	atggctccatttttgcccgaccaggtcgactacaaacacgtcgacaccttatgttattatgtgacgggatcatccacgaaaccaccgtcgaccaaatcaaagacgtt attgctcctgacttccctgctgacaagtacgaagagtacgtcaggacattcaccaaaccctccgaaaccccagggttcagggaaaccgtctacaacacagtcaac gcaaacaccacgggacgcaatccaccagttcattatcttgaccaatgttttggcatccaggtcttggctccagcttgaccaactcgttgacgcctatcaaggacatga gcttggaagaccgtgaaaaattgttggcctcgtggcgcgactcccccaatcgctgccaaaaggaagttgttcaggttgtttctacgcttacctttggtcacgttcacgag attggccaatgagttgcattgaaagccattcattatccaggaagaagaaccgtgaaaaggcttatgaaaacccaggagattgaccctttaagtaccagttttggaa aaaccgaagtttacggcgctgagttgtactccagatatatgatgatcattattggatcgtgccggtgctggttgtggcccacacttggccaacgatggcttca agagttggtttggaaaaggcaaatacttagcaactccgagttgaacttgatgacaaggacggcgttcaagaattataccaaagtggaggtacttgactacagt caaccaacagttgtttgttctgctggttccacttttggtggcggtaccactgtcaattggtcagcctgtcttaagacgccattcaaggtgcgtaaggaatggtatgatgag tttggtgttgacttgctgctgatgaagcatacgataaagcgcaggattatgtttggcagcaaatgggagcttctacgaaggcatcaccactctttggctaacgagat tattattgaaggtggtaagaaattaggttacaaggccaaggtattagaccaaaaacagcggtggtcatcctcagcacagatgcggtttctgttatttgggctgtaagcac ggtatcaagcagggttctgttaataactggtttagagacgcagctgcccacggttcccagttcatgcaacaggttagagtttttgcaaatactttaacaagaagggggatc gcttacggtatctgtgtgaggatgttgtaacggcggccaagttcaccattactggcccaaaaaagtttgttgtgctgcggtgctttgaacactccatctgtgttggtcaa ctccggcttcaagaacaagaacatcggtaagaacttaacttgcacccagtttctgtcgtgtttggtgatttttggcaaagacgttcaagcagaccacttccacaactcca tcatgactgcccttgttcagaagccgctgatttagacggcaagggccatggatgcagaattgaaaccatcttgaacgctccattcatccaggcttcattcttaccatgg agaggtagtaacgaggctagacgagacttgttggttacaacaacatggtggcgatgttgctccttagtcgtgacacccaccagtggttccgtttctgctcatccaacca aacctgaagctttggttgtcgagtacgacgtgaacaagtttgacagaaaactcgatcttgcaggcattgttggtcactgctgacttgttgtatatccaaggtgccaagaga atccttagtccacaggcatgggtgccaattttgaatccgacaagccaaaggataagatcaatcaaggacgaggactatgtcgaatggagagagccaaggttgcc aagattccttcgacacctacggctcacctatgattcggcacatcaaatgtcttcttgccgtatgtcaggtaagggtcctaaatacggtgctgttgacaccgatggtaga ttgtttgaatgttcgaatgtttatgttgccgatgcaagtctttgccaactgcaagcggtgccaaccctatggtccaccaccatgactctgccagacatgttgcgttaggtttg gcagactccttgaagaccaaagccaagttgtag	101

Table 19

Gene (Organism)	Nucleic acid sequence of the gene protein-encoding sequence (from START codon to STOP codon except where otherwise noted)	SEQ ID NO
FAS1 fatty acid synthase (*Candida viswanathii* ATCC20336)	ATGTCTACTCATAGACCTTTCCAATTGACCCACGGTTCCATCGAACACACCTTGTTGGTGCCAAACGAGTTGTTC TTCAACTATTCACAGTTAAAAGACGAATTCATAAAGACCTTGCCTGAACCAACCGAAGGTTTCGCTGGGCGACGA TGAACCTTCCAGTCCTGCTGAATTGTACGGCAAATTCCTCGGCTACATCAGTGACAACACCGTTCAATTCCCCC AGATCTTACAATTGTCCTTGCAAGACTTCCAGCAGCGATTCTTGGACAACCACGACAACATCCACTCCTTTGCC GTCAGATTATTAGAAGATGAAGCTTATCCAACAACAATCACCAAAGTCAAGGAAAATATCATCAAGAACTACTAC AAAGCCATCAAGTCCATCGACAAGGTCGAGTCAAACTTGTTGTACCACTGCAAACATGACGGCCAAGTTGGCCG CTATATTCGGTGGTCAAGGTAACACCGACGACTACTTTGAAGAATTGCGTGAATTGTACACCTTATACCAGGGC TTGATTGAGGACCTCCTTATCTCCATTGCCGACAAGTTGGACGAGTTATACCCTTCTTTTGACAAGATCTACACC CAGGGTTTGAACATCTTGGGCTGGTTGAAGCACCCAGAAACCACCCCTGACCAAGATTACTTGTTGTCCGTACC AGTGAGTTGTCCTGTTATCTGTATCATCCAATTGTGTCACTACACCATCACCTGCAAAGTTCTTGGTTTGACCCC TGGTGAATTTAGAGACTCGTTGAAGTGGTCCACCGGTCACTCCCAAGGTTTGGTTACTGCTACCGCTATTTCCA GTTCCGACTCCTGGGACTCCTTCAAACAAAACTCCATTGCTGCCGTCTCCTTGATGCTTTTCATTGGTGCCAGAT GTTTGATGGCTTACCCAAGAACTACCTTGCCACCAACCATGTTGCAAGACTCCTTGGAACACGGTGAAGGTAGA CCATCTCCAATGTTGTCAGTTAGAGACTTGACCATCACCCAAGTTGAGAAGTTTATTGAACAGACCAACTCTCAC TTGCCAAAGGAAAAGCACATTGCCGTCAGTTTGGTCAATGGTGCCAGAAATTTGGTTCTTTCTGGTCCCCCGGA GTCCCTTTACGGTTTCAACTTGAACTTGAGAAACCAAAAGGCTCCAATGGGATTGGACCAATCACGTGTTCCAT TCAGTGAACGTAAGTTGAAGTGTTCCAACAGATTCTTGCCAATTTTTGCACCATTCCACTCTCACTTGTTGGCTG ATGCCACTGAGTACATTTTGGATGATGTCAAAGAACACCGTTTGTCTTTCCAGAAATTGAAGATTGCAGTCTACG ATACCTACGACGGCTCCAACTTCCAAGAGAGCAAGGAACCAATTATTGACAGACTCGTCAAGTTGATCACCGAG TTGCCAGTTCACTGGGAAACCGCCACCAACCACAGGGCCACCCACATTTTGGATTTCGGCCCAGGTGGTGTCT CCGGTTTGGGTGTTTTGACCCACAGAAACAAGGAAGGTACTGGTGCTAGAATAATTGTTGCTGGTGCTCTTGAC TCCAACCCAATTGACGATGAGTATGGTTTCAAGCACGAAATCTTCCAGACTTCTGCCGACAAGTCCATCAAGTG GGCTAGCGATTGGTTGGAAGAATTCAAACCAACTTTGGTCAAGACTTCCCAGGGAAAGATCTACGTCAACACCA AGTTCTCGCAATTGTTGGGCAGAGCTCCCTTGATGGTCCCAGGTATGACACCAACCACTGTTAACCCAGACATC ATTGCCGCCTCTTTGAATGCTGGCTATCACATTGAATTAGCCGGTGGTGGTTATTTCGCCGGCAGGATCATGAC CAAAGCCATTGACCAAATTGTTGCCGACATCAAGCCAGGTTACGGTTTGGGTATCAACTTGATTTACGTCAACC CATTCATGTTGCAATGGGGTATTCCATTGATTAAGGAGTTGAGAGAAAAGGGTTATCCAATCCAATCTTTGACCA TTGGTGCTGGTGTTCCATCTTTGGAAGTTGCCACTGAATACATTGAAGAGTTGGGTCTCACGCACTTGGGCTTG AAACCAGGTTCGATTGACGCCATCAGCCAAGTCATCACCATTGCCAAGGCTCATCCAAAGTTCCCAATTGTCTT GCAATGGACTGGTGGTAGAGGTGGTGGCCACCACTCTTTTGAAGATTTCCACCAACCAATCCTCCAGATGTACT CCAAGATCAGAAGATGCCCAAACATTGTCTTGGTTGCTGGTTCCGGGTTTGGTTCTGACGAAGACACCTACCCA TACTTGACTGGTTCTTGGTCCAAGAGATTCAACTACCCACCAATGCCATACGATGGTGTCTTGTTTGGTTCCAGA GTCATGACCGCCAAGGAGGCCCACACTTCGTTGGAAGCTAAGAAATTGATTGCCTCGTGCCCAGGTGTCCCAG ATGAGAAGTGGGAGCAAACCTACAAGAAGCCAACCGGTGGTATCATCACTGTTAGATCTGAAATGGGTGAGCC AATCCACAAGATTGCCACCAGAGGTGTCATGTTCTGGAAGGAATTGGACGACACTATCTTCAACCTTCCAAAGC	102

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1728 of 3077

Table 19

Gene (Organism)	Nucleic acid sequence of the gene protein-encoding sequence (from START codon to STOP codon except where otherwise noted)	SEQ ID NO
	CAAAAGCCTTGGAAGCCATCAAGAAGAAGAGACTACATCATCAAGAAGTTGAACAGCGACTTCCAAAAGCCA	
	TGGTTTGGTAAGAATGCTTCTGGTGTTTGTGACTTGCAAGAAATGACCTACGAGGAAATCGCCAACAGATTGGT	
	TGAGTTGATGTACGTCAAGAAGTCTCAAAGATGGATTGACGTTTCCTTAAGAAACTTGTACGGTGACTTTTTGAG	
	AAGAGTTGAAGAAAGATTCACTTCTGCTGCCGGCGTGGTTTCCTTGTTGCAAAACTTCATCCAGTTGAACGACC	
	CAGAAACATTCAGTGCCGAGTTCTTCAACAAGTTCCCACAAGCCAAGGAACAATTGATTTCCGAAGAAGACGCT	
	GACCACTTTTTGTTGTTGGCTGCCAGACCAGGGCAAAAGCCGGTTCCATTCGTGCCAGTCTTGGACGAAAGATT	
	TGAATTCTTCTTCAAGAAAGATTCTCTTTGGCAATCTGAAGACTTGGAAAGTGTTGTCGACGAAGACGTTCAAAG	
	AACTTGTATCTTGCACGGTCCAGTTGCCTCCCAATTCACCAAGAAGGTTGATGAACCAATTGGCGAAATCTTGG	
	ACTCTATCCACGAGGGCCATATTGCCAAGTTGATCAAGGATGAATACGCTGGTGATGCATCCAAGATCCCAGTT	
	GTTGAGTACTTCGGTGGTTTCAAGACCGACAAGGTTAATGCTAACAATGTTCAAGTCAATGCTACCAGAAAGGA	
	AACCGTCTACGAAATTGGTTCCAAGTTGCCAGCCAGGCAAGACTGGTTGGACTTGTTGGCCGGTACTGAATTG	
	AACTGGTTGCACGCTTTCATCTCCACCAACAGAATTGTCCAAGGCTCCAAGCACGTCGCCAACCCATTGCACGA	
	CATTTTGGCTCCCGTTGCCAGATCCAGTGTTTCCATTGACAAGGCTACCAAGAAATTGACTGCTTATGAAAAGG	
	TCAAGGGTGAGTTGGTTCCAGTTGTTGAAATTGAATTGGTCAAGCCAAACACCATTCAATTGTCTTTGATTGAAC	
	ACAGAACTGCTGATGGCAAACCAGTTGCTTTGCCATTCTTGTACAAGTACGACCCAACTGATGGGTTTGCACCA	
	GTCTTGGAGATCATGGAAAACAGAAACGACAGAATCAAGGAATTCTACTGGAAGTTGTGGTTCGGTGCTTCCGT	
	CCCTTACGACAATGACATCGATGTCGAAGAGCAAATCTTGGGTGACGAAATCACCATTTCTTCTCAAGACATTG	
	GTGAATTCACACACGCTATTGGTAACAAGTGTGAAGCCTTTGTCAACAGACCAGGTAAGGTCACTTTGGCTCCA	
	ATGGATTTCGCCATTGTTGTTGGTTGGAAAGCTATCATCAAGTCCATCTTCCCAAAGACCGTTGACGGTGACTT	
	GTTGAAGTTGGTCCACTTGTCCAACGGTTACAAGATGATCCCTGGTGCAGCTCCATTGCAAAAGGGCGATGTTG	
	TTTCCACTAGATCTGACATCAAGGCTGTTTTGAACCAACCAAGTGGTAAGTTGGTTGAAGTTGTTGGTACCATCT	
	TCCGTGAAGGCAAGCCAGTTATGGAAGTCACTTCACAATTCTTGTACCGTGGTGAATACGACGACTACTGCAAC	
	ACCTTCCAAAAGGTCACTGAAACTCCAGTTCAAGTCTCATTCAAGTCTCCTAAGGATTTGGCTGTTTTGAGATCC	
	AAGGAATGGTTCCATTTGGAAAAGGATGTCGAGTTTGATGCTTTGACTTTCAGATGTGAATCCACTTACAAGTTC	
	AAGTCTGCCAACGTCTACTCGTCCATCAGAACGACCGGTCAAGTTTTCTTGGAGTTGTCCACCAAAGAAGTTAT	
	CCAAGTTGGTTCTGTTGACTATGAAGCTGGTACCTCTTATGGTAACCCAGTCACTGACTACTTGAACAGAAACG	
	GTAAGACCATTGAAGAGGCTGTTACTTTTGAGAATGCCATCCCCTTGTCGTCTGGTGAAGAGTTGACCACCAAG	
	GCTCCGGGTACCAACGAGCCATATGCTATTGTTTCTGGTGACTACAACCCAATCCACGTTTCCAGAGTCTTTTCT	
	GCTTACGCCAAGTTGCCAGGTACTATCACCCACGGTATGTACTCTTCTGCCGCCATCAGAGCCTTGGTTGAAGA	
	GTGGGCTGCCAACAACGTTGCCCCAAGAGTCAGAGCCTTCAAGTGTGAATTTGTTGGTATGGTTTTGCCAAACG	
	ACACTTTGCAAACCACTATGGAACACGTTGGTATGATCAATGGCCGTAAGATCATCAAGGTCAAGACTGTCAAT	
	GCCGAGACCGAGACTCCAGTCTTGCTTGCTGAAGCCGAAATTGAACAACCAACCACCACCTATGTTTTCACTGG	
	TCAAGGTTCCCAAGAACAAGGCATGGGTATGGATTTGTACAACTCTTCTGAAGTTGCCCGTAACGTTTGGGATA	
	CCGCCGACAAGCATTTCATCAACCACTATGGCTTCTCCATCTTGGACATTGTGCAAAACAACCCTAAGGAATTG	
	ACTATCCACTTTGGAGGTGCTAAAGGTAGAGCTATCAGAGACAACTACATTGGTATGATGTTTGAAACAATTGGT	

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1729 of 3077

Table 19

Gene (Organism)	Nucleic acid sequence of the gene protein-encoding sequence (from START codon to STOP codon except where otherwise noted)	SEQ ID NO
	GAAGACGGTTCTTTGAAGTCCGAAAAGATCTTCAAGGACATTGACGAAAACACCACTTCCTACACCTTTGTTTCT GACACTGGGTTGTTGTCTGCTACTCAATTCACCCAACCCGCTTTGACTTTGATGGAGAAGGCTGCCTACGACGA TATCAAGTCTAAAGGATTGATTCCAAGTGACATCATGTTTGCTGGTCACTCTCTTGGTGAATACTCTGCTTTGAC TTCCTTGGCCAACGTTATGCCTATTGAATCCTTGGTTGATGTTGTCTTCTACAGAGGTATGACCATGCAAGTTGC TGTTCCAAGAGACGAGTTTGGTAGATCCAACTACGGTATGGTTGCTGTCAACCCAACCAGAGTCAGCCCAACAT TTGACGATGCCGCCATGAGATTTGTTGTTGACGAGACTGCCAAGAGAACCACCTGGTTGTTGGAAATTGTCAAC TACAATGTTGAAAACCAACAGTACGTTGCTGGTGACTTGAGAGCCTTGGATACCTTGACCAACGTGTTGAA TGTTTTGAAGATCAACAAGATTGATATTGTCAGATTGCAAGAACAATTATCCCTCGACAAGGTCAAGGAGCACTT GTACGAGATTGTTGATGAAGTTGCTGCCAAGTCCATTGCTAAGCCACAACCAATTGAATTAGAAAGAGGTTTTG CTGTTATCCCATTGAAGGGTATTTCTGTCCCATTCCACTCTTCCTACTTGATGTCTGGTGTCAAGCCATTCCAGA GATTCTTGTGCAAGAAGATTCCAAAGGCTTCCATCAAACCACACAAGATTTGATTGGCAAGTACATTCCTAACTTGA CTGCTAAGCCATTTGAACTTACTAAGGAATACTTCCAGGATGTCTACGACTTGACTAAATCTGAAAAGATCAAGG CTATCTTGGACAACTGGGAAAAATACGAATAG	
FAS2 fatty acid synthase (*Candida viswanathii* ATCC20336)	ATGAAGCCAGAGATTGAACAAGAAATTATCCCACACCTTGTTAACAGAATTGTTAGCTTATCAGTTCGCTTCTCCA GTCAGATGGATCGAAACCCAAGATGTCTTCTTGAAGCAACACACAACACCGAAAGAATCATCGAAATCGGCCCTTC CCCAACCTTGGCCGGTATGGCCAACAGAACCATCAAGGCCAAATACCAATCCTACGACGCCGCGTTGTCCTTG CAAAGAGAAGTCTTATGCTACTCTAAGGACGCCAAGGAGATCTACTACAAGCCAGATCCAGCAGATCTTGCTCC AAAGGAAGAACCAAAGAAGGAAGCTGCCGCCGCTCCAGCCGCTACACCAGCTGCTGCTGCTGCTGC ATTCGGTTCCACTCCTGAAAAACCAGAAGATACCCCATTGGAAGAATTGGCCGAACAGTTCCAAGACTCCTTCA GTGGTCAATTGGGTAAGACTTCTACTTCATTGATTGGTAGATTGATGTCTTCTAAGATGCCTGGTGGTTTCTCAA TCACCGCTGCCAGAAAATACTTGGAATCCAGATTCGGTTTGGGTGCCGGTAGACAAGACTCTGTCTTGTTGGTT GCTTTGACCAACGAACCTGCAAGCAGATTGGGTTCTGAGGCCGAAGCTAAGACCTTCTTGGACACCATGGCTC AGAAATATGCCTCATCTGCTGGTATTTCCTTGTCGTCAGCTTCTGCCGGTGCCGGTGCTGGAGGTGCCGCCGG TGGCGCCGTTGTTGACAGTGCTGCTTTGGACGCCTTGACTGCTGAAAACAAGAAATTGGCTAGACAACAATTAG AGGTCTTGGCTAGATACTTGCAAGTCGACTTGAACTCAGGAGCTAAGTCTTTTATCAAAGAAAAAGAAGCTTCC GCTGTTTTGCAGAAAGAATTGGACTTGTGGGAAGCCGAACATGGTGAATTCTACGCCAGAGGTATCAAACCAAC TTTCTCAGCTTTGAAAGCAAGAACCTATGATTCCTACTGGAACTGGGCCAGACAAGATGTTTTTGTCCATGTACTT TGATATTTTGTTTGGTAAGTTGACCTCCGTTGACAGAGAAACCATCGACCAATGTATCCAAATTATGAACAGATC CAACCCAACTTTGATCAAGTTTCATGCAATACCACATTGACCACTGCCCAGAATACAAGGGTGAGACTTACAAGT TGGCCAAGAGATTGGGTCAACAGTTGATTGACAACTGTAAGCAAACCTTGAATGAAGACCCAGTGTACAAGGAC GTTTCTAGAATCACTGGTCCAAAGACCACCGTCTGCGCCAAGGGTAACATTGAATACGAAGAAGCCGAAAAGG	103

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1730 of 3077

Table 19

Gene (Organism)	Nucleic acid sequence of the gene protein-encoding sequence (from START codon to STOP codon except where otherwise noted)	SEQ ID NO
	ATTCTGTTAGAAAGTTTGAACAGTACGTCTACGAAATGGCCCAAGGTGGTGAAATGACCAAGATTGCCCAACCA	
	ACTATTCAAGAAGACTTGGCCAGAGTTTACAAAGCCATCTCCAAGCAAGCTTCCAGAGACAGCAAGTTGGAATT	
	GCAGAAAGTCTACGAGCAATTGTTGAAGGTTGTTGCTGGTTCAGACGAAATTGAAACTCAGCAATTAACCAAGG	
	ACATCTTGCAAGCTCCAACTGGCTCCAACACCCCAACTGATGAAGATGAAATTTCCACCGCCGACTCTGACGAT	
	GAAATTGCTTCATTGCCAGACAAGACTTCAATTGCCCAACCAGTTTCTTCAACTGTTCCACCCCAGACCATCCCA	
	TTCTTGCACATTCAAAAGAAGACCAACGAAGGCTGGGAATACGACCGCAAGTTGTCTGCCCTTTACTTGGACGG	
	TTTGGAATCCGCTGCTGTCAACGGTCTCACCTTCAAGGACAAGTACGTTTTGGTTACCGGTGCTGGTGCTGGAT	
	CCATTGGTGCCGAAATCTTGCAAGGTTTGATCAGTGGTGGTGCCAAGGTTGTTGTTACCACCTCTAGATTCTCC	
	AAGAAGGTTACTGAGTACTACCAAAACATGTACGCCAGATACGGTGCTGCCGGTTCTACTTTGATTGTTGTTCC	
	ATTCAACCAAGGTTCTAAACAAGATGTTGACGCTTTGGTTCAATACATCTACGACGATCCAAAGAAGGGTGGTTT	
	AGGCTGGGACTTGGATGCCATTATCCCATTCGCTGCTATCCCAGAAAATGGTAACGGTATCGACAACATTGATT	
	CTAAATCCGAATTTGCCCACAGAATTATGTTGACCAACCTTTTGAGATTGTTGGGTGCTGTCAAATCCAAGAAGA	
	CTACCGACACCAGACCAGCTCAATGTATCTTGCCAATGTCTCCTAACCACGGTACTTTCGGTTTCGATGGGTTG	
	TACTCTGAATCCAAGATTTCCTTGGAAACCTTGTTCAACAGATGGTACTCCGAAGATTGGGGCTCCAAGTTGAC	
	CGTCTGTGGTGCCGTTATTGGTTGGACCAGAGGTACTGGTTTGATGAGCGCCAACAACATCATTGCCGAAGGT	
	ATCGAAAAGATTGGTGTCAGAACCTTCTCCCAAAAGGAAATGGCTTTCAACATCTTGGGTTTGTTGACTCCAGA	
	GATTGTCAAGTTGTGCCAAGAAGAACCAGTTATGGCCGACTTGAACGGTGGTTTGCAATTCATTGAAAACTTGA	
	AGGATTTCACTTCCAAGTTGAGATCTGACTTGATGGAATCCGCTGAAGTTAGAAGAGCTGTCTCCATTGAATCC	
	GCCATCGAACAAAAGGTTGTCAATGGTGACAATGTTGATGCCAACTACACCAAGGTTACCGTTCAACCAAGAGC	
	CAACATGAAATTCGACTTCCCAACCTTGAAATCGTACGATGAGATCAAGAAGGTTGCTCCAGAATTGGAAGGCA	
	TGTTGGACTTGGAATCCGTCGTTGTTGTCACCGGTTTTGCCGAAGTTGGTCCATGGGGTAACGCCAGAACCAG	
	ATGGGAAATGGAATCCAAGGGTGAATTCTCCTTGGAAGGTGCCATTGAAATGGCCTGGATCATGGGTTTCATCA	
	AGTACCACAACGGTAACTTGAAGGGTAAGCCTTACTCTGGTTGGGTTGATGCCAAGACCCAAACTCCAATCGAT	
	GACAAGGACATCAAGGCCAAGTACGAAGAAGAGATCTTGGACCACTCTGGTATTAGATTGATTGAGCCAGAATT	
	GTTCAATGGCTACGATCCAAAGAAGAAGCAGATGATCCAAGAAGTTGTCATCCAACATGACTTGGAACCATTTG	
	AAGCCTCCAAGGAAACTGCTGAACAATACAAACACGAACACGGTGACAAGTGTGAGATCTTTGAAATTGAAGAA	
	TCCGGTGAATACACTGTTAGAATCTTGAAAGGTGCTACCTTGTTTGTTCCAAAGGCTTTGAGATTTGACAGATTG	
	GTTGCTGGTCAAATTCCAACTGGTTGGGATGCTCGTACCTACGGTATTCCAGAAGATACCATTAACCAAGTTGA	
	TCCTATCACTTTGTACGTCTTGGTTGCTACCGTTGAAGCTTTGTTGTCTGCTGGTATCACCGACCCATATGAATT	
	CTACAAGTACGTCCACGTTTCCGAAGTTGGTAACTGTTCTGGTTCCGGTATGGGTGGTGTCTCTGCCTTGAGAG	
	GAATGTTCAAGGACAGATACGCCGACAGACCAGTGCAAAACGATATCTTGCAAGAATCTTTCATCAACACCATG	
	TCCGCCTGGGTTAACATGTTGTTGTTCTTCGGGTCCAATCAAGACCCCAGTTGGTGCCTGTGCTACCGC	
	TGTTGAATCCGTTGACATTGGTATTGAAACTATTTTGTCTGGTAAGGCTAAGGTTGTTATGGTTGGTGGTTACGA	
	TGACTTCCAGGAAGAAGGTTCTTATGAATTCGCCAACATGAATGCCACTTCCAACTCCCTTGACGAGTTTGCTC	
	ACGGCAGAACTCCAAAGGAGATGTCCAGACCAACTACCACACCACCAGACACGGTTTCATGGAGGCCCAAGGTTC	

Table 19

Gene (Organism)	Nucleic acid sequence of the gene protein-encoding sequence (from START codon to STOP codon except where otherwise noted)	SEQ ID NO
	TGGTATCCAAGTTATTATGACTGCTGACTTGGCCATCAAGATGGGTGTTCCAATTCACGCTGTGTTGGCCATGT CTGCTACTGCTACCGACAAGATTGGTAGATCTGTTCCAGCTCCAGGTAAGGGTATTTTGACCACTGCCAGGGAA CACCACGGTAACTTGAAGTACCCATCTCCAGCTTTGAACATCAAGTACAGAAAGAGACAATTGAAGGCTAGATT AGACCAAATCAAGGCTTGGGAAGAAGCTGAAATTGCTTACTTGCAAGACGAAGCTGAGTTGGCCAAGGAAGAA ATGGGCGATGAGTTCTCCATGCACGAATTCTTGAAGGAAAGAACTGAAGAAGTGTACCGTGAATCCAAGAGACA AGTTTCTGACGCTAAGAAGCAATGGGGTAACCAATTCTACAAGTCTGACCCAAGAATTGCTCCATTGAGAGGTG CCTTGGCTGCTTTCAACTTGACCATTGACGATCTTGGTGTTGCTTCCTTCCACGGTACTTCTACCGTCGCCAAC GATAAGAACGAATCCGCCACTATTAACAGCATGATGCAACACTTGGGCAGATCTGAAGGTAACCCAGTGTTTGG TGTTTTCCAGAAGTACTTGACTGGTCATCCAAGGGTGCTGCTGGTGCTTGGATGTTGAACGGTGCCATCCAGA TCTTGGAGTCTGGTATTGTTCCAGGTAACAGAAATGCCGATAACGTTGACAAGGTCTTGGAAGAATACGAGTAC GTCTTGTACCCATCCAGATCCATCCAAACTGACGGTATCAAGGCCGTTTCCGTGACCTCTTTCGGTTTCGGTCA AAAAGGTGCTCAAGCTGTTGTCGTCCACCCAGACTACTTGTTTTGCTGTTTTGGATAGATCTACTTATGAGGACTA CGCCACCAGAGTTTCTGCCAGAAACAAGAAGACTTACCGTTACATGCACAATGCTATTACTAGAAACACTATGTT TGTTGCTAAGGATAAGGCTCCATATGCCGATGAATTGGAACAACCAGTTTACTTGGACCCATTAGCCCGTGTTG AAAACGCTAAGGAAAAGCTTGCCTTCAGCAACAAGAGTATCCAATCCAACCAAGCTTATGCTGGTGAAAATGCC AGAACCACTGCCAAGGCTTTGGCTGCCTTGAACAAGTCATCCAAGGGTGTTGGTGTCGACGTTGAATTATTGTC TGAGCTCAACTTGGAGAATGAAACTTTTGTTGCAAGAAACTTCACTCCTGGTGAAATCCAATACTGCTCCAAGAG TGCCAACCCACAAGCTTCATACACCGGCACTTGGTCTGCTAAAGAAGCTGTCTTCAAGGCATTAGGTGTTGAAT CTAAAGGTGCTGGTGCCAGCTTGGTTGACATTGAGATCACTCGTGACGTCAACGGCGCTCCACAAGTTGTCTT GCACGGTGATGCGGCAAAATCAGCCGCCAAAGCTGGTGTCAAGAACGTCAAGATTTCCATCTCCCATGACGAC TTCCAAGCCACTGCTGTTGCCTTGAGTGAATTCTAG	
HPD1 3-hydroxy-propionate-dehydrogenase (*Candida viswanathii* ATCC20336)	Atgttgagatcttcagtccgtactttctccacccagtccagagtattagccaactacggtttcgtaggcttggtctcatggccagcacatggccagacacgtctacaa ccagttgcagccagcagacaagttgtatgtccacgagacgtcaacccccagcacaccacccagttcgtcaccgacgtgaccaccagaagccacagaacgccac acaattgacgcccttgtcctccttgaaagagttcaccaccgagccagagtcccagttggacttcatcgtcaccatggtcccgagggcaagcacgtcaaggccgttg tctccgagctagtcgaccactacaatgcgtcgggaaaatacgacccatccaagaagttgacctttgtggactcctcaccatcgacatcccacctcagggaggt ccaccagctcgttgccgacaagttacaaggcgccacgttcatcgacgccccggtttcgggtggtgtcgctggtgccaggaacggaacctttgtcgttcatggtgtcgc gggacaccaaggaagacgtcgaccctaacctcgtcacgctttgaactacatgggcagcaacatcttcccatgtggtggaacccacgggaccggcttggctgcca agttggcaaacaactacttgttggcgatcacgaacatcgccgtcgcagatagcttccagttgggttgaacttgcagaactacgccaagttggtgt cgacctccacaggtaagtcctgggctagtgtcgataactgcccaatcccggtgtctaccctgaaaagaacttgacttgtgataacggatacaaggggtgggtttgtca cgaagttgacgagaaaggatgtgtcttggctacggagtcgtctaaggctaacaaccaagttccttatgcttggcgaagtcggtagatactggtacgacaaggcttgtg aagatgaaaagtacgccaacagagacttgtgttctttttcgaattcttgggtgatcttaaaaaataa	104

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1732 of 3077

Table 19

Gene (Organism)	Nucleic acid sequence of the gene protein-encoding sequence (from START codon to STOP codon except where otherwise noted)	SEQ ID NO
ALD6 malonate semialdehyde dehydrogenase (*Candida viswanathii* ATCC20336)	atgttatccagagttctttttcaagactaaaccaagagttcctactaaatcaatcaccgccatggccatcagaaacaaatccatcgtgacttatcctccaccacctccac atacccaaccgacccacacgacccgtcacgggagccatacatcacgccatcttcgtgaacaacgagttcatcaagtcggactccaacacctggttcgacgtgca cgacccggccacgaactacgtcgtgtccaaggtgccacagtcgacgccggaggagttggaagaggcgatcgtcggcccatgccgcgttccccaagtggcgc gacaccagcatcatcaagcgtcagggggatcgcgttcaagtttgtgcagttgttgcgagaacatggacagaatcgcaagcgtcattgtcttggaacagggtaaga cgtttgtcgatgcccagggtgacgtgactagaggattgcaggttgctgaggctgcgtgcaacatcactaatgacttgaaggtgagtcgttggaagtgtctactgatat ggagaccaagatgattagagaacctttggtgttgtgggatccatctgtcctttaacttcccagctatggtcccattgtggtctttgccttggttttggtcacgggtaacact gctgtgattaagccttccgagagagtcccgggcgcaagtatgattatttgtgaatggccgccaaggctggtgttccacctggtgtgttgaacattgtccacggtaagca cgacaccgtcaacaagttgattgaggacccaagaatcaaggcattgactttttgttggtggtgacaaggccggtaagtacatttacgaaaaggttccagtttgggca agagagtgcaggccaacttgggtgctaagaacccacttggtgttgccagaacgcacacaagcagagtttttgtcaatgccgtcaacggtgccgctttcgggtgctctg gacagagatatggctatttctgtctggtcaccgtgggtaagaccaaggaatgggtgcaggatgtcatcaaggacgccaagttgttgaacaccggaagtggatttt gacccaaagagtgacttgggtccagtcatcaaccccagagtccttgactctgctgcgtgctgaagaaatcattgctgattccgtgcctgttggaattggacgg aagaggatacagaccagaagacgctagattcgccaagggtaacttcttgggtccaaccatcttgaccaacgtcaagccaggcttgagagcatacgacgaagaag atttcgctcctgtttgtctgtggttaacgtcgacaccattgacgaagccattgagttgatcaacaacaacaagtacggtaacggtgtttcattatttacttcctccggtggc tcagcccagtatttcaccaagagaatcgacgtcggtcaagtcggtatcaatgtcccaatcccttgttccattgcctatgttctccttcactggttccagaggctccttcttggg tgacttgaacttctacggtaaggccggtatcaccttcttgaccaagccaaagaccatcactagtgcctggaagaccaacttgattgatgacgagatcttgaaaccatc tacctcgatgcctgtccaacagtaa	105
ECI1 enoyl-CoA isomerase (*Candida viswanathii* ATCC20336)	Atgtccgacgaggaatcagatatcttatacgaggtcagagacagaaccgccatcatcaccttgaacatcccaagagattgaacgcattgaacggtgctcaatact tgaagttgggtaagttcttggagagagccaacaacgaagaggacaccgtcttgacatttgatccaggccctgggcagattcttctccgccggtgccaatttcgccgac aacgatatggccaaggtcgaaatgtccaagttgttcagtcacgagtactggttgaaagattcgtcgccagaaacatctggttgaccaacttgttcaacgaccacaa gaagatcttggctgctgtcaatggtccagttatcggtttgagcactggttgttgttggtcgatttggtctacgtccacgacttgaacaagttctacctcttggcccca ttgccaacttgggtttggttgccgaaggtgcttcctctgccacttgttcaacagattgggctggtcaaaggcttctgaggccttgttgttggccaagccaatcggcggcc aagactgttacaacgccggtttcatcaacaagcactacgacgtaagtttcctccactgaagagttcaacgaacacgtctacaaggagttgacggaagctttgaa aacttgcacgatgactccatttgcagaacaagcaattgttgaagttgtccagagaccaggccatcaactag	106
ECI2 enoyl-CoA isomerase (*Candida viswanathii* ATCC20336)	Atgtccgacgacctatcacctacgaagtcaagaccgagctgccgtgatcaccttgaacaacccaagaagctcaacgccttatcgatccgcagtacgacacc atctgcaagctcttagaacgagccaacgccgaagaagacaccgtcatcacttgtccagtccacgggccgggtgttctctgccgggccaacgccgactccatc gtgggcaggatgccgagctcgagacctggttgaacatgtcggtggccaagcagacgttcttggtgcagacgttcctcgcacacaagaagatccttgccgtcgcctt gaacggcccgtgattggcttatcggcggcgttcgtggcgctctgcgacttggtctacgtgcacaacgccgcaaagacgttcttcttgacccggttcgccaacatcgg gatccttgccgaggccggcacctcagccacgttgcccatgcgcgtgggtggtccaggccgcggaagcgttgttgttgtcaaagaggatttcgggagatgacttg cagagagcgggttcttcaataaggactacaagggcagttcaagtccgcggaggagttaacgaggtcgtcttgaaggagttgcttgacgccacggaaaacttg catgaggactcgatcatccagaacaaggagttgttgaaggctatttcaagccaaagatcagtgaggtcaactcgcaggaggtgtgtgtacaagtgg acctctgggtgccaatggatagatttaaaaaattgcttaatggtgagttgaaacataaattatag	107

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1733 of 3077

Table 19

Gene (Organism)	Nucleic acid sequence of the gene protein-encoding sequence (from START codon to STOP codon except where otherwise noted)	SEQ ID NO
DCR1 dienoyl-CoA reductase (*Candida viswanathii* ATCC20336)	atgccaaacacttagatcacaactacttaagaaaaagcgtctggaagtcagatatcttgccggcaaagtcatcttcatcactggtggtgccggtaccatctgtcgtgt gcaagccgaagccatggtcttgttgggtgctaatgctgccatcattggtagaaaacgtcgaaaagactgaagaagccgctaaggagattgcttccttgagaccagga gccaaggtcattggtattggtgccgttgacgtgcgtaagatccaaactatcaaggatgccgttgacagaactgttgccgagttgggcagaatcgactatgtcattgctg gtgccgccggtaacttccttgtgacttcaaccacttgtctgccaacgccttcaagtccgtgattgacattgatttgttgggttctttcaacaccgtcaaggttaccttgacc agttggagaaagaacaaaggtgccgtgttgtttgtcagtgccacttgcactacaccggtgttcctatgcaatcgcacgttggtgctgccaagtctggtgttgatgccttga gcaacgccctgctgttgaattgggtccattgggtatcagattcaactgtattgcccaggtgccattgctggtactgaaggtatgctgagattggcccccacctacccgac acgccattggagacgaagattccattgcagagacaaggtactactgaggacattgctgatgccactgttttgttctcgccagcttcctcctacattactggtgatgtttt ggttgttgatggtgctatgtggcaaactggcggaggttattcttaacgacttctacccagacatcattattcaccagaaacgccgacccagaaggtaagttgtaa	108
DCR2 dienoyl-CoA reductase (*Candida viswanathii* ATCC20336)	atgccaaacacttaactgaagcatacaaacaagaaagttcttggaagccagatttgttcaagggcaaagtcgtcttcatcaccggtggtgcaggttcgatctgccgt gtccaggccgaagccatggtcttgttgggtgccaacgctgccatcattggtagaaacgtcgaaaagaccgaatccgccgccaaagaaatcgccagcttgagatca ggagccaaggttcttggtattggtggcattgatgtgcgtaaagttgacagcttgaagagcgccgtcgataaagaccgttgccgaattgggcagaatcgactttgtcattg ctggtgccgccggtaacttccttgtgacttcaaccacttgtctccaacgccttcaaatccatcattgacattgatttgggttctttcaacaccgtcaaggtcactttga ccaattgagaaagaacaaggtgccatcttgtttgtcagtgccaccttgcactactacggtgttccttccaaattggtgttggtgctgccaaggctggtgtcgacgcctt gagcaatgccctgccgttgagttgggtccattgggcatcagatccaactgtattgccccaggtccaatcgacggtaccgaaggtgttgaaagattggtcagagcttc caaggccgaagccgccaagaaggttccattgcaaagattgggtactacacaagacattgctgacggtaccgtctacttgttctccccagccgcctcctttgtcactgg tgacgttttggttgttgatggtgctagctggcagatctcctctggtgttggtgccaccaactacccagtctccatctcaatgcgatcgacgcaccgaaggtggtaagtt gtaa	109
MCR malonyl-CoA reductase (*Sulfolobus islandicus*)	atgaggagaactcttaaggccgcccattttgggatcgactggattggttggtatcgaatatgttagaatgttagcaaaccaccctacataaaggtgggttacctagcgg gtaagggctcagtcggcaagccctacgaagaagttgtcagatggcaaaccgtcggccaagttcctaagaaatcgctgatatggaagtcaagcctaccgatccc aagttgatggacgacgtcgacattatctttccgcttgcccaaggcgctgctggtcctgttgaagaggaatttgccaaacatggatttcctgtcatttccaactctcctg atcacagattcgacccagatgtcccctctcctcatcccagaaatcaatccgcatacaatctcgcttattgacaaacaaaggaaaggcatgactggaggggcttcatc gttaccactccttgtgcaccgccgcaaggtgctgccattccttggccctatctatatgaatttcagaatcgactccagtctcatcactaccattcaatcgctctctggcgc tgggtatccaggtatcccatcattagacgttgtcgacaacgtcttgccattaggcgataattatgacacaacaaaacagttaaagagatttccaggattttgtccgagaca aagaggatggtgaacgatgacaacgatctatcgttgggcgccactacccacagaatctcgactattcacgggcattacgaggtcatttatgtaacctcagagaaaga cgtttccgtcgaaaagataagagagacgttagactcatttagaggtgaaccccagaagttgaagttgcctaccgcgcctgacaagcctatactttgaccaacccaag acgcaagaccacaagtctacttgacaggtgggccggagaccctccaggtatgtccgtggtcgtgggtagattgagtcaaatcaacaggagaaaccattaggctcgt ctcggtcgttcataatacccgtcagggaagctgctggtggtggaatcttggccgctgaactactcgtagagaagggttacattgacaagaggtaa	323

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1734 of 3077

Table 19		
Gene (Organism)	**Nucleic acid sequence of the gene protein-encoding sequence (from START codon to STOP codon except where otherwise noted)**	**SEQ ID NO**
2PS 2-pyrone synthase (*Gerbera hybrida*)	atgggatcctactcaagtgacgacgtcgaagtcattcgtgaagccggacgagcgcaaggtctcgccaccattttagccattggcactgctactcctccgaattgtgttg ctcaagctgactatgcggattattatttccgtgttactaagtctgagcatatggtggacttaaaagaaaaattcaaacggatttgcgaaaaaacggcaataaagaaac ggtacctagccttgacagaggattatctccaagaaaacccaacgatgtgcgaatttatggctccaagtctcaacgctcggcaagatctagtcgttacaggcgttcca atgttaggcaaagaggccgcggttaaggccattgacgagtggggactaccaaaaagtaagatcacacacttgatcttttgtacaacagctggcgtggatatgccgg gtgctgattatcaattggttaaattgttaggtttgagtccttcagttaaacggtatatgctttaccaacagggatgcgccgccggcggcacggttttgcgactagccaagg atttagctgagaacaacaagggctcacgggttttaatcgtttgtagtgaaatcactgctatcctctttcatggaccgaatgaaaaccacttagattcattggttgctcaag ctctctttggagatggagctgcggcgttgattgtcggttcaggccctcaccttgccgtagagcgaccaatatttgaaatcgtctcaactgaccaaacgatccttcccgat actgaaaaggcgatgaagctccaccttagagaaggagggcttaccttccagcttcatagagacgtaccgcttatggttgcgaagaacatagaaaacgcggcaga aaaagcactttccccactagggataactgactggaactcagtgttttggatggtccacccaggtggtcgggccatacttgatcaggtcgaacggaaactaaacttaa aggaggacaagctcagggcttctaggcatgtcttaagcgagtacggaaacctcatttctgcttgcgtccttttatcattgatgaagtcaggaagagatccatggcaga ggggaagagcacgacaggtgagggtcttgactgtggtgtgcttttcggattcggacccggtatgactgtggaaactgtggtgttacgttctgttcgggtgactgctgcag tggccaatggaaactaa	325

EXAMPLE 27: Nucleic acid sequences of plasmids and other nucleic acids referenced herein

Table 20		
Plasmid	**Nucleic acid sequence**	**SEQ ID NO**
pAA073	aaacgccagcaacgcggccttttacggttcctggccttttgctggccttttgctcacatgttctttcctgcgttatcccctgattctgtggataaccgtattaccgcctttgagt gagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtcagtgagcgaggaagcgagttcggcatggcagatcatcatgcctgcaggagctccaa ttgtaatatttcgggagaaatatcgttggggtaaaacaacagagagagagagggagagatggttctggtagaattataatctggttgttgcaaatgctactgatcgact ctggcaatgtctgtagctcgctagttgtatgcaacttaggtgttatgcatacacacggttattcggttgaattgtggagtaaaaattgtctgagttgtgtcttagctactggct ggcccccccgcgaaagataatcaaaattacacttgtgaatttttgcacacacaccgattaacatttccctttttttgtccaccgatacacgcttgcctcttctttttttttctctgtgct tcccctcctgtgactttttccaccattgatatataaaatcaactccatttccctaaaatctccccagattctaaaaacaacttcttctcttctgcttttccttttttttttgttatatttattta ccatccctttttttgaatagttattccccactaacattgttcaaatcttcacgacataagaagagcccgggtctagatgtgtgctcttccgagtgactcttttgataagagtc gcaaatttgatttcataagtatatattcattatgtaaagtagtaaatggaaaattcattaaaaaaaaagcaaatttccgttgtatgcatactccgaacacaaaactagcc ccggaaaaacccttagttgatagttgcgaatttaggtcgaccatatgcgacgggtacaacgagaattgtattgaattgatcaagaacatgatcttggtgttacagaac atcaagttcttggaccagactgagaatgcacagatatacaaggcgtcatgtgataaaatggatgagatttatccacaattgaagaaagagtttatggaaagtggtca accagaagctaaacaggaagaagcaaacgaagaggtgaaacaagaagaagaaggtaaataagtattttgtattatataacaaacaaagtaaggaatacagat ttatacaataaattgccatactagtcacgtgagatatctcatccattccccaactcccaagaaaaaaaaaaagtgaaaaaaaaaatcaaacccaaagatcaacct	110

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1735 of 3077

Table 20

Plasmid	Nucleic acid sequence	SEQ ID NO
	ccccatcatcatcgtcatcaaacccccagctcaattgcatggttagcacaaaaacatacacagaaagggcatcagcacacccctccaaggttgcccaacgttt attccgcttaatggagtccaaaaagaccaacctctgcgcctcgatcgacgtgaccacaaccgccgagttccttcgctcatcgacaagctcggtcccccacatctgtct cgtgaagacgcacatcgatatcatctcagacttcagctacgagggcacgattgagccgttgctgtgcttgcagagcgccacgggttcttgatattcgaggacagga agtttgctgatatcgggaaacaccgtgatgttgcagtacacctcgggggtataccggatcgcggcgtggagtgacatcacgaagcgcacggagtgactgggaag ggcgtcgttgaaggttgaaacgcggtgcggaggggtagaaaaggaaagggcgtgttgatgttggcggagttgtcgagtaaaggctcgttggcgcatggtga atataccgtgagacgatcgagaattgcgaagagttcgtgattgggttcatcgcgcagcgggacatgggggtagagaagaaggtttgattggatc atcatgacgcctggtgtgggattggatgataaaggcgatgccgttgggccagcagtataggactgttgatgaggtggttctgactggtacgatgtgattattgtcggga gagggttgtttggaaaaggagaagacctgaggtggaggaaagagatacagggatgctggatggaaggcatacttgaaggagaactggtcagttagaataaat attgtaataaataggtctatatacatacactaagcttcattgtagtcttcgaagttgtctcgtagtttagttccatgattcgaaaaccaataacgcaatggat gtagcagggatggttgttagtgcgttcctgacaaacccagagtacgccgcctcaaaccacgtcacattgccgccttgcttcatccgcatcacttgcttgaaggtatcca cgtacgagttgtaatacaccttgaagaacggcttcgtctacgcgcgagacgaaagggcctcgtgatacgcctatttttataggttaatgtcatgataataatggtttctta gacgtcaggtggcacttttcgggggaaatgtgcgcggaacccctatttgtttatttttctaaatacattcaaatatgtatccgctcatgagacaataacctgataaatgcttc aataatattgaaaaaggaagagtatgagtattcaacatttccgtgtcgcccttattcccttttttgcggcatttttgcttccttgtttttgctcaccgagaaacgctggtgaaagt aaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaacagcggtaagatccttgagagttttcgccccgaagaacgtttccaatgat gagcactttaagttctgtatgtggcgcggtattatcccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctcagaatgacttggttgagtact caccagtcacagaaaagcatcttacggatggcatgacagtaagagaattatgcagtgctgccataaccatgagtgataacactgcggccaacttacttctgacaac gatcggaggaccgaaggagctaaccgcttttttgcacaacatgggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccatacccaacga cgagcgtgacaccacgatgcctgtagcaatggcaacaacgttgcgcaaactattaactggcgaactacttactctagcttccccggcaacaattaatagactggatg gaggcggataaagttgcaggaccacttctgcgctcggctccttcggctggctgtttattgctgataaatctggagccggtgagcgtgggtctcgcggtatcattgcag cactgggccagatggtaagccctcccgtatcgtagttatctacacgacgggggagtcaggcaactatggatgaacgaaatagacagatctgctgagatagggtgcct cactgattaagcattggtaactgtcagaccaagtttactcatatatactttagattgatttaaaacttcatttttaatttaaaaggatctagttgaagatcctttttgatcttc atgaccaaaatcccttaacgtgagtttctgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcctttttttctgcgcgtaatctgctgcttgc aaacaaaaaaacaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagataccaaatac tgttcttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctgccagtggcgataa gtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacggggggttcgtgcacacagcccagcttggagcgaacga cctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttccgaagggagaaaggcggacaggtatccggtaagcggcagggtcgga acaggagagcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgactgagcgtcgatttttgtgatgctcgtcag ggggcgggagcctatggaa	
pAA298	cccgtagaaaagatcaaaggatcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatca agagctaccaactctttttccgaaggtaactggcttcagcagagcgcagataccaaatactgtccttctagtgtagccgtagttaggccaccacttcaagaactctgta gcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggg cgcagcggtcgggctgaacggggggttcgtgcacacagcccagcttggagcgaacgactacaccgaactgagatacctacagcgtgagctatgagaaagcg ccacgcttcccgaaggagaaaggcggacaggtatccggtaagcggcagggtcgaaacaggagagcgcacgagggagcttccagggggaaacgcctggta tctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggcctttttacg	111

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1736 of 3077

Table 20

Plasmid	Nucleic acid sequence	SEQ ID NO
	gttcctgggcttttgctggccttttgctcacatgttctttcctgcgttatcccctgattctgtggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccga	
	acgaccgagcgcagcgagtcagtgagcgaggaagcggaagagcgcccaatacgcaaaccgcctctccccgcgcgttggccgattcattaatgcagctggcac	
	gacaggtttcccgactggaaagcgggcagtgagcgcaacgcaattaatgtgagttagctcactcattaggcaccccaggctttacactttatgcttccggctcgtatgt	
	tgtgtggaattgtgagcggataacaatttcacacaggaaacagctatgaccatgattacgccaagctatttaggtgacactatagaatactcaagctatgcatcaagc	
	ttggtaccgagctcggatccactagtaacggccgccagtgtgctggaattcgcccttcgttaaacaaaaatcagtctgtaaaaaaggttctaaataaatattctgtcta	
	gtgtacacattctcccaaaatagtgaaatccagctctacaatttggctttaccggtacaaatcaaagaccaatcgtcttcagtcaactcctggtacttgtctgccattcaac	
	cagtagatcaaatcctgccgtcttcaccctttggcaacttttggttctgaattggttctaggaacctgtggttgtgcgaagcctcaatggtgccaatcttgatgaacgca	
	ggttgagcatacacaggcaaagacttggtcacgtgagagtgaatcaatttcaagatttctcatggctcaactcgtcctggcttcacagacggccaaaacaggctcta	
	ccttcgtggtttggcaccttgacaccggacaacggacgctccaaggcctggagcccatcaattcgttctcgacctcggtggcgggagacgttttcggaacttccaac	
	ggaaagtgtcacctaatctgtcgacaaagtacacaacaatttgtcctcgtccatctccaacaagtcaccggatctgtaccacgcgtcacctttttgaaaacattggtgagg	
	atttgctgttggtggcggacttgttaccataataacctggaaggatttctgcacgtcgttaggttcaagattctcatcaacaactcacctggctcgttgtaaggcggcct	
	cggtacagaaccggtcttgggtccttgtagatttcactctcgtcttctgggtccatttctgctggtagacaataacaagctgatgagggaccgtact	
	tacgacaggcgccgacaccgtactcaccgtactgcaagttggtggtggcggatagggggactcggtggcggcgtagaactcaccgatacttcaatgtggaatctgc	
	gcttgaactcagaccatatatctggacgcaaccgttaccgtaggcaattctgacattgtgtctgtcttggtctggatgaggcttggagttcaacaagtaacgacagac	
	ctcaccgacgtattgcacgtgggtggcaccacataatctggcctgggtccagaacgaagtagcggagaatttctgggacacggagacacagccaccgacaatca	
	aagtaggacacaacccaacatggccgcggtggagtggtacaaggggcatggcggtcaagacgttcgatttcgagtcaatcttcatgatgtggccaaagaaaacc	
	gaggccatgaaggctttttccaggacatgataccggcttttggcaaaccggtggtaccgaggtgtaaatcaatgcacaagcggaggagtcagtatcggttggtct	
	tctggtcttgtcctcggctctgtgtttttggagtcgacttgagtctcaatctgtcaaacaaggcaaactcgtcaatgtagtttatttgcacatgtggcaattcctctctgatctga	
	gcctcggtatctctgattgggaatcacagtccgggtcaacgaaaactgcgaagcgttgacaatcttaagacagtggatcaatggcttgtcctggtgttgaagttca	
	agaacgcaggcaaggcaccaatgttccacaatgccaacacagacaatgaaaagcggcttgttcatacaagaaacaccgatggtgtcgttggcagtgacgccc	
	gtactcgttcttcaagatgtatgagtactcaaaaccatgtcgtacaattcctgtaggtgtattcttctaggtcaaactgatcgtcgtagatccactagtaacggccgcca	
	gtgtgctggaattcgccctttgggctaacgaaaaggaaaccgctgacgttaaaggtatctacggttgtttcggtatgacccccgggggatctgacgggtacaacgaga	
	attgtattgaattgatcaagaacatgatcttggtgttacagaacatcaagttcttggaccagactgagaatgcacagatatacaaggcgtcatgtgataaaatggatga	
	gatttatccacaattgaagaaagagtttatggaaagtggtcaaccagaagctaaacaggaagaagcaaacgaagaggtgaaacaagaagaagaaggtaaat	
	aagtatttgtattatatacaaacaaagtaaggaatacagatttatacaataaattgccatactagtcacgtgagatatctcatccattcccaactcccaagaaaaa	
	aaaaaagtgaaaaaaaaaatcaaacccaaagatcaacctcccatcatcatcgtcatcaaacccccagctcaattcgcaattggttagcacaaaaaacatacaca	
	gaaagggcatcagcacacccctccaaggttgcccaacgtttattccgcttaatggagtccaaaaagaccaacctgcgcctcgatcgacgtgaccacaaccgcc	
	gagttccttcgctcatcgacaagctcggtcccacatctgtctggtgaagacgcacatcgatatcatctcagacttcagctacgaggcacgattgagccgttgcttgt	
	gcttgcagagcgccacgggttctgatattcgaggacaggaagttgctgatatcggaaaacaccgtgatgttgcagtacacctcgggggtatacccggatcgcgggcgt	
	ggagtgacatcacgaagcggcacggagtgactgggaagggcgtcgttgaagggttgaaacgcggtgcggaggggtagaaaaggaaagggcgtgttgatgt	
	tggcggagttgtcgagtaaaggctcgttggcgatggtgaatataccgtgagacgatcgagattgcgaagagtgatcgtgattgggttcgtgattgggttcatcgcgcag	
	cgggacatgggggtagagaagaaggtttgattggatcatcatgacgcctggtgtgggttggatgatataaggcgatgcgttgggccagcagtataggactgttg	
	atgaggtggttctgactggtaccgatgtgattattgtcgggagagggttgtttggaaaaggaagagaccctgaggtggaggaaagagatacagggatgctggatg	
	gaaggcatacttgaagagaactggtcagttagaataaatattgtaataaatagtggttagtgcgttcctgacaaaccccagagtacgccgcctcaaaccacgtcac	
	agtttagttctcatgatttcgaaaaccaataacgcaatggatgtagcagggatggttagtgcgttcctgacaaaccccagagtacgccgcctcaaaccacgtcac	

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1737 of 3077

Table 20

Plasmid	Nucleic acid sequence	SEQ ID NO
	attcgcccttgcttcatccgcatcacttgcttgaaggtatccacgtacgagttgtaatacaccttgaagaacggcttcgtctacggtcgacgacgggtacaacgagaat	
	tgtattgaattgatcaagaacatgatcttggttacagaacatcaagttcttggaccagactgagaatgcacagatatacaaggcgtcatgtgataaaatggatgag	
	atttatccacaattgaagaaagagtttatggaaagtggtcaaccagaagctaaacaggaagaagcaaacgaagaggtgaaacaagaagaagaaggtaaata	
	agtatttttgtattatataacaaacaaagtaaggaatacagatttatacaataaattgccatactagtcacgtgagatatctcatccattcccaactcccaagaaaaaa	
	aaaaagtgaaaaaaaaaaatcaaacccaaagatcaacctcccatcatcatcgtcatcaaacccccagctcaattcgcagagctcggtaccccgggatctggaat	
	ccctcggcgtcggtcttgggggtgggggcattctttcttggtctgggaacgccaacgctttgttgtttgggttcttgaacacggactgctcgaaaaagtaccagtatgatg	
	ccttacctctgctggcttccacaagtatgtgggaggcattgacagcgactgtcttggctaacagcacgtcgtcggcaattaaatatttggcttccaatatactgactacca	
	aggatggcagcacgcggctatttctaatcctgacatgtttctcgtacgtagtagtgaatgaaggtggaataatatcaagggcgaattctgcagatatccatcac	
	actggcggccgctcgagcatgcatctagaggcccaattcgccctatagtgagtcgtattacaattcactggccgtcgttttacaacgtcgtgactgggaaaaccctg	
	gcgttacccaacttaatcgccttgcagcacatcccccttcgccagctggcgtaatagcgaagaggcccgcaccgatcgcccttcccaacagttgcgcagcctatac	
	gtacggcagtttaaggtttacacctataaaagagagagccgttatcgtctgtttgtggatgtacagagtgatattattgcacgccgggggcgacggatggtgatcccc	
	tggccagtgcacgtctgtctgcagataaagtctcccgtgaacttaccccggtggtgcatatcgggtgatgaaagctggcgcatgatgaccaccgatatggccagtgtg	
	ccggtctccgttatcgggggaagaagtggctgatctcagccaccgcgaaaatgacatcaaaaacgccattaacctgatgttctgggaatatataaatgtcaggcatga	
	gattatcaaaaaggatcttcacctagatccttttcacgtagaaagccagtcgcgaaaacggtgctgaccccggatgaatgtcagctactgggctatctggacaagg	
	gaaaacgcaagcgcaaagagaaagcaggtagcttgcagtgggcttacatggcgatagctagactgggcggttttatggacagcaagcgaaccggaattgccag	
	ctgggcgccctctggtaaggttgggaagccctgcaaagtaaactggatggctttctcgccgccaaggatctgatggcgcagggggatcaagctctgatcaagaga	
	caggatgaggatcgttcgcatgattgaacaagatggattgcacgcaggttctcggccgcttgggtggagaggctattcggctatgactgggcacaacagacaat	
	cggctgctctgatgccgccgttccggctgtcagcgcaggggcgcccggttcttttgtcaagaccgacctgtccggtgccctgaatgaactgcaagacgaggcag	
	cgcggctatcgtggctggccacgacgggcgttccttgcgcagctgtgctcgacgttgtcactgaagcgggaagggactggctgctattgggcgaagtgccggggc	
	aggatctcctgtcatctcaccttgctcctgccgagaaagtatccatcatggctgatgcaatgcggcggctgcataccgctgatccggctacctgcccattcgaccacca	
	agcgaaacatcgcatcgagcgagcacgtactcggatggaagccggtcttgtcgatcaggatgatctggacgaagagcatcagggctcgcgccagccgaactg	
	ttcgccaggctcaaggcgagcatgccgacgggcgaggatctcgtcgtgaccatggcgatgcctgcttgccgaatatcatggtggaaaatggccgcttttctggattc	
	atcgactgtggccggctggtgtggcggaccgctatcaggacatagcgttggctaccctgatattgctgaagagcttggcggcgaatgggctgaccgcttcctcgt	
	gctttacggtatcgccgctcccgattcgcagcgcatcgccttctatcgccttcttgacgagttcttctgaattattaacgcttacaatttcctgatgcggtatttttctccttacgc	
	atctgtgcggtatttcacaccgcatacaggtggcacttttcgggggaaatgtgcgcggaaccctatttgtttatttttctaaatacattcaaatatgtatccgctcatgagac	
	aataaccctgataaatgcttcaataatagcacgtgaggagggccaccatggccaagttgaccagtgccgttccggtgctcaccgcgcgcgacgtgccggagcg	
	gtcgagttctggaccgaccggctcgggttctcccgggacttcgtggaggacgacttcgccggtgtggtcgggacgacgtgacccttgttcatcagcgcggtccagg	
	accaggtggtgccggacacaccctggcctgggtgtggcgcggcctggacgagctgtacgccgagtggtcggaggtcgtgtccacgaacttccgggacgccc	
	tccgggccggccatgaccgagatcggcgagcagccgtgggcgggggttcgccctgcgcgaccgggcccggcaactgcgtgcacttcgtggccgaggagcag	
	gactgacacgtgctaaaacttcatttttaatttaaaaggatctaggtgaagatcctttttgataatctcatgaccaaaatcccttaacgtgagttttcgttccactgagcgtc	
	agac	

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1738 of 3077

Table 20

Plasmid	Nucleic acid sequence	SEQ ID NO
pAA1164	cccgtagaaaagatcaaaggatcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatca agagctaccaactctttttccgaaggtaactggcttcagcagagcgcagataccaaatactgtccttctagtgtagccgtagttaggccaccacttcaagaactctgta gcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcg cgcagcggtcgggctgaacggggggttcgtgcacacagcccagcttggagcgaacgactacaccgaactgagatacctacagcgtgagctatgagaaagcg ccacgcttccgaaggggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttccagggggaaacgcctggta tctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggcctttttacg gttcctggccttttgctggccttttgctcacatgttctttcctgcgttatcccctgattctgtggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccga acgaccgaggcgcagcgagtcagtgagcgaggaagcggcccaatacgcaaaccgcctctcccccgcgttggccgattcattaatgcagctggcac gacaggtttcccgactggaaagcgggcagtgagcgcaacgcaattaatgtgagttagctcactcattaggcaccccaggctttacactttatgcttccggctcgtatgt tgtgtggaattgtgagcggataacaatttcacacaggaaacagctatgaccatgattacgccaagctatttaggtgacactatagaatactcaagctatgcatcaagc ttggtaccgagctcggatccactagtaacggccgccagtgtgctggaattcgcccttaaacgttgggcaacctggaggggtgtgctgatgcccttctgtgtatgttttt gtgctaacaatgcgaattgagctgggggtttgatgacgatgatgtgggaggttgatctttgggttgatttttttttcacttttttttttctgggagttgggggatggatga gatatctcacgtgactagtatggcaatttatgtataaatctgtattcctactttgtttgtatataatacaaaatactattaccttcttcttgtttgcctcttgttgctcttc ctgtttagcttctggttgaccacttccataaactctttctcaattgtggataaatctcatccatttttatcacatgacgccttgtatatctgttcagtctggtccaagaaac ttgatgttctgtaacaccaagatcatgttcttgatcaattcaatacaattctcgttgtaccgtcgcatatggtcgcaactaatcgcaactatcaactaagggttttccggg gctagtttgttgtcggagtatgcatacaacgggaaatttgctttttttttaatgaatttccatttactacttttacatgaatatatactattgaaatcaaatttgcgactcttatc aaaagagtcactctctctattcttaactggacaagatttcagcagcttcttcggacttttcacctctttctctgactcaaggttctgacgggttcaacatgtctccaaagcct tggagtatggagccttggtgttttctggtggttcaaagccttcgacagtttcgaagtagtgttcgtagacgttaccatcatatctaccaatagcagcattggtcatcatgtca gtcaagttgaaaccatcagtcaaaccaacaacgtttggtctaacctggcacacaaggctgggaaggtctcggaggaaatctttctgctgacatctgggtaatgat accgtactgcaagaaggtggcaccaaagttggtcaagatccagtcggcatacaagaaacccaaggtctcaagacttcagaccattctgggggtgatacgcttgaa gaaaccgtcaatcaagaatctgtgagcgttcaacttggaaaccaaaaccaaacttggaccggtaatgtcgaaaccctgtcttcagcaacagcaccaatgtctctg gccaagccggaaaggggcgtggtcgaaagccttaaccaacttggctgggtcgtccaagtcggcaacgctggagagaaccaatccctttgttctggctcctcaaca agtctctaaccattggcttggcaacgttcatggccaagatgttgttgtcacctccaggttacattggacaacccagtcggagtaggctttaccgaaaccgttgtaaga cgagtaaccgtgaccacccacaagcttgacgagcttcgtcaatggcttcagcagtcaaccaggtacaggtggacttcaaacaacgggaagcgacgaacaatttctt ggattcgacaatggcagcgtcaatggcttcctttcaccagcactgacagcttcgtccaacttgtcggttgcgttcatggtttgttctaagtacaaggcacctttgggag aacaagtaagcggcagccaagaatgggaacaatcttttgatgcaatgggtagtcaatcaacttggtttcaatggtatcgtcttcttcttgaattgtctctgtggat ggcgtatctcaaggcaatggtgatgaatctactggtcattcgtaggagtccatcatcatggtgactctaccaccaatcaaagccgagtaagacaattgttcagatggt ggcatggtgacttcaccagacgggaaacctacagtacttttgcaacatataaagaatcttggaattctgacgttggagaacttggatccaaccgttatcgatacccgtctc tacccatcttggccaccaatgtcaccaacagtcacacctggctcgaggtcgtggttggagtcctcaatgggacacaaaggtcttgacaccgtagtcctttccttgac cttcaatctggcgtagacggagcagtggtggcggagtgcgcagcaccacggatccacccacttggtggcaccaatgtgtgggtgttgatgacaaactcgtcagtg tcttcgtcaaaagtggcggtggtttcaatacctggacgttggaaccgtgggccaactcagtcatcaccaaaacaaccaacaccataggatacctctcaacttgtgatacccttattt atggtccagtagaagaactgggagttggtaccgttaccacggacaggaaacaaacaaaacccaagttgacaccgattctggttccgacttgtgggtcaaagaca ccgaggatcgacaatctctgggcctgttggtctgggtactcgtctcaaagtatctggagagtctggcaatcttcttggcggtgactcttctttgttggtcttggtcaagttgt agtatgaagcgtcgaccttcaagatagggtctttttgcatttgttcgaccatgttgctgatgatctcggatcttctgggaacctccaagaagtagttcaactccttgggtt gaacttggtgagttctcttcttttttgaagttcggtaggcatgatgattgttatagtttctttctttttgaggatgaccagatgtttgataataatcagataggtaggaaaag	112

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1739 of 3077

Table 20

Plasmid	Nucleic acid sequence	SEQ ID NO
	aataaaaaaaaaactaatatatggtaggaaagaagaaaaaaaatggtggatgtgaaagaattgtgaaaaggggggccaatgcctgaagtttatataaagattcaggct	
	ccagagtttgccctgttgattctgcgggggccacacacaagcacactgcagacaggaatagtgggggtagtgttgcatgtaataaaaatgcctgtgttttttttttctgt	
	ttcttctgatgccattgtcgcacgggcacaaacttggctgtctgccccccaggcgtaataaccgaggattatctcacaactacagttgtcactttgccagcaaaatag	
	gaatcaacaagcaatgggttatataacactgcttatacagactactaccaatcccaatacacaccactgagcctactactcacagaaattgaggcgaccacaccacc	
	aattgatcgggaatatgtctccctgattatttttccacacttaattacccataaaacggttacatcaaggataatgaagtgggatatgcaaagtttcaatttttcctctcc	
	gctctaaccaagggattctccatctctatccgtggtccaaacagggccaaacaccacccaagctctactctagaaccttctggtatctacgtgccagcttaattacatc	
	aactcaacgactcttctcccttctgcaggcatgatgatctgccatgccgaactagacgaagccgttcttcaaggtgtattacaactctgtagtgcgatacctcaagcaa	
	gtgatgcggatgaagcaaagggcgaatgtgacgtggtttgaggcgggctactctggtttgtcaggaacgcactaaccaccatccctgctacatccattgcgttattg	
	gttttcgaaatcatgagaactaaactagcagacaacttcgaagactacaatgacgtcctagaagcttagtatgtatgatatagacctatttattacaatatttatctaactg	
	accagttcttcaagtatgccttccatcagcatcctgtatctctttccctcacctcaggtgtctcttcctttccaaacaacctctcccgacaataatcacatcggtacc	
	agtcagaaccacctcatcaacagtcctatactgctggcccaacgcatcgccttatcatccaacccacaccaggcgtcatgatgatccaatcaaacccttctctcta	
	cccccatgtccgctgcggcatgaacccaatcacgaactcccgatcactcttcgcaatctcgatcgtctcacgggtatattcaccatgcgccaacgagccttactc	
	gacaactccgccaacatcaacacgccccttccttttctacccctccgcaccggcgtttcaaccttcaacgacgcccttcccagtcactcctgcggcgttcgtgatgtc	
	actccacgccgcgatccggtatacccccgaggtgtactgcaacatcacggtgtttccgatatcagcaaacttcctgtcctcgaatatcaagaacccgtggcgctctgc	
	aagcacaagcaacggctcaatcgtgccctctagctgaagtctgagatgatatcgatgtgcgtcttcacgagacagatgtggggaccgagcttgtcgatgagcgaaa	
	aggaactcggcggttgtggtcacgtcgatcgaggcgcagaggttggtcttttggactccattaagcggaaaaagggcgaattctgcagatatccatcacactggcgg	
	ccgctcgagcatgcatctagagggccccaattcgccctatagtgagtcgtattacaattcactggccgtcgttttacaacgtcgtgactgggaaaaccctggcgttaccc	
	aacttaatcgccttgcagcacatcccccttccagctggcgtaatagcgaagaggcccgcaccgatcgccctccccaacagttgcgcagcctatacgtacggca	
	gtttaagtttacacctataaaagagagagcgcgttatcgtctgtttgtggatgtacagagtgatattattgacacgccgggggcgacggatggtgatccccctggccagt	
	gcacgtctgctgtcagataaagtctcccgtgaacttaccggtggtgcatatcgggatgaaagctggcgcatgataccaccgatatggccagtgtgccggtctcc	
	gttatcgggagaagtggctgatctcagccaccgcgaaaatgacatcaaaaacgccattaacctgatgttctgggaatataaatgtcaggcatgagattatcaa	
	aaaggatctcactagatccttttcacgtagaaagccagtccgcagaaacggtgctgaccccggatgaatgtcagctactgggctatctggacaagggaaaacg	
	caagcgcaaagagaaagcaggtagcttgcagtgggcttacatggcgatagctagactgggcggtttatggacagcaagcgaaccggaattgccagctgggggc	
	gccctctggtaaggttgggaagccctgcaaagtaaactggatggctttctcgccgccaaggatctgatggcgcagggggatcaagctctgatcaagagacaggatg	
	aggatcgtttcgcatgattgaacaagatggattgcacgcaggttctccggccgcttgggtggagaggctattcggctatgactgggcacaacagacaatcggctgct	
	ctgatgccgccgtgttccggctgtcagcgcaggggcgcccggttctttttgtcaagaccgacctgtccggtgccctgaatgaactgcaagacgaggcagcgcggct	
	atcgtggctggccacgacgggcgttccttgcgcagctgtgctcgacgttgtcactgaagcgggaagggactggctgctattgggcgaagtgccggggcaggatctc	
	ctgtcatctcaccttgctcctgccgagaaagtatccatcatggctgatgcaatgcgggcggctgcataccgcttgatccggctacctgcccattcgaccaccaagcgaaa	
	catcgcatcgagcgagcacgtactcggatggaagccggtcttgtcgatcaggatgatctggacgaagagcatcagggctcgcgccagccgaactgttgccagg	
	gctcaaggcgagcatgccgacggcgaggatctcgtcgtgacccatggcgatgcctgcttgccgaatatcatggtggaaaatggccgcttttctggattcatcgactg	
	tggccggctgggtgtggcggaccgctatcaggacatagcgttggctacccgtgatattgctgaagagcttggcggcgaatgggctgaccgcttcctcgtgctttacggg	
	tatcgccgctcccgattcgcagcgcatcgccttctatcgccttcttgacgagttcttctgaattattaacgcttacaatttcctgatgcggtattttctccttacgcatctgtgcg	
	gtatttcacaccgcatacgtcaaagcaaccatagtacgcgccctgtggttacctgttgttgtgaaaggacagcatcggagatgtgacagacgcgacggcgaggcag	
	gataaatgcttcaataatagcacgtgaggagggccacctgccgcgagcagaaggacccatggctgactaatgcgccgtgtagatcctgatatcgcacgagacccg	
	ggaccgaccggctcgggttctcccgggacttcgtggagacgaccacgtcgtgtggtccgggacgacgtgaccctgttcatcagcgcggtccaggaccaggtgg	

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1740 of 3077

Table 20

Plasmid	Nucleic acid sequence	SEQ ID NO
	tgccggacaacacctggcctgggtgtggtgcgccggcctggacgagctgtacgccgagtggtcggaggtcgtgtccacgaacttccgggacgcctccgggccg gccatgaccgagatcggcgagcagccgtgggggcgggagttcgccctgcgtcgcccgccggcgacccgccgcaactgcgtcactgcgtggccgaggagcaggactgaca cgtgctaaaacttcatttttaatttaaaaggatctaggtgaagatccttttgataatctcatgaccaaaatcccttaacgtgagtttcgttccactgagcgtcagac	
HDE1 promoter from Candida strain ATCC20336	Aagggaagaagtcgttgagttgatgtaattaagctggcacgtagataccagaaggttctagagtagagcttgggtggtgtttggccctgtttggaccacggatag agatggagaatccctggttagaggcggagaggaaaaaattgaaacttgcatatcccactcattatcctgatgtaaccgtttatgggtaattaaagtgtggaaaaaa taatcagggagacatatccgcgatcaattgggtggtggtcgctcaatttctgtgagtagtaggctcagtggtgtgtattgggattggtagtagtctgtataagcagtgttata taacccattgcttgttgattcctatttgctggcaaaagtgacaactgtagttgtgagataatcctcggttattacgcctgggggggcagacagccaaagttgtgcccgtg cgacaatggcatcagaagaagaacagaaaaaaaaaacacaggcatttttatccacatgcacactaccccactattcctgtctgcagtgtgcttgtgtggccccccg cagaatcaacagggcaaactctggagcctgaatctttatataaacttcaggcattggcccccctttcacaattcttacatccaccatttttttcttcctcaccatatta gttttttttatctttcctacctatctgattattatcaaacatctggtcatcctcaaaagaaaagaaagaaactataacaatcaatc	113
PGK promoter from pAA105	Ttgtccaatgtaataatttttccatgactaaaaagtgtgttgttggtgtaaagaagaaagtggaagggacgttggtgatggtgagttcgtctatcccttttatagttgcttgt atagtaggctacttcttctagggactcgatgggggaaggttcttgatatttgcttagttcgagaaggttccagatgagcgagcgaatttttggtacgacattttggtggatgat ctgcacgacatttttgtgattcttgcgacacgctgcactaccaagtgtagtctggctgaacggatcacaagataaacctctgaaaaattatccagggcatgcaacaac aattatacatagaagagggagtcacgatatacacctgtgaaggaatcatgtggtcggctcctgaacttttgaattcatgcaattattaagaagaagcacaggtgag caacccaccatacgttcatttgcaccacctgatgatttaaaagccaaagaaagaaaaaaaaaaaaaagaaacagcggtgggaattgttacaacccacgcgaacc cgaaaatggagcaatcttcccgggggcctccaaataccaactcacccgagagagagaaagagacaccacccaccgagacggagtatatcccaccaaggta agtaactcaggtttaatgatacaggtgtacacagctccttccctagccattgagtgggtatcacatgacactggcaggttacaacccacgtttagtagttatttttgtgcaat tccatgggatcaggaagtttggttggtgggtcgtctactgattcccctttgtctctgaaaatctttttccctagtggaacacttggctgaatgatataaattcaccttgatt cccaccctccctttctctctctctgttacacccaattgaatttctttttttttttttacttccctccttctttatcatcaaagataagtaagtttatcaattgcctattcaga	114
PGK terminator from pAA105	Aaagtacggtgttgttgacaagttgtccacagtttctaccggtggtggtcttctttggaattgttagaaggtaaggaattgccaggtgtgactgctctttccaacaaggcc taaatcagaaatatttattctaataaatagggatgcaaaaaaaaaaaaaccccctttaataaaaaaaaaaaaagaaacgatttttatctaatgaagtctatgtatctaacaaatgt atgtatcaatgtttattccgttaaacaaaaatcagtctgtaaaaaaaggttctaaataaaatattctgtctagtgtacacattcctcccaaaatagtgaaatccagctctacaatt tagctctaccagtaca	115
POX4 terminator from *Candida* strain ATCC20336	Gaatagaagagagtgactcttttgataagagtcgcaaatttgatttcataagtatatattcattatgtaaagtagtaaatggaaaattcattaaaaaaaaaaagcaaatttc cgttgtatgcatactccgaacacaaaactagcccggaaaaaccttagttgatagttgcgaatttaggtcgac	116
POX4 promoter from	Gagctccaattgtaatatttcgggagaaatatcgttgggtaaaacaacagagagagagaggagatggttctggtagaattataatctggttgttgcaaatgct actgatcgactctggcaatgtctgtagctcgctagttgtatgcaacttaggtgttatgcatacacacggttattcggttgaattgtggagtaaaaattgtctgagttgtgtctt	117

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1741 of 3077

Table 20

Plasmid	Nucleic acid sequence	SEQ ID NO
pAA073	agctactggctggccccccgcgaaagataatcaaaattacacttgtgaatttttgcacacacaccgattaacatttccctttttgtccaccgatacacgcttgcctcttctttttttctctgtgcttcccccctcctgtgacttttccaccattgatataaaatcaactccattttcctaaaatcatcccagattctaaaaacaacttctctctgctttccttttttttgttatatttattaccatccctttttttgaatagttattcccactaacattgttcaaatcttcacgacata	
G6Pl promoter from pAA2218	Aaaatcagaggctactccggagatggcacattatcacgtgggcagtcccatctctctgacaatggcactgacaatgccatgtcattctagaagtcgctcgttccattg ctagtcgtgtaggtgtcttactcagaccatatgtttggctgtgtgtggtaaggggtaagaatatctctaggaggaggcaactctactcgaaagtacgaaatggccga cgccaatccgacctgccattgtgcaggacgcgggtcgagatccggaagtgccctcctctgagttgttggtttgttccagttggcataagaaattaataatcgtttagtc aatggtacagtattaataattcctgacaaaaccgtgcctacaagcagtgtgagttccgggggtagcactagtggtcgtggtggtaggcccacagagagtatttcc ggaaccgactgatcctttgattccaatattccaagatttgggtttgtggtgtgtaatgcatgaggttatttttgtcaacctccttcctctactcaggttgttgctcctcctcctct ccttcccttttgttcatttaaatacacacacacacctcccctcccacccacaagaacttttttttttttacttttcttctctctacttttcttcttcttttcattacctgttagattatatt atcaacctcaattggtttattatctaccacccatacaaccccccaccctgccacatccaccaaca	118
GPD promoter from *Candida* strain ATCC20336	Cggaagttgtttaccgacctgaccgtaaatttgctgctgaaagaaaacgtgtcaaacaagaccaattggctcaattgaccctgtggaaatgctttgttgaccaccaatg cttccaccaaacgttacttttttttgcaatcggatggtatgggtctgggggttcacctgttttgtaaagctacagaaggtggcatatttctctgatcaggtgttttttttttcggctg ctgctgctcgtggtggtagtggtgtgtgtgtgtgtgcgtgctggaaggacgcttttgctctgactcctcccaatcagaagttgctatagtggtgaa acaacaatggatgataatgcccgggcggtgcgtgtccgacacaaaccactacattttttagctgggaagcctactgccactcgacccacccacccatggtcaaca aaaaaattctgacaaattataaaataaccctgaattcccccttggaaaaatttttggtatttctctctctctttccttctttttctctccatcaatcaatttgacgttcag taactcaattaattacatcacatccctcaattaaagaatttaatcgat	119
TEF1 promoter from *Candida* strain ATCC20336	Ctagcaaaggcttgatcagagaaagcaacaaaaaaaaaaaaaactctaatactccagaatacactcctttagaaacacacaacaaccaagcctagactaccatggg actacgatgaagacgatttagattacatttctcaaggagaagaaggaagagtttgacgaaaacaagttgaacaacgaagagtacgactgttgcatgacatgcttcc ggagttgaagacaaaattgaaagattacaatgatgagatccacgattacgatttaaaggaagcgttatactacaactatttcgagatagaccctaccattgaagaatt gaagacgaaattcaaaaagagtacgtatatacaactaacatcaacgccttctagtttctgttctccaatgcttctccggtttcttctctgtaccaacaa ggaaaaaaaaaaaaaaatctggcaaaaaaaaccaaaccaaccaatgttcttactcaccaacgccctacaatc	120
PEX11 promoter from *Candida* strain ATCC20336	Gaagatgaagcgtatgagtattatgagtactgtcggacgttggaaggtggcagagttaagcccgagaaaagcaaggaaggagtgggagatgatgagtgatgcgg ccaaagaggatgtgaaggctgcgtatctgttttgatagctggtggtagccgaatagaggaaggcaagcttgttcatattggatgatgtggtagatggtggctgccaa agtggttgtaaatagaaaaaaagtgggttgggtctgttgatagttagtggtggcggctgtctgtgattacgtcagcaagtagcacctcggcagttaaaacagcagcaa cagaaaaaaaaatgtgtgaaagtttgattcccccacagtctaccacacccagagttccatttatccatatatcacaagcaataggaaaaataaaaaattatcaacaa atcacaacgaaaagattctgcaaaattatttttcacttcttcttttgactccttcattttccccttaaacccatacacaacgcagcc	121
pAA2214	cgaattccagcacactggcggccgttactagtggatccgagctcggtaccaagcttgatgcatagcttgagtattctatagtgtcacctaaatagcttggcgtaatcatg gtcatagctgtttcctgtgtgaaattgttatccgctcacaattccacacaacatacgagccggaagcataaagtgtaaagcctgggtgcctaatgagtgagctaactc acattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattggg cgctcttccgcttcctcgctcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatccacagaatcaggg gataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggcccaggaaccgtaaaaaggccgcgttgctggcgttttccataggctccgcccccctgacg agcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttccccctggaagctccctcgtgcgctctcctgttcc	122

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1742 of 3077

Table 20

Plasmid	Nucleic acid sequence	SEQ ID NO
	gaccctgccgcttaccggataacttcgtcgccttctccctcgggaagcgtggcgctttctcatagctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaag	
	ctgggctgtgtgcacgaacccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaaccggtaagacacgacttatcgccactggcagc	
	agccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaaggacagtatttggtatct	
	gcgctctgctgaagccagttaccttcgggaaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggtttttttgtttgcaagcagcagattac	
	gcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacgggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaaa	
	aaggatctcacctagatccttttaaattaaaaatgaagttttagcacgtgtcagtcctgctcctcggccacgaagtgcacgcagttgccggccgggtcgcgcagggc	
	gaactccgccccccacggctgctcgccggcccggaggcgtcccggaagttcgtggacacgacctccgaccactcggcgtacagctcgt	
	ccaggccgcgcacccacaccacccaggccagggtgttgtccggcacccacctggtcctggaccgcgctgatgaacagggtcacgtcgtcccggacccacacccggcga	
	agtcgtcctccacgaagtccgggagaccggcggtcggtccagaactcgaccgctccgggcgagcgtgcgcgcggtgagcaccggaacgcggcactggtca	
	acttggccatggtggcctcctcacgtgctattattgaagcatttatcaggtgattgtgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaataggggt	
	tccgcgcacatttcccgaaaagtgccacctgtatgcggtgtgaaataccgcacacgatgcgtaaggagaaaataccgcatcaggaaattgtaagcgttaataattc	
	agaagaactgtcaagaaggccgatagaaggcgatgcgctgcgaatcgggagcgggcgataccgtaaagcacgaggaagcggtcagcccattcgccgccaag	
	ctcttcagcaatatcacgggtagccaacgctatgtcctgatagcggtccgccacacccagccggccacagtcgatgaatccagaaaagcgggccatttccaccatg	
	atattcggcaagcaggcatcgccatgggtcacgacgagatcctcgccgtcgggcatgctcgccttgagcctggcgaacagttcggctggcgcgagcccctgatgct	
	cttcgtccagatcatcctgatcgacaagaccggcttccatccgagtacgtgctcgctcgatggcgatgtttcgcttggtggtcgaatgggcaggtagccggatcaagcgt	
	atgcagccgccgcattgcatcagccatgatggatacttctcggcaggagcaaggtgagatgacaggagatcctgccccggcgcacttcgcccaatagcagcccagtc	
	ccttccgcttcagtgacaacgtcgagcacagctgcgcaaggaacgcccgtcgtggccagccacgatagccgcgctgcctgctgcttgcagttcattcagggcaccg	
	gacaggtcggtcttgacaaaaagaaccgggcgcccctgcgctgacagccggaacacggcggcatcagagcagccgattgtctgttgtgcccagtcatagccga	
	atagcctctccacccaagcggccgggagaacctgcgtgcaatccatcttgttcaatcatgcgaaacgatcctcatcctgtctcttgatcagagcttgatccctgcgcca	
	tcagatccttggcggcgagaaagccatccagtttacttgcagggcttcccaacctaccagagggcgccccagctggcaattccggttgcttgtctgtccataaaac	
	cgcccagtctagctatcgccatgtaagcccactgcaagctacctgctttctcttgcggtttcccttgtccagatagcccagtagctgacattcatccgggggtcag	
	caccgtttctgcggactggctttctacgtgaaaaggatctaggtgaagatcctttttgataatctcatgcctgacatttatattcccccagaacatcaggttaatggcgtttttg	
	atgtcatttttgcggtggctgagatcagccactcttcccgataacggagaccggcacactggccatatcggtggtcatcatgcgccagctttcatcccgatatgca	
	ccaccgggtaaagttcacgggagacttttatctgacagcagacgtgcactggccagggggatcaccatccgtcgccccggcgtgtcaataatatcactctgtacatc	
	cacaaacagacgataacggctctctcttttataggtgtaaaccttaaactgccgtagtataggctgcgcaactgttgggaagggcgatcggtgcgggcctcttcgct	
	attacgccagctggcgaaaggggggatgtgctgcaaggcgattaagttgggtaacgccagggtttcccagtcacgacgttgtaaaacgacggccagtgaattgtaa	
	tacgactcactatagggcgaattgggcccgacgtcgagagtgctccgaaaatacctcaaggggtggttagagttacatgataaacaacaacaggccaaaaaaggaga	
	gggcgtttctatagaggggttttcttgttcaaatctagggggtgagaggcggaagtcctggcggaacagaagtgagctgggtaatcatacatctcaagtaatg	
	ttgcataccggagagaagaaaaaactggccctcttctctggctgaggagtaaactagaggacggcaggtcgggatagttttttttgactgtgtagctattgcctatgg	
	ccagcaccagcacatcacaaactccagctctggttgacacacctgcccgcgagcatgcggcccatgcagcggaaaagagacctctttgacggaaatatgtctta	
	ggggtgtttttctactgggggttggtaggacttcggaaacacggagtcctcttcttcggaacgcccccctggttgccggccggttcgtcgcttgatcccccctgcagga	
	aaatcagaggctactccggagatggcacattatcacgtgggcagtccatctctctgacaatggcactgacaatgcatgtcattctagaagtgctcgttccattgct	
	agtcgtgtaggtgtcttactcagaccatatgtttggctgtgtgtggtaaggggtaagatatctactagggaggaggcaactctactgaaagtacgaaatggccgac	
	gccaatccgacctgccattgtgcaggacgcgggtcgagatccggaagtgccctccttctgaggttgttggttgttccagttggcataagaaattaataatcgtttagtca	
	atggtacagtattaataattcctgacaaaaccgtgcctacaagcagtgtgagttccgggggtagcactagtggtcgtggtgtagggcccacagagagtatttccg	

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1743 of 3077

Table 20

Plasmid	Nucleic acid sequence	SEQ ID NO
	gaaccgactgatccttgattccaatattccaagatttgggtttgttgtgtaatgcatgaggttatttttgtcaacctccttcctctctacggttgttgctcctcctcctcttc cttccctttgttcatttaaatacacacacacactccctccacccacaagaacttttttttttttttactttcttctcttctacttttcttctttcattacctgttagattatatta tcaacctcaattggtttattatctaccacccatacaaccccccacccctgccacatccaccaacaatggacgacgttgattctgctttagccgataatgttaaatcctcgc tgccggtggttttggtgtatttgtgccgttgaccggtcatccattcggactggtcaaggtcagattgcaaactgggttgtacaaatcgtcagtgcaatgtgtcaaagaa acaatagccaaagacgggttgtttgggttatacagaggtgtctgcctccattgttggtgtcacccccaatgtttgctgttcctctgggggttacgacgttggtagaaaatt ggtttctccttcactggcaaatcagtcgacaagtttgaaatcaaagacatttcgactgccggtttcatcagtgccatcccaacccacctggttgccgctccattcgaaag agtcaaggtcatgatgcagatccaagaaggtgccaagtccaatgggtgccgttgttgctgaaatgtacagaacccggtggtatcagatccatcttcaaggggta ctgtggctacttggccagaagatggtccaggttctgccttgtatttcgctactatgaatgggtcaagaaggaattgactgcccctggtgaagactttgtccttgttgctatta ccactgctggtggtttccgctggtattgccatgtggttgggtgttttcccaattgacaccatcaagtctaccgagtcttccaatgtcgcaggccac taagaacatctacgctaaaggtggtatcaaggcttttctccctggtgttggtcctgcctggcagagcttcccagccaacgctgctacctttctaggggttgaattggc cagaaagttcttggacaaagtatatataagagtgcgactcttttgataaggagtgccgaaatttgaatttcataatgtataattcattgtaaagtagtaaatggaaaaattcattaaa aaaaaaagcaaatttccgttgtatgcatactccgaacacaaaaactagcccgggaaaaaccctagttgatagttgccgaatttaggtcgacctggacgcgtgcggccg ccagtgtgatggatatctgcagaattcgccctcccgaaatattacaattggagctccaataaatattgtaataaatagggtctatatacatacactaagcttctaggacgt cattgtagtcttcgaagttgtctgctagtttagttcctcatgatttcgaaaaccaataacgcaatggatgtagcagggatgtggttagtgcgttcctgacaaacccagagt acgcgcctcaaaccacgtcacattcgccctttgcttcatccgcatcacttgcttgaaggtatccacgtacgagttgtaatacacctgaagaacggcttcgtctcgacg ggtacaacgagaaattgtattgaattgatcaagaacatgatcttggtgttacagaacatcaagttcttggaccagactgagaatgcacagatatacaaggcgtcatgtg ataaaatggatgagatttatccacaattgaagaaagagtttatggaaagtggtcaaccagaaagctaaacaggaagaagcaaacgaagaggtgaaacaagaaag aagaaggtaaataagtatttgtattatataacaacaaacaaagtaaggaatacagatttatacaataaattgccatactagtcacgtgagatatctcatccattccccaact cccaagaaaaaaaaaaaagtgaaaaaaaaaatcaaacccaaagatcaacctcccatcatcatcgtcatcaaacccccagctcaattcgcaatggttagcacaa aaacatacacagaaagaaggcatcagcacaccctccaaggttgcccaacgtttattccgcttaatggagtccaaaaagaccaacctctgcgcctcgatcgacgtga ccacaaccgccgagttccttcgctcatcgacaagctcggtcccacacatctgtctgtgaagacgcacatcgatatcatctcagacttcagctacgagggcacgattg agccgttgcttgtgcttgcagagcgccacgggttcttgatattcgaggacaggaagtttgctgatatcggaaacaccgtgatgttgcagtacacctcgggggtataccg gatcgcgggcgtggagtgacatcacgaacgcgcacggagtgactgggaagggcgtcgttgaagggttgaaacgcggtgcgggaggggtagaaaaggaaagg ggcgtgttgatgttggcggagttgtcgagtaaaggctcgttggcgcatggtgaatataccgtgagacgatcgagaagttgatcgtgaattggg ttcatcgcgcagcgggacatgggggtagagaagaaggtttgattggatcatcatgacgcctggtgtgggttggatgataaaggcgatgcgttgggccagcagt atagggactgttgatgaggtggttctgactggtaccgatgtgattattgtcgggagagggtgtttggaaaggagaccctgaggtggagggaaagagatacag ggatgctggatggaaggcatacttgaagagaactggtcagttagaataaatattgtaataaatggtctatatacatacactaagcttctaggacgtcattgtagtcttc gaagttgtctgctagtttagttcctcatgatttcgaaaaccaatagcaatgggatgtagcagggatggtggttagtgcgttcctgacaaacccagagtacgccgcctca aaccacgtcacattcgcccttgcttcatccgcatcacttgcttgaaggtatccacgtacgagttgtaatacacacactagttggtagttctgggactaccggcagcaa ttccatccgtgcgttatctataatccggatctgtgaaaaaaaaggaaaaaattctatatagaaaagaaattagatagaagcctggcgagtgtgttctagaagacttc cggattcagatgtcgcatggctgcgatatcaaaaaaaaaaaagagagactccaacattacacaaaacgtggtctccgctccatatctgattgcggatttagagtg acactcttccctcgttgattatcaatttcgcaactgtaggatgggctggaggtattgttgatgcatgcaccacttatcattgtcggatccactaccgcgtataatacccac acacacacacacatacacaaacgcttcgtcacatctctatcggtatgtatgggggagagatagaacaggaaagtcatgaggccggccaaggg	
pAA1116	taaacgttgggcaaccttggagggtgtgctgatgcccttctgtgtatgttttgtgctaaccattgcgaattgagctgggggtttgatgacgatgatgatgggggaggttg	123

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1744 of 3077

Table 20

Plasmid	Nucleic acid sequence	SEQ ID NO
	atctttgggtttgatttttttttttcactttttttttttctgggagttgggggaatggatgagatatctcacgtgactagtatggcaatttattgtataaatctgtattcctacttgttgttta	
	tataatacaaaaatacttatttacctctctctctgtttcacctcttcgtttgcttctcctgtttagcttctggttgaccacttccataaactctttctcaattgtggataaatctcatc	
	catttatcacatgacgccttgtatatctgtgcatctcagtctggtccaagaacttgatgttctgtaacaccaagatcatgttctgatcaattcaatacaattctcgttgtacc	
	cgtcgcatatggtcgacctaaattcgcaactatcaactaagggttttttcgggggctagttttgtttggagtatgcatacaacgggaaatttgctttttttttaatgaatttttcca	
	ttactacttacataatgaatatatacttatgaaatcaaatttgcgactcttatcaaaagagtcactctcttctaatctaactggacaagatttcagcagcttctttcggacttttt	
	cacctctttctgacttcaaggtctggacggttcaacatgtcttccaaagcctggagcctggttttctggtgttttctggttgaagcctgacagtttcgaagtagt	
	gttcgtagacgttaccatcatatctaccaatagcagcattggtcatcatgtcagtcaagttgaaaccatcagtcaaaccaacaacgtttggtctaacctggcacacaa	
	ggctgggaagtgctcgggaggaaatctttctgctgacatctgggtaatgataccgtactgcaagaaggtggcaccaaagttggtcaagatccagtcggcatacaag	
	aaaccccaaaggtctcaagacttcagaccattctgggtgatacgcttgaagaaaccgtcaatcaagaaatctgtgagcgttcaacttggaaaccaaaaccaaaacttg	
	gaccggtaatgtcgaaaccctgtctcagcaacagcaccaatgtctctggccaagccggaaaggggcgtggtcgaaagccttaaccaacttggctgggtcgtccaa	
	gtcggcaacgctggagagaaccaatccctttgttctggctccttcaacaagtctctaaccattggcttggtcaaccatgttgttgtcacctccaggt	
	acattggacaacccagtcggagtaggctttaccgaaaccgttgtaagacgagtaaccgtgaccaccacaagcttgacgagcttcgtcaatggcttcagcagtcaac	
	caggtacaggtggacttcaaacaaccgggaagcgacgaacaatttcttggattcgacaatggcagcgtcaatggctcctttcaccagcactgacagcttcgtccaa	
	cttgtcgttggttgcgttcatggtttgttctaagtacaaggcacctgggagaacacaagtaagcggcagccaagaatgggaacaatctctttgatgcaatgggtagtcaa	
	tcaactggtttcaatggtatcggtgtccttcttcgaattgtctctgtggatggcgtatctcaaggcaatggtgatgaatctactggttcattctgtaggagtccatcatcatg	
	gtgactctaccacaatcaaagccgagtaagacaattgttcagatggtggcatggtgacttcacccagacgggaaacttacagtactttgcaacatataaggaatctt	
	gggattctgacgttggagaactggatccaaccgttatcgataccgtctctaccatctggcaccaatgtcaccaacagtcacacctggctcgaggtcgtggttggagt	
	ctctcaatgggacaacaaaggtcttgacaccgtagtcctttccttgacctcaagatcgtcgctagacggagcagtgggtggcggagtgcgcagcaccacggatccac	
	cacttggtggcaccaatgtgtgggggtgttgatgacaaactcgtcagtgtcttcgtcaaaagtggcggtggtttcaatacctggacgttggaacggtgggccaactcag	
	tcataccaaaacaacaaccatagatacctctaacttgtcgatacccttatttatggtccagtagaagaactgggagttggtaccgttaccacggacacaggaaacaaac	
	aaaccccaagttgacaccgattctggttccgacttgtgggtcaaagacaccgaggatcgacaatctctgggctgttggtctgtgctcaaagtatctggag	
	agtctggcaatcttcttggcggtgacttctcttgttgtcttggtctagtatgaagcgtcgacctcaagatagggtctttttgcattgttgctgatg	
	atctcggatcttcttgtgggaacctccaagaagtagttcaactcctttgggttgaacttggttgagttcttcttttttgaagttcgtggtaggcattatgtcgtgaagatttgaaca	
	atgttagtgggaataactattcaaaaaaaaggggatggtaaataaatatacaaaaaaaaaaggaaaagcagaagagaggaagaagttgttttagaatctgggggaga	
	ttttagggaaatggagttgatttatatcatggtggaaaaagtcacaggaggggggaagcacagagaaaaaaaaagaggcaagcgtgtatcggtggacaaa	
	aaagggaaatgttaatcggtgtgtgcaaaaatcacaagtgtaatttgattatcttcgcgggggggccagccagtagctaagacacaactcagacaatttttactcc	
	acaattcaaccgaataaccgtgtgtatgcataacacctaagttgcatacaactagcgagctacagacattgccagagtcgatcagtagcatttgcaacaaccagatt	
	ataattctaccagaaccatctctcctctctctctgttgtttacccaacgatatttctcccgaaatattacaattggagctctgcaggcatgatgatctgccatgccga	
	actagacgaagccgttcttcaaggtgtattacaactcgtagtggatacctcaagcaagtgatgcggatgaagcaaagggcgaatgtgacgtggtttgaggcggc	
	gtactctggtttgtcaggaacgcactaaccacctcctgctacatccattgcgttattggtttcgaaatcatgagaactaaactagcagacaacttcgaagactac	
	aatgacgtcctagaagcttagtgtatgtatatagacctatttattacaatatttattctaactgaccagttctcttcaagtatgcctccatccagcatcctgtatctcttcct	
	ccacctcagggtcttcttccttttcaaacaacccctcctccgacaataatcacatcggtaccagtcagaaccacctcatcaacagtcctatactgctggcccaacgcat	
	cgccttatcatcaaccccacaccaggcgtcatgatgatccaatcaaaccctcttctacccccatgtccgctgcgcgatgaacccaatcacgaactcccgat	
	cactcttcgcaatctcgatcgtctcacgggtatattcaccatgcgccaacggagccttactcgacaactccgccaacatcaacacgcccctttcctttctacccctccg	
	caccgcgtttcaaccctcaacgacgccctccagtcactccgtgcgcggttcgtgatgtcactccacgccgcggatccggtataccccgaggtgtactgcaacatca	

Table 20

Plasmid	Nucleic acid sequence	SEQ ID NO
	cggtgtttccgatatcagcaaacttcctgtcctcgaatatcaagaacccgtggcgctctgcaagcacaagcaacggctcaatcgtgccctcgtagctgaagtctgag	
	atgatatcgatgtgcgtcttcacgagacagatgtgggacgagcttgtcgatgagcgaaaggaactcggcggttgtggtcacgtcgatcgaggcgcagaggttgg	
	tctttttggactccattaagcggaaaaggcgaattctgcagatatccatcacactggcggccgctcgagcatgcatctagaggcccaattcgccctatagtgagtc	
	gtattacaattcactggccgtcgttttacaacgtcgtgactgggaaaaccctggcgttacccaacttaatcgccttgcagcacatccccctttcgccagctggcgtaata	
	gcgaagaggcccgcaccgatcgcccttcccaacagttgcgcagcctatacgtacggcagtttaaggtttacacctataaaagagagagccgttatcgtctgtttgtgg	
	atgtacagagtgatattattgacacgccgggggcgacggatggtgatccccctggccagtgcacgtctgctgtcagataaagtctcccgtgaactttaccccggtggtgc	
	atatcgggggatgaaagctggcgcatgatgaccaccgatatgtgccagttgtccggttctccgttatcgggaagaagtggctgatctcagccaccgcgaaaatgaca	
	tcaaaaaacgccattaacctgatgttctgggggaatataaatgtcaggcatgagattatcaaaaaggatctcacctagatccttttcacgtagaaagccagtccgcaga	
	aacggtgctgacccccggatgaatgtcagctactgggctatctggacaagggaaaacgcaagcgcaaagagaaagcaggtagcttgcagtgggcttacatggcg	
	atagctagactgggcggttttatggacagcaagcgaaccggaattgccagctgggcgccctctggtaaggttgggaagccctgcaaagtaaactggatggctttc	
	tcgccgccaaggatctgatggcgcaggggatcaagctctgatcaagagacaggatgaggatcgtttcgcatgattgaacaagatggattgcacgcaggttctccg	
	gccgcttgggtgggagaggctattcggctatgactgggcacaacagacaatcggctgctctgatgccgccgtgttcggctgcatgcagcgggcgccgggttctttt	
	gtcaagaccgacctgtccggtgccctgaatgaactgcaagacgaggcagcgcggctatcgtggctggccacgacgggcgttccttgcgcagctgtgctcgacgtt	
	gtcactgaagcgggaagggactggctgctattgggcgaagtgccgggggcaggatctcctgtcatctcaccttgctcctgccgcgagaaagtatccatcatggctgatgc	
	aatgcgccggctgcatacgcttgatccggctacctgcccattcgaccaccaagcgaaacatcgcatcgagcgagcacgtactcggatggaagccggtcttgtcga	
	tcaggatgatctggacgaagagcatcagggctcgcgccagccgaactgttcgccaggctcaaggcgcgatgccgacggcgaggatctcgtcgtgaccatcat	
	ggcgatgcctgcttgccgaatatcatggtggaaaatggccgcttttctggattcatcgactgtggccggctgggtgtggcggaccgctatcaggacatagcgttggcta	
	cccgtgatattgctgaagagcttggccgcgaatgggctgaccgcttcctcgtgctttacggtatcgccgctcccgattcgcagcgcatcgccttctatcgccttcttgacg	
	agttcttctgaattattaacgcttacaatttcctgatgcggtatttttctcctacgcatctgtgcggtatttcacaccgcatacaggtggcacttttcgggggaaatgtgcgcgg	
	aaccctatttgttatttttctaaatacattcaaatatgtatccgctcatgagacaataaccctgataaatgcttcaataatagcacgtgaggagggcgcaccagtggccaa	
	gttgaccagtgccgttccggtgctcaccgcgcgcgacgctgccggagcggtcgagttctggaccgaccggctcgggttctcccgggacttcgtggaggacgacttc	
	gccggtgtggtcccgggacgacgtgaccctgttcatcagcgcggtccaggaccaggtggtgccggacacaccctggcctgggtgtgcgcggcctggacga	
	gctgtacgccgagtggtcggaggtcgtgtccacgaacttccgggacgcctccgggccggccatgaccgagatcggcgagcagccgtgggtgcgggagttcgcc	
	ctgcgcgaccccggccggcaactgcgtgcacttcgtggccgaggagcaggactgacacgtgctaaaacttcatttttaatttaaaaggatctaggtgaagatccttttg	
	ataatctcatgaccaaaatcccttaacgtgagttttcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcctttttttctgcgcgtaatctg	
	ctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagatac	
	caaatactgtccttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctgccagtg	
	gcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacggggggttcgtgcacacagcccagcttggagc	
	gaacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttccgaagggagaaaggcggacaggtatccggtaagcggcagg	
	gtcggaacaggagagcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgct	
	cgtcaggggggcggagcctatggaaaaacgccagcaacgcggccttttacggttcctggccttttgctggccttttgctcacatgttctttcctgcgttatcccctgattct	
	gtggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtcagtgagcgaggaagcggaagagcgcccca	
	atacgcaaaccgcctctccccgcgcgttggccgattcattaatgcagctggcacgacaggtttcccgactggaaagcgggcagtgagcgcaacgcaattaatgtg	
	agttagctcactcattaggcaccccaggctttacactttatgcttccggctcgtatgttgtgtggaattgtgagcggataacaatttcacacaggaaacagctatgaccat	
	gattacgccaagctatttaggtgacactatagaatactcaagctatgcatcaagcttggtaccgagctcggatccactagtaacggccgccagtgtgctggaattcgc	

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1746 of 3077

Table 20

Plasmid	Nucleic acid sequence	SEQ ID NO
	cctt	
pAA2311	ttccgcttaatggagtccaaaaaagaccaacctctgcgcctcgatcgacgtgaccacaaccgccgagttccttcgctcatcgacaagctcggtcccccacatctgtctc gtgaagacgcacatcgatatcatctcagacttcagctacgaggcacgatttgagccgttgcttgtgcttgcagagcgcccacgggttcttgatattcgaggacaggaa gtttgctgatatcggaaacaccgtgatgttgcagtacacctcgggggtatacccggatcgcggcgtggagtgacatcacgaacgcgcacggagtgactgggaagg gcgtcgttgaagggttgaaacgcggtgcggagggggtagaaaaggggcgtgttgatgttggcggagttgtcgaagtaaaggctcgttggcgcatggtgaat atacccgtgagacgatcgagattgcgaagagtgatcgggagttcgtgattgggttcatcgcgcagcgggacatgggggtagagaagaaggtttgattggatcat catgacgcctggtgtgggttggatgataaaggcgatcgcgttgggccagcagtataggactgttgatgaggtggttctgactggtaccgatgtgattattgtcgggaga gggttgtttggaaaaggaagagaccctgaggtggagggaaaagagatacagggatgctggatggaaggcatacttgaaagaggtcagttagaataaatatt gtaataaatagggtctatatacatacactaagcttctaggacgtcattgtagtcttcgaagttgtctgctagtttagttccatgatttgtcgaaaaccaatacgcaatggatgt agcagggatggtggttagtgcgttcctgacaaacccagagtacgccgcctcaaaccacgtcacaattcgcccttgcttcatccgcatcacttgcttgaaggtatccacg tacgagttgtaatacaccttgaagaacgggcttcgtctagttcgggcagatcatcatgcctgcagaaaaatcagaggctactccggagatggcacattatcacgtg ggcagtcccatctctctgacaatggcactgacaatgccatgtcattctagaagtcgctcgttccattgctagtcgtgtaggtgtcttactcagaccatatgtttggctgtgtg tgtggtaagggggtaagatatctctaggaggaggcaactctactcgaaagtacgaaatggccgacgccaatccgacctgccattgtgcaggacgcgggtcgaga tccggaagtgccctccttctgaggttgttggttccagttggcataagaaattaataatcgtttagtcaatggtacagtattaatattccctgacaaaaaccgtgccctac aagcagtgtgagttccgggggtagcactagtggtcgtggtgtagggcccacagagagtatttccggaaccgactgatccttgattccaatattccaagatttgggttt gtggttgtgtaatgcatgaggttatttttgtcaacctccttcctctctacggttgttgctcctcctcctcttcccttgttcatttaaatacacacacaccactcccctccc acccacaagaacttttttttttttttacttcttctcttctcttctttcattacctgttagattatatttatcaacctcaatttgtttattatctaccacccatacaaccccca cccctgccacatccaccaacaatggacgacgttgattctgctttagccgataatgttaaatccttcgctgccggtggttttggtgttgattttgtgccgtgttgaccggtcatcc attcgacttggtcaaggtcagattgcaaactgggttgtacaaatcgtcagtgcaatgtgtcaaagaaacaatagccaaagacgggttgtttgggttatacagaggtgt cttgcctccattgttgggtgtcacccaaatgtttgctgtttccttcgggggttacgacgttggtaagaaattggtttcttccttcactggcaaatcagtcgacaagtttgaaatc aaagacatttcgactgccggtttcatcagtgccatcccaacaccacctggttgccgctccattcgaaagagtcaaggtcatgatgcagatccaagaaggtgccaagtc caagtccatgggtgccgttgttgctgaaatgtacagaaaccggtggtatcagatccatctcaaggtactgtggctacttggccagagatggtccaggttctgccttgta tttcgctacttatgaatgggtcaagaaggaattgactgcccctggtgaagacttgtccttgtttgctattaccactgctggtggtttcgctggttattgccatgtggttggtgttt tcccaattgacaccatcaagtctacccagcagtcttccaacgtcaaggtctccattgtgcaggccactaagaaacatctacgctaaaggtggtatcaaggcttcttcttcct ggtgttggtcctgccttggccagagcttccagccaacgctgctacctcctagggttgaattggccagaaagttcttggacaaagttatataataagaatagaaga gagtgactctttgataagagtcgcaaatttgatttcataagtatatattcattgtaaagtagtaaatggaaaattcattaaaaaaaaagcaaatttccgttgtatgcata ctccgaacacaaaactagcccccggaaaaaccttagttgatagttgcgaatttaggtcgaccatatgcgacgggtacaacgagaattgtattgaatttgatcaagaa catgatcttggtgttacagaacatcaagttcttggaccagactgagaatgcacagatatacaaggcgtcatgtgataaaatggatgagatttatccacaattgaagaa agagtttatggaaagtggtcaaccagaagctaaacaggaagaagcaaacgaagaagaaggtaaataagtatttttgtattatataacaa acaaagtaaggaatacagatttatacaataaattgccatactagtcacgtgagatatctcatccattcccaactcccaagaaaaaaaaaaagtgaaaaaaaaaa tcaaacccaaagatcaacctcccccatcatcatcgtcatcaaacccccagctcaattcgcaatggttagcacaaaaacatacacagaaagggcatcagcacaccc ctccaaggttgcccaacgtttaaagggcgaattccagcacactggcggccgttactagtggatccgagctcggtaccaagcttgatgcatagcttgagtattctatagt gtcacctaaatagcttggcgtaatcatggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccacacaacatacgagccggaagcataaagtgtaaagc ctgggtgcctaatgagtgagctaactcacattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgc	124

Table 20

Plasmid	Nucleic acid sequence	SEQ ID NO
	gcgggagaggcggtttgcgtattgggcgctcttccgcttcctcgctcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcg gtaatacggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggc gtttttccataggctccgcccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttccccct ggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcatagctcacgctgtaggtatc tcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggt aagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacgg ctacactagaaggacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcg gtggttttttgtttgcaagcagcagattacgcgcagaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgt taagggatttggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttagcacgtgtcagtcctgctcctcggccacgaagtgcacg cagttgccggccgggtcgcgcagggcgaactccgcccccacgggctgctcgccgatctcggtcatggccggcccggaggcgtccccggaagttcgtggacacga cctccgaccactcggcgtacagctcgtccaggccgcgcaccccacacccaggccagggtgttgtccggcaccgcctgatgtccttggaccaggcgctgatgaacagggtc acgtcgtcccggaccacgtccgcgaagttcgtcctccacgaagtcccgggagaaccccgagccggtcggtccagaactcgaccgctcgggcgacgtcgcgcgcg gtgagcaccggaacggcactggtcaacttggccatggtggccctcctcacgtgctattattgaagccatttatcaggtttattgtctcatgagcggatacatatttgaatgt atttagaaaaataaacaaatagggttccgcgcacatttcccgaaaagtgccacctgtatgcggtgtgaaataccgcacagatgcgtaaggagaaaataccgc atcaggaaattgtaagcgttaatattcagaagaactcgtcaagaaggcgatagaaggcgatgcgctgcgaatcgggagcgggcgataccgtaaagcacgaggg aagcggtcagcccattcgccgccaagctcttcagcaatatcacgggtagccaacgctatgtcctgatagcggtccgccacacccagccggccacagtcgatgaat ccagaaaaggcggccatttccaccatgatattcggcaagcaggcatcgccatgggtcacgacgagatcctcgccgtcgggcatgctcgccttgagcctggcgaac agttcggctggcgcgagccccgatgctcttcgtccagatcatcctgatcgacaagaccggcttccatccgagtacgtgctcgctcgatgtttcgcttggtggtc gaatgggcaggtagccggatcaagcgtatgcagccgccgcattgcatcagccatgatggatacttctcggcaggagcaaggtgagatgacaggagatcctgcc ccggcacttcgcccaatagcagccagtccttccgcttcagtgacaacgtcgagcacagctgcgcaaggaacgcccgtcgtggccagccacgatagccgcgct gcctcgtcttgcagttcattcagggcaccggacaggtcgtcttgacaaaaagaaccgggcgcccctgcgtgacagccggaacacggcggcatcagagcagcc cgattgtctgttgtgcccagtcatagccgaatagcctctccacccaagcggccgggagaacctgcgtgcaatccatcttgttcaatcatgcgaaacgatcctcatcctgt ctcttgatcagagcttgatcccctgcgccatcagatccttggcggcgagaaagccatccagtttacttgcagggcttcccaacctaccagagggcgccccagctgg caattccggttcgctttgctgtccatagaaaccgcccagtctagctatcgccatgtaagcccactgcaagctacctgctttctcttgctttcgctttcgttcctgtccagatag cccagtagctgacattcatccgggggtcagcaccgtttctgcggactggctttctacgtgaaaaggatctaggtgaagatcctttttgataatctcatgcctgacatttatatt ccccagaaacatcaggttaatggcgtttttgatgtcattttcgcggtggctgagatcagccacttcttcccgataacggagaccggcacactggccatatcggtggtcat catgccgccagcttttcatcccgatatgcaccaccgggtaaagttcacgggagacttttatctgacagcagacgtgcactggccagggggatcaccatccgtcgcccc ggcgtgtcaataatatcactctgtacatccacaaacagacgataaccgctcctcttttataggtgtaaaccttaaactgccgtacgtatagggctgcgcaactgttggga agggcgatcggtgcgggcctcttcgctattacgccagctggcgaaaggggatgtgctgcaaggcgattaagttgggtaacgccagggttttcccagtcacgacgttt gtaaaacgacggccagtgaattgtaatacgactcactatagggcgaattgggccctctagatgcatgctcgagcggccgccagtgtgatggatatctgcagaattcg ccctt	125
IGR5-5' genomic DNA contained in	tcatgacttcctgttctatctcccccatacatacccgatagagatgtgacgaagcgtttgtatgtgtgtgtggtattatacgcggttagtggatccgacaat gataagtggtgcatgcatcaacaatacctccagcccatctacagttgcgaaattgataatcaacgagggaagagtgtcactctaaaatccgcaatcagatatggg agcggaagaccacgttttgtgtaatgttggagtctctttttttttgtgatatcgcagccatgcgacatctgaatccggaagtcttctagaaacacactcgccaggcttct	

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1748 of 3077

Table 20

Plasmid	Nucleic acid sequence	SEQ ID NO
pAA2153	atctaatttctttctatatagaatttttcctttttttcacaagatccggattatagataacgcaccggatggaattgctgccggtagtcccagaactaccaactagtgt	
IGR5-3' genomic DNA contained in pAA2153	ggatcaagcgacgaaccggccggcaaccaggggggcgttccgaagaagaggactccgtgtttccgaagtcctaccaaccccccagtaagaaaacaccccctaagacatatttccgtcaaagaggtctctttccgctgcatgggccgcatgctcgcgggcaggtgtgtcaacgagagctggagtttgtgatggtgctggtgctggccataggcaatagctacgacagtcaaaaaaaaactatccgacctgccgtcctctagtttactcctcagccagagaaagaggccaagttttttcttctccggtatgcaacattacttgagatgtatgattacccagctcacttctgttccgccaggacttccgcctctcaaccccccctagatttgaacgaaagaaaaccccctctatgagaaacgccctctcttttttggcctgttgtttatcatgtaactctaaccaccctgaggtatttttcggagcactctcgagcatgg	126
pAA208	tttccataggctccgcccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttcccccctgg aagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcatagctcacgctgtaggtatctc agttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaa gacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggcta cactagaaggacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtg gtttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacgggtctgacgctcagtggaacgaaaactcacgttaa gggattttggtcatgagattatcaaaaggatcttcacctagatccttttaaattaaaaatgaagtttaaatcaatctaaagtatatatgagtaaacttggtctgacagtta ccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtagataactacgatacgggagggcttaccatct ggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatcagcaataaaccagccagccggaaggccgagcgcagaagtggtcctg caacttatccgcctccatcccagtctattaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggt gtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctcc gatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatggcagcactgcataattctcttactgtcatgccatcgtaagatgctttctgtgactggtgagt actcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctc atcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttactt tcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttgaatactcatactcttcctttttcaata ttattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtattagaaaaataaacaaataggggttccgcgcacatttccccgaaaagtgccacc tgacgtctaagaaaccattattatcatgacattaacctataaaaatagggcgtatcacgaggccctttcgtctcgcgcgtttcggtgatgacggtgaaaacctctgacac atgcagctcccggagacggtcacagcttgtctgtaagcggatgccgggagcagacaagcccgtcagggcgcgtcagcgggtgttggcgggtgtcgggctggct taactatgcggcatcagagcagattgtactgagagtgcaccatatgcggtgtgaaatacccgcacagatgcgtaaggagaaaataccgcatcaggcgccattcgcc attcaggctgcgcaactgttgggaagggcgatcggtgcgggcctcttcgctattacgccagctggcgaaaggggatgtgctgcaaggcgattaagttgggtaacg ccagggtttcccagtcacgacgttgtaaaacgacggccagtgaattcgagctcggtacccgggtcgaaagctgggtctagctactggctggcccccgcg aaagataatcaaaattacacttgtgaatttttgcacacacccgattaacatttccctttttgtcaccgatacacgttgcctcttttttttctgtgctccccctcctgtg actttttccaccattgatataaaatcaactccatttcctaaaatctccccagattctaaaaacaacttcttcttcgtttccttttttttgtatatttatttaccatcccttttttt gaatagttattccccactaacattgttcaaatcttcacgacataatgacttttacaaagaaaaaacgttagtgtatcacaaggtcctgacctagatcatccatcaaaag gaaagagacagctccaaatggaacccctcaacaaatgaactacttctggaaggctccgtcgaaagaagtgagttgatgaaggcttggcccaacaaatggaaag agacccaatcttgttcacagacggctcctactacgacttgaccaaggaccaacaaagagaattgaccgccgtcaagatcaacagaatcgccagatacagagaa caagaatccatcgacacttcaacaagagattgtccttgattgtatcttgacccacaggtcggtaccagaattggtgtcaacctcggttgttccttgtatcagagg	127

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1749 of 3077

Table 20

Plasmid	Nucleic acid sequence	SEQ ID NO
	taacggtaccacttcccaattgaactactcgggctaacgaaaaggaaaccgctgacgttaaaggtatctacggttgtttcggtatgaccccgggggatctgacgggta	
	caacgagaattgtattgaattgatcaagaacatgatcttggtgttacagaacatcaagttcttggaccagactgagaatgcacagatatacaaggcgtcatgtgataa	
	aatggatgagatttatccacaattgaagaaaggagtttatggaaagtggtcaaccagaagctaaacaggaagaagcaaacgaagaggtgaaacaagaagaag	
	aaggtaaataagtatttgtattatataacaaacaaagtaaggaatacagatttatacaataaattgccatactagtcacgtgagatatctcatccattcccaactccc	
	aagaaaaaaaaaaaaagtgaaaaaaaaaatcaaacccaagatcaacctcccatcatcatcgtcatcaaacccccagctcaattcgcaatggttagcacaaaaa	
	acatacacagaaaaggcatcagcacacccctccnaggttgcccaacgtttattccgcttaatggagtccaaaaaagaccaacctctgcgcctcgatcgacgtgacc	
	acaaccgccgagttccttcgctcatcgacaagctcggtcccacatctgtctcgtgaagacgcacatcgatntcatctcagactttcagctacgaggcacgattgag	
	ccgttgcttgtgcttgcagagcgcccacgggttcttgatattcgaggacaggaagtttgctgatatcgggaaacaccgtgatgttgcagtacacctcgggggtatacccgga	
	tcgcgggcgtggagtgacatcacgaacgcgcacgggagtgactgggaagggttgaaacgcggtgcggaggggtagaaaaaggaaaaggggg	
	cgtgttgatgtnggcggagttgtcgagtaaaggctcgttggcgcatggtgaatatacccgtgagacgatcgagattgcgaagagtgatcgggagttcgtgattgggtt	
	catcgcgcagcgggacatgggggtagaagaagaggtttgattggatcatcatgacgcctggtgtgggggttggatgataaaggcgatgcgttgggccagcagta	
	taggacgttgatgaggtggttctgactggtaccgatgtgattattgtcgggagagggttgtttggaaaaggaagaagaccctgaggtggaggggaaagagatacagg	
	gatgctggatggaaggcatacttgaagagaactggtcagttagaataaatattgtaataaatagggtctatatacactacaggttctaggacgtcattgtagtcttcg	
	aagttgtctgctagtttagttctcatgatttcgaaaaccaataacgcaatggatgtggttagtgcgttcctgacaaacccagagtacgccgcctcaa	
	accacgtcacattcgcccttgctcatccgcatcacttgcttgaaggtatccacgtacgagttgtaatacacctgaagaacggcttcgtctacggtcgacgacgggta	
	caacgagaattgtattgaattgatcaagaacatgatcttggtgttacagaacatcaagttcttggaccagactgagaatgcacagatatacaaggcgtcatgtgataa	
	aatggatgagatttatccacaattgaagaaaggagtttatggaaagtggtcaaccagaagctaaacaggaagaagcaaacgaagaggtgaaacaagaagaag	
	aaggtaaataagtatttgtattatataacaaacaaagtaaggaatacagatttatacaataaattgccatactagtcacgtgagatatctcatccattcccaactccc	
	aagaaaaaaaaaaaaagtgaaaaaaaaaatcaaacccaagatcaacctcccatcatcatcgtcatcaaacccccagctcaattcgcagagctcggtacccgg	
	ggatcctctagagtcgacctgcaggcatgctaatgtttatatcgttgacggtgaaaaatttaacttactgtaatatttagcattggacttgaagtccaaaatgaagtata	
	aatattggttgaaaattgaaaatgatttacaagccatatacaaatgttccaatcaaattggagatattgtactaaatatccaggcagaaccacacaacaatgac	
	gacaaagtgcaagccaaagagtagaaatcattaatctttcatattgtcatttttttattattattcagtacatatatacatgtgtcaactatatatttcatgaaatatacata	
	cgtcaaataaaacatatccaatcactaaaaaaaaaaaaaaaaatattaaccaaaactaatctaacttaattaaatgttcatatatttcctaccagacatcttctttggtatct	
	atttgaacttccaagatttaaccaattcagcagctttatcttcgtggatatctctacttttggcaacaatcttggcaattgcagcttgaacatctctagccattctactggcat	
	caccacaaacatataaataattgctccactagataataactcatcaacaagagcactatttctaaaatcttatcttagttggatcttgtctgtgagaa	
	ggcattaaacatttcgaaatttctcccaatactgaggcatattcactccattcttgtttgtacaagaaatcttgttcggaatttctacatccataaaacaatacagtcttacc	
	aacattaacaccatttttaacttgttgaactctttctctacaaaaacctctcaatggtgcaacaccggtacctggaccaatcaaaataactgtgggtagtgctatttctttggt	
	aatttgaaattagatcttctaacgtgaactggcaatctgaacttgctaaatttgcctctggaccattcaaatcataatgaaccattgggtttcaccagtttttgtttgttcaat	
	ttcaatattcttcaacaagttggtgacaacaccagtaactggtcttccatcagcttcttcttcggcttcaacaacagcagtaacattaatggtttgctttcacttaatgagga	
	agaagaaatggagtaataacgaggagttaagtgttggacatttttcaataggaattcgaatggaacatcggaccatggtctgttgttggaagcaaataataaagcat	
	cggcaatgttgaatttttctacgggtgacttaccttgcttatcaccaccaattctagtaaatgactttttagtttcttcatcagggcaaatccagcaatagata	
	agaaaaattgtctagaaaacaggacctgaaattccaaatggtgtctaataacagctccataagtgattggaattagggaatggtggaataagtggaatccaaagct	
	ttcaattcaataacagtatcaagtttgtcttctaaccaaaaacatttggcaaattgcttaatgtttcatcagagttagatggccagattgcaagatgatcaccggtggtata	
	tttcaagtttgattcagaaatatcaaattcaacatgaacacaatgtctgtcttcagaagtaaacaattcttagtttaacaattctagccaaaaatggatgagtatgatca	
	aatggaccttagttaagtcaacaccttcagatttaatgtattgacatttggttcacccaaggagacatctggatcattgccagataaaatcatctcttcagtcaatttaac	

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1750 of 3077

Table 20

Plasmid	Nucleic acid sequence	SEQ ID NO
	atttggttcgtatttcaactctttttcttcaaaattcaaatcattcttaaggaatcaaacacgttatcctccaggccaagaaatcttcatctaaagtaccagtaccatcgtca	
	ccttcaccgtattcagcaaatctgtcaccacctttttctcccaacaatctgtcaaatttctaccaatagcattgaagaattcataagttgaattaccccaaaccaaaaaca	
	gtatatttcaaagtactcaaggtgtcagcttcttcagtcaaccaagtgtgaaattcgtcagcattatcggttggttcacctcaccataagtagcaacaataagaaaac	
	caagatatcttcagtaatatctccgaagttttcgaaatcataatcagcgaaatcagcaaccatggttttcaaaccaaatcttgaatgcaattctcttgacaatttgttggca	
	taatcttctgctgtacctgtttgggatccaaatataataacgtatttttattgttctcttcaaagcttgtaagatatctctgaattaccatctccactatcagtattaaggaac	
	ccggtatcttgttggtcaagaaactgattcttgcaaataagctgcaattgcaaccaccaatgttataataacatataaatctaacttatctaatgccatgacttgag	
	cgattgaaacaatagtcttagttgatagagcagtaatggatcaagaaaagaaaagaaataaaaggtagaaataaaaaaaaaacaaagaaaaaattggaaatttg	
	aaatgtttaaaatcttttttttttatttttctcgttcagttcgtttgaagatttgtcgtatgaaatttttttatctttttttttttttgttcctcttctgatgtttcaattacgggaccccaacgtcgttt	
	atgtattacgatactgggaccatactgacttgaaatatgtatcatttctgacagtcaggtaatgtgtaccaactagaccaaatttcatctctatataattattattgtatatata	
	actgtccttctttttcgtctattagagtttttatctcaggaggatttatcggtgcaatttcgatatcggatcatttattcactagaggtgagacacaaaagcagaaatgcta	
	caccgaccacgctaccaccaccattatttttgggggtttactctacgcaacccccccccatcgaagcattgaatagacatgacagctcttccgcttcctcgctcactgac	
	tcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatccacagaatcaggggataacgcaggaaagaacatgga	
	cgctgcgttgaaaagaacgacaccaaggctatctttaagtctattgacgacatgaagtcatttgttgtgactctggttcctgaagtccactgccacttggttgggtgct	
	gaagccattgaccaatgtagacaagcctgtggtggtcacggttactcgtcctacaacggcttcggtaaagcctacaacgattgggttgtccaatgtactttgggaaggt	
	gacaacaatgtcttggccatgagtgttggtaagccaattgtcaagcaagttatcagcattgaagatgcggcaagaccgtcagaggttccaccgctttcttgaaccaa	
	ttgaaggactacactggttccaacagctccaaggttgtttgaacactgttgctgacttggacgacatcaagactgtcatcaaggctattgaagttgccatcatcagattg	
	tcccaagaagctgcttctattgtcaagaaggaatcttcgactatgtcggcgctgaattggttcaactctccaagttgaaggctcaccactacttgttgactgaatacatc	
	agaagaattgacaccttgaccaaaaggactggttccatacttgatcacccgtggtaagttgtacgctgccactattgtcttggacagatttgccggtgtcttcttgacttt	
	caacgttgcctccaccgaagccatcactgctttggcctctgtgcaaattccaagttgtgtgctgaagtcagaccaaacgttgttgcttcacaccgactcctccttccaacaat	
	ccgacatgattgtcaattctgctattggtagatacgatggtgacatctatgagaactacttttgacttggtcaagttgcagaacccaccatccaagaccaaggctcctttac	
	tctgatgcttttggaagccatgttgaacagaccaacctggacgaaagagaaagatttgaaaagtctgatgaaaccgctgctatcttgtccaagaagcttggcggtaatc	
	atggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccacacaacatacgagccggaagcataaagtgtaaagcctggggtgcctaatgagtgagcta	
	actcacattaattgcgttgcgctcactgccgcttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtatt	
	gggcgctcttccgcttcctcgctcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatccacagaatca	
	ggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtt	
pAA918	ggccgcgagcactaggtttgataatttggttcttacagtttatgtatttgattcctcttttttaatatttttattattcttatgatgtaacgacagtcccactataa	128
	ttaacttaaacttgctgtaaatcagatgacaagtgtttccctgtttgcagggagagctctactagtttccttaattcatcttggttatgttgtctgcaccgcaaggataagctg	
	ggcccgaatccaggccacgcatacccaactatcgctgcccggaaaaaatacaaactagtaaactgcaaaaaaccatatataacaatataggaatttctctgtatcat	
	tcttcgaaaaacctcacttctctctctctctcaccctttccctatcatcatcatcatcaccaactcaaaaccctaagcagccaaagtaaagaa	
	cagagaatgattccacaacatcacactatgttcttcgccgttaatttaagaattcactggccgtcgttttacaacgtcgtgactgggaaaaccctggcgttacccaactta	
	atcgccttgcagcacatcccccctttcgccagctggcgtaatagcgaagaggcccgcaccgatcgcccttcccaacagttgcgcagcctgaatggcgaatgggcgcct	
	gatgcggtatttctccttacgcatctgtgcggtatttcacaccgcatatggtgcactctcagtacaatctgctctgatgccgcatagttaagccagccccgacacccgcc	
	aacacccgctgacgcgccctgacgggcttgtctgctcccggcatccgcttacagacaagctgtgaccgtctccgggagctgcatgtgtcagaggttttcaccgtcatc	
	accgaaacgcgcgagacgaaagggcctcgtgatacgcctatttttataggttaatgtcatgataataatggtttcttagacgtcaggtggcactttcgggggaaatgtgc	

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1751 of 3077

Table 20

Plasmid	Nucleic acid sequence	SEQ ID NO
	gcggaaccctatttgttatttttctaaatacattcaaatatgtatccgctcatgagacaataaccctgataaatgcttcaataatattgaaaaaggaagagtatgagtat	
	tcaacatttccgtgtcgcccttattcccttttttgcggcattttgccttcctgttttgctcacccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacg	
	agtgggttacatcgaactggatctcaacagcggtaagatccttgagagttttcgcccccgaaagaacgttttccaatgatgagcacttttaaagttctgctatgtggcgcgg	
	tattatcccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctcagaatgacttggttgagtactcaccagtcacagaaaagcatcttacggatg	
	gcatgacagtaagagaattatgcagtgctgccataaccatgagtgataacactgcggccaacttacttctgacaacgatcggaggaccgaaggagctaaccgcttt	
	ttgcacaacatgggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccatacccaaacgacgagcgtgacaccacgatgcctgtagcaat	
	ggcaacaacgttgcgcaaactattaactggcgaactacttactctagcttcccggcaacaattaatagactggatggaggcggataaagttgcaggaccacttctgc	
	gctcggccttccggctggctggtttattgctgataaatctggagccggtgagcgtgggtctcgcggtatcattgcagcactggggccagatggtaagccctcccgtat	
	cgtagttatctacacgacgggggagtcaggcaactatggatgaacgaaatagacagatcgctgagataggtgcctcactgattaagcattggtaactgtcagaccaa	
	gtttactcatatatactttagattgatttaaaacttcatttttaatttaaaaggatctaggtgaagatcctttttgataatctcatgaccaaaatcccttaacgtgagtttcgttcc	
	actgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtgg	
	tttgtttgccggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagataccaaatactgtccttctagtgtagccgtagttaggccaccact	
	tcaagaactctgtagcaccgccctacatacctcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgata	
	gttaccggataaggcgcagcggtcgggctgaacggggggttcgtgcacacagcccagcttggagcgaacgactacaccgaactgagatacctacagcgtga	
	gctatgagaaaagcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcgaacaggagagcgcacgagggagcttccagg	
	gggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagc	
	aacgcggcctttttacggttcctggccttttgctcacatgttctttcctgcgttatcccctgattctgtggataaccgtattaccgccttgagtgagctgatacc	
	gctcgccgcagccgaacgaccgagcgcagcgagtcagtgagcgaggaagcggaagagcgcccaatacgcaaaccgcctctcccgcgcgttggccgattca	
	ttaatgcagctggcacgacaggtttcccgactggaaagcgggcagtgagcgcaacgcaattaatgtgagttagctcactcattaggcaccccaggctttacacttat	
	gcttccggctcgtatgttgtgtggaattgtgagcggataacaatttcacacaggaaacagctatgaccatgattacgccaagctttaattaaagatatcacaggggt	
	agagaccttggttatgggctgattggaggaagaaatgggtgatttggcacttgacagcgcggagagtggttaacacctggtttcctcatttgggttctgacatttgata	
	agttgaaagaacaatgcagaattcacatggctaatttggcctcggttccacaacgcactcagcattaaaaaaaaaaatacgcaatggcagctcggtcgacgcagca	
	gaagcgccgacgtaccgtcgcgttgcccgcccatgcctcgccgaccctccacgccatcgtttgtcccattgtttgtggtagtgcgccgtgacacacaaaaacttgtcc	
	tgtcacatgctgaagttacaccaacatactactatgggattacgtaatcaaaaatttcacagtttaacaaaaaaaaaaatcatacacatcaacattgggacatctgcc	
	ctccccacaaaaacttgtcttctgcatcaatcatatataaacatcatgaaataagcctaaactcactcttttttttcatccttcctacttcttcttttcatagtaactactttttttttat	
	taccacacttattcatacaccacgctatcggccggccgcatgcggatccgagctcgcggccgccagtgtgatggatatctgcagcggccgctaaatattgtaataa	
	ataggtctatatacatacactaagcttctaggacgtcattgtagtcttcgaagttgtctgctagtttagttctcatgatttcgaaaaccaatatacgcaatggatgtagcagg	
	gatggtggttagtgcgttcctgacaaacccagagtacgccgcctcaaaccacgtcacattcgccctttgcttcatcgcgcatcacttgcttgaaggtatccacgtacgagt	
	tgtaatacaccttgaagaacggcttcgtctcgacgggtacaacgagaattgtattgaattgatcaagaacatgatcttggttgttacagaacatcaagttcttggaccag	
	actgagaatgcacagatatacaaggcgtcatgtgataaaatggatgagatttatccacaattgaagaaaggagtttatggaaagtggtcaaccagaagctaaacag	
	gaagaagcaaacgaagaggtgaaacaagaagaaggtaaatagtatttgtattatataacaaacaaagtaaggaatacagatttatacaataaattgccat	
	acaagtcacgtgagatatctcatccattcccaactccagaaaaaaaaaaaagtgaaaaaaaaaaatcaaacccaagatcaacctcccccatcatcatcgtcat	
	caaacccccagctcaattcgcaatggttagcacacaaaaacatacacagaaaaggcatcagcacacccctccaaggttgcccaacgtttattccgcttaatggagtcc	
	aaaaagaccaacctctgcgcctcgatcgacgtgaccacaaccgccgagttccttcgctcatcgacaagctcggtcccccacatctgtctcgtgaagacgcacatcg	
	atatcatctcagacttcagctacgagggcacgattgagcgttgctgtgcttgcagagcgccacgggttcttgatattcgaggacaggaagttgctgatatcggaaa	

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1752 of 3077

Table 20

Plasmid	Nucleic acid sequence	SEQ ID NO
	caccgtgatgttgcagtacacctcgggggtataccggatcgcggcgtggagtgacatcacgaacgcgcacggagtgactgggaagggcgtcgttgaagggttga aacgcggtgcggagggggtagaaaaggaaagggcgtgttgatgttggcggagttgtcgagtaaaggctcgttggcgcatggtgaatatacccgtgagacgatc gagattgcgaaagagtgatcgggagttcgtgattgggttcatcgcgcagcgggacatgggggggtagagaagaaggtttgattggatcatcatgacgcctggtgtgg ggttggatgataaaggcgatgcgttgggccagcagtataggactgttgatgaggtggttctgactggtaccgatgtgattattgtcgggagaggttgtttggaaaagg aagagaccctgaggtggagggaaagagatacagggatgctggatggaaggcatacttgaagagaactggtcagttagaataaatattgtaataaataggtctata tacatacactaagcttctaggacgtcattgtagtcttgaagttgtctctagtttagttctcatgatttcgaaaaccaataacgcaatggatgtagcagggatggtggtta gtgcgttcctgacaaacccagagtacgccgcctcaaaccacgtcacattcgcccttgcttcatccgcatcacttgcttgaaggtatccacgtacgagttgtaatacac cttgaagaacggcttcgtctgc	

EXAMPLE 28: Nucleic acid sequences of oligonucleotides referenced herein

Table 21

Oligo-nucleotide No.	Nucleic acid sequence	SEQ ID NO
oAA0124	CACACACATATGCGACGGGTACAACGAGAATT	129
oAA0125	CACACAACGCGTAGACGAAGCCGTTCTTCAAG	130
oAA0173	ATGATCTGCCATGCCGAACTC	131
oAA0174	AGCGAGTTCGGCATGGCAGATCATCATG	132
oAA0570	CACACAGCTCTTCTAGAATGGCTTTAGACAAGTTAGA	133
oAA0571	CACACAGCTCTTCGTTTCCAAACATCTTCTTGGTATCTATTTT	134
oAA0208	CACACACTGCAGGAGCTCCAATTGTAATAT	135
oAA0209	CACACATCTAGACCCGGGCTCTTCTTATGTCGTGAAGATTTGAACAAT	136
oAA0216	GAGCCCGGGTCTAGATGTGTGCTCTTCCGAGTGACTCTTTTGATAA	137
oAA0217	CACACACATATGGTCGACCTAAATTCGCAACTATCAA	138
oAA0519	CACACAGCTCTTCCATAATGACTGCACAGGATATTATCGCC	139
oAA0520	CACACAGCTCTTCCCTCTCTTCTATTCCTAATACATCTCAATGTTGGCACCG	140
oAA0572	CACACAGCTCTTCCATAATGGCTTTAGACAAGTTAGATTTGT	141
oAA0573	CACACAGCTCTTCCCTCTCTTCTATTCCTACCAAACATCTTCTTGGT	142
oAA2206	TTCCGCTTAATGGAGTCCAAA	143
oAA2207	ATGATCTGCCATGCCGAACTAGACGAAGCCGTTCTTCAAG	144
oAA2208	CTTGAAGAACGGCTTCGTCTAGTTCGGCATGGCAGATCAT	145
oAA2209	TAAACGTTGGGCAACCTTGG	146
oAA7409	AAAATCAGAGGCTACTCCGGAGATGGCACATTATCACGTG	147
oAA7410	TGTTGGTGGATGTGGCAGGGGTGGGGGTTGTATGGGTGGTA	148
oAA7403	CCGGTTCGTCGCTTGATCCAAAATCAGAGGCTACTCCGGAGATGGCACATTATCACGTG	149
oAA7404	CGTCGTCCATTGTTGGTGGATGTGGCAGGGGTGGGGGTTGTATGGGTGGTA	150
oAA7381	AGAGTCACTCTTATATAACTTTGTCCAAGAACTTTCTGGCC	151
oAA7382	CACCAACAATGGACGACGTTGATTCTG	152
oAA7379	ACTGGCGGCCGCACGCGTCCAAGGTCGACCTAAATTCGCAAC	153
oAA7380	AAGTTATATAAGAGTGACTCTTTTGATAAGAG	154

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1754 of 3077

Table 21		
Oligo-nucleotide No.	Nucleic acid sequence	SEQ ID NO
oAA7377	CTTGGACGCGTGCGGCCG	155
oAA7378	GGATCAAGCGACGAACCG	156
oAA2586	CACACAGCTCTTCCATAATGTTTAACTTTAAGTTGTCGCAACA	157
oAA2587	CACACAGCTCTTCCCTCTCTTCTATTCTTACAACTTAGGCTTAGCATCAGTCA	158
oAA2372	CCACCAAATCCATTATGCCA	159
oAA2373	CCGAAACAACCGTAGATACCTTTAATGGGTTACCCTTGATGACTTCTG	160
oAA2374	CAGAAGTCATCAAGGGTAACCCATTAAAGGTATCTACGGTTGTTTCGG	161
oAA2375	GATTTCTTCGTCGTGGGCAGGGTACCGAGCTCTGCGAATT	162
oAA2376	AATTCGCAGAGCTCGGTACCCTGCCCACGACGAAGAAATC	163
oAA2377	AGTAAGCCAAACCGAAACCG	164
oAA3312	AATTCGCAGAGCTCGGTACCCAACAAGTTGTTGTAGGCACTCAA	165
oAA3313	TTGAGTGCCTACAACAACTTGTTGGGTACCGAGCTCTGCGAATT	166
oAA3314	GACGAAGAATTCCAAAGGCTTGTTAAAGGTATCTACGGTTGTTTCGG	167
oAA3315	CCGAAACAACCGTAGATACCTTTAACAAGCCTTTGGAATTCTTCGTC	168
oAA4719	AAAGAAAGAAAGAAACTATAACAATCAATCATGCCAATTTTGAAAAAACCATTC	169
oAA4720	TCTTATCAAAAGAGTCACTCTCTTCTATTCTTACAACTTAGGCTTAGCATCAG	170
oAA4722	GAATAGAAGAGAGTGACTCTTTTG	171
oAA4723	GATTGATTGTTATAGTTTCTTTCTTTC	172
oAA5319	TTGACTGATGCTAAGTAAGAATAGAAGAGAGTGACTCTTTTG	173
oAA5320	TCTCTTCTATTCTTACTTAGCATCAGTCAATAATCCCTTAG	174
oAA6472	AAAAGAAAGAAAGAAACTATAACAATCAATCAATGTTTAACTTTAAGTTGTCGCAACA	175
oAA6473	TCAAAAGAGTCACTCTCTTCTATTCTTACTTAGCATCAGTCAATAATCCCTTAGTC	176
oAA6367	GAATAGAAGAGAGTGACTCTTTTGATAAGAGTC	177
oAA6368	TGATTGATTGTTATAGTTTCTTTCTTTCTTTTG	178
oAA9946	AAAGAAAGAAAGAAACTATAACAATCAATCATGTCAACTTACCAGTTCCAAG	179
oAA9947	TCTTATCAAAAGAGTCACTCTCTTCTATTCCTAATACTCGCTTAATCTCAAC	180
oAA6669	TGATTGATTGTTATAGTTTCTTTCTTTCTTTTG	181
oAA6771	TCAACTTACCAGTTCCAAGAAACTTTAGAA	182

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1755 of 3077

Table 21		
Oligo-nucleotide No.	Nucleic acid sequence	SEQ ID NO
oAA6772	CAAAAGAAAGAAAGAAACTATAACAATCAATCAATGCTTTCCCGTACCACTTTAAGAGT	183
oAA6773	TTCTAAAGTTTCTTGGAACTGGTAAGTTGATTGGTCGGTCTTGCTGTTGAA	184
oAA6774	CAAAAGAAAGAAAGAAACTATAACAATCAATCAATGTCTGCATTAAGATCATTCCAACG	185
oAA6775	TTCTAAAGTTTCTTGGAACTGGTAAGTTGAATAGGTTCTGACTGAGTTCTTTAAGGTGG	186
oAA6776	CAAAAGAAAGAAAGAAACTATAACAATCAATCAATGTTTAACTTTAAGTTGTCGCAACA	187
oAA6777	TTCTAAAGTTTCTTGGAACTGGTAAGTTGAACCCTTTGCGTGGCTGGT	188
oAA5511	ATGGACGACGTTGATTCTGCTTTAGCCGATAATGTT	189
oAA5512	TTATATAACTTTGTCCAAGAACTTTCTGGCC	190
oAA5553	TTGTTTCTTTGACACATTGC	191
oAA5554	TGCTATTACCACTGCTGG	192
oAA5555	CAATGTGTCAAAGAAACAAGGTACCGAGCTCTGCGAA	193
oAA5556	CAGCAGTGGTAATAGCATTAAAGGTATCTACGGTTGTTTCG	194
oAA5698	GTAACCCCAGAAGGAAAC	195
oAA5699	TTCAAGGGTACTGTGGCTAC	196
oAA5700	GTTTGCTGTTTCCTTCTGGGGTTACGACGGGTACAACGAGAATTGTATTG	197
oAA5701	CCAAAGTAGCCACAGTACCCTTGAATGCGAATTGAGCTGGGGG	198
oAA7624	AGATCATCATGCCTGCAGAAAATCAGAGGCTACTCC	199
oAA7625	GTCACTCTCTTCTATTCTTATATAACTTTGTCCAAGAACTTTC	200
oAA7512	GAATAGAAGAGAGTGACTCTTTTG	201
oAA7265	CTGCAGGCATGATGATCTG	202
oAA6369	AAAAGAAAGAAAGAAACTATAACAATCAATCAATGGCTCCATACCTCAAAAAATTGAGA	203
oAA6370	AGAGTCACTCTCTTCTATTCCTACAACTTGGCCCGTCTACCTTTGTTCAACTTCATGGT	204
oAA0784	CACACAGCTCTTCCATAATGAGATGCCAAGTATCTCAACCATCAAG	205
oAA0785	CACACAGCTCTTCCCTCTCTTCTATTCCTAGTTCAATGATTTAAGGAACTTTTCC	206
oAA0347	CACACACTGCAGTTGTCCAATGTAATAATTTT	207
oAA0348	CACACATCTAGACCCGGGCTCTTCTTCTGAATAGGCAATTGATAAACTTACT	208

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1756 of 3077

Table 21		
Oligo-nucleotide No.	**Nucleic acid sequence**	**SEQ ID NO**
	TATC	
oAA0351	GAGCCCGGGTCTAGATGTGTGCTCTTCCAAAGTACGGTGTTGTTGACA	209
oAA0352	CACACACATATGAATTCTGTACTGGTAGAGCTAAATT	210
oAA5668	GCTGCCAGTGTTGGTGACGACT	211
oAA5669	TTGGACGGCGTTGTATCTGGAA	212
oAA5670	GCTGTTTCTGACTTGTCGTTTGT	213
oAA5671	TGCAGCATGCTTCATAGTGGTC	214
oAA5672	GCTTTTGTTGTTGATTCCAAGAG	215
oAA5673	CAAGTCAGAAACAGATGCAGCA	216
oAA5690	GCTGTTTCTGACTTGGCTTTTGTTGTTGATTCCAAGAGCGAATCCACTTCCAGA	217
oAA5691	TGCAGCATGCTTCATAGTGGTC	218
oAA6145	TTCAAGTTGGCTCAAATCGCAGGTGCCAGATACAACGCCGTCCAATCTGC	219
oAA6146	GCAGATTGGACGGCGTTGTATCTGGCACCTGCGATTTGAGCCAACTTGAA	220
oAA6691	GGTGCTCGCTCATTGGCTGATGGTGGCTTGTTG	221
oAA6692	CAACAAGCCACCATCAGCCAATGAGCGAGCACC	222
oAA7257	CTCTACGCGTTGTACTGGTAGAGCTAAATTGTA	223
oAA7258	CATCATGCCTGCAGGTCGACT	224
oAA7259	TCATGACTTTCCTGTTCTATCTCTCCCC	225
oAA7260	CCATGCTCGAGAGTGCTCCGAAAA	226
oAA6371	AAAAGAAAGAAAGAAACTATAACAATCAATCAATGCCAGAATCAACTCAACAATCTC	227
oAA6372	GACTCTTATCAAAAGAGTCACTCTCTTCTATTCTTATTTCTTGCAGGCATGGACA	228
oAA6470	AAAAGAAAGAAAGAAACTATAACAATCAATCAATGACTACAGACTCAAACACCCACA	229
oAA6471	GACTCTTATCAAAAGAGTCACTCTCTTCTATTCTTACTTTCTACCGAATTGAGCACC	230
oAA3582	ATGCTTACACTAACTTCTGGTCCCAACC	231
oAA3583	GACTATTGTCTCTAGCGGTCTAGATCCT	232
oAA3584	CACATCGAGCCATGATTGAGAAC	233
oAA3585	CGTGAGCAAACAATTGGAACG	234

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1757 of 3077

Table 21		
Oligo-nucleotide No.	Nucleic acid sequence	SEQ ID NO
oAA3586	ATGAGTCACCCAACGCCAGA	235
oAA3587	TCAGAGCTTCGGCGACTCAG	236
oAA3588	ACCATGACGCTGATTGCACC	237
oAA3589	CGAGCTACAATTTCTGCACACCT	238
oAA3590	ATGCCCACATTCAACTACAAGGAC	239
oAA3591	CGCTCTGTTGCAAGTATCGTCA	240
oAA3592	ATGTCAACCACCGGAATCTACAAC	241
oAA3593	CTACAACTTTGGAGTTGTGACCTGCTTC	242
oAA3594	CAGACCCCGAATAATGGACAAA	243
oAA3595	CCTCTCTATAATTTCGGGGTGATCAC	244
oAA3596	GATCATGGAGAGATTACAGGCCG	245
oAA3597	CGTTTTCCCCTACAACTTTGGC	246
oAA6449	AAAGAAAGAAAGAAACTATAACAATCAATCATGAGTCACCCAACGCCAGAAG	247
oAA6450	TCTTATCAAAAGAGTCACTCTCTTCTATTCTTACGACTCAGCTATGTACAACCTC	248
oAA2127	CACACAGCTCTTCCATAATGGTCGCCGATTCTTTAGT	249
oAA2128	CACACAGCTCTTCCCTCTCTTCTATTCTTAAGTGGCCTTCCACAAGT	250
oAA3156	TTGATAGTTGCGAATTTAGGTCGACAACGATATAAACATTAGCATGCCTG	251
oAA3354	CATTATGGTGTTCCGGTTATGAATTCGGTTGAATTGTGGAGTAAAAA	252
oAA3159	CCCGAAATATTACAATTGGAGCTCAAGGCTATCTTTAAGTCTATTGACGACA	253
oAA3356	GTCAGATTTCACGCCTTTACGAATTCATCAGACTTTTCAAATCTTTCTCTTTCG	254
oAA3157	CAGGCATGCTAATGTTTATATCGTTGTCGACCTAAATTCGCAACTATCAA	255
oAA3158	TGTCGTCAATAGACTTAAAGATAGCCTTGAGCTCCAATTGTAATATTTCGGG	256
oAA3355	CATTATGGTGTTCCGGTTAT	257
oAA3357	GTCAGATTTCACGCCTTTAC	258
oAA951	CCTACTTCCACAGCTTTAATCTACTATCAT	259
oAA952	TTTAAGAAAACAACTAAGAGAAGCCAC	260
oAA3557	CATTGAAGAGTGTTTTGCCAGAA	261
oAA3558	CCGAAACAACCGTAGATACCTTTAAGGCAAAAATGCAATCACTCTGT	262

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1758 of 3077

Table 21		
Oligo-nucleotide No.	**Nucleic acid sequence**	**SEQ ID NO**
oAA3559	ACAGAGTGATTGCATTTTTGCCTTAAAGGTATCTACGGTTGTTTCGG	263
oAA3560	CGAAATATTACAATTGGAGCTCGGTACCGAGCTCTGCGAATT	264
oAA3561	AATTCGCAGAGCTCGGTACCGAGCTCCAATTGTAATATTTCG	265
oAA3562	CAGTGGCGGCTTTAACCTTCCGCATATGGTCGACCTAAAT	266
oAA3563	ATTTAGGTCGACCATATGCGGAAGGTTAAAGCCGCCACTG	267
oAA3564	GGGTCCACTCGTCATCCAAC	268
oAA5770	AAAGAAAGAAAGAAACTATAACAATCAATCATGATTGAACAACTCCTAGAATATTG	269
oAA5771	TCTTATCAAAAGAGTCACTCTCTTCTATTCCTAGTCAAACTTGACAATAGC	270
oAA5768	GAATAGAAGAGAGTGACTCTTTTG	271
oAA5769	GATTGATTGTTATAGTTTCTTTCTTTC	272
oAA2125	CACACAGCTCTTCCATAATGGTCAACATATCGAAATT	273
oAA2126	CACACAGCTCTTCCCTCTCTTCTATTCTTATATTTCAAATACTTTAT	274
oAA2159	CACACAGCTCTTCCATAATGACAGTGGAGAATGCAAAA	275
oAA2162	CACACAGCTCTTCCCTCTCTTCTATTCCTACTTCTTGGCGACACTCTTCTTTTCATTCTCCTT	276
oAA2679	GACGGGTTATAACAAGCAGGACA	277
oAA2680	CCGAAACAACCGTAGATACCTTTAATCAAGATCTTCCAGATAATGGTCAA	278
oAA2681	TTGACCATTATCTGGAAGATCTTGATTAAAGGTATCTACGGTTGTTTCGG	279
oAA2682	CAACAAATAACAAGTTCTTGCCCTTGGTACCGAGCTCTGCGAATT	280
oAA2683	AATTCGCAGAGCTCGGTACCAAGGGCAAGAACTTGTTATTTGTTG	281
oAA2684	CTTACTAAGAGATGGACGGTTTGAA	282
oAA2914	AATTGACGGGTTATAACAAGCAGG	283
oAA2915	AATTCGCAGAGCTCGGTACCGGTGATGTACTGGTCTATGTTTTGGAT	284
oAA2916	ATCCAAAACATAGACCAGTACATCACCGGTACCGAGCTCTGCGAATT	285
oAA2917	GGTCATCGTCATCTATATCAGCAATATTTAAAGGTATCTACGGTTGTTTCGG	286
oAA2918	CCGAAACAACCGTAGATACCTTTAAATATTGCTGATATAGATGACGATGACC	287
oAA2919	CACCTGGTACATCAACTCTTCCATA	288
oAA3051	CACACAGCTCTTCCATAATGTCC	289
oAA3052	CACACAGCTCTTCCCTCTCTTCT	290
oAA3053	CACACAGCTCTTCCATAATGTCAG	291

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1759 of 3077

Table 21		
Oligo-nucleotide No.	**Nucleic acid sequence**	**SEQ ID NO**
oAA6658	CACACAGCTCTTCCATCATGTCCGCCAACGAAAACAT	292
oAA6659	CACACAGCTCTTCCAGCTCATGATCTAGTCTTGGCCTTAG	293
oAA6660	CACACAGCTCTTCCATCATGTCAGCGAAATCTATTCACGAAGCCGACGGTAAGGC	294
oAA6661	CACACAGCTCTTCCAGCTCACACGCCTAAAGGAGTAGAAGCTTC	295
oAA540	CACACAGCTCTTCCATAATGCCTACCGAACTTCAAAAAGAAAG	296
oAA541	CACACAGCTCTTCCCTCTCTTCTATTCTTAACTGGACAAGATTTCAGCAGC	297
F98G fwd	GTCTTTGACCCACAAGTCGGAACCAGAATCGGTGTCAAC	298
F98G rev	GTTGACACCGATTCTGGTTCCGACTTGTGGGTCAAAGAC	299
W429F fwd	GTAAAGCCTACTCCGACTTTGTTGTCCAATGTACCTG	300
W429F rev	CAGGTACATTGGACAACAAAGTCGGAGTAGGCTTTAC	301
oAA0179	GAATTCACATGGCTAATTTGGCCTCGGTTCCACAACGCACTCAGCATTAAAAA	302
oAA0182	GAGCTCCCCTGCAAACAGGGAAACACTTGTCATCTGATTT	303
oAA1023	GATATTATTCCACCTTCCCTTCATT	304
oAA1024	CCGTTAAACAAAAATCAGTCTGTAAA	305
oAA2055	TGCCATCCTTGGTAGTCAGTTATT	306
oAA2056	CCGAAACAACCGTAGATACCTTTAATGGCTTGTCCTTGGTGTTGA	307
oAA2060	TGTCGCCATTCAACCAGTAGAT	308
oAA2069	AATTCGCAGAGCTCGGTACCGACTTGTTGAAGATGGACGAGGA	309
oAA2057	TCAACACCAAGGACAAGCCATTAAAGGTATCTACGGTTGTTTCGG	310
oAA2068	TCCTCGTCCATCTTCAACAAGTCGGTACCGAGCTCTGCGAATT	311
oAA2070	TTGATCCACTGTCTTAAGATTGTCAA	312
oAA2075	AGGATTTTGCTGTTGGTGGC	313
oAA2074	AATTCGCAGAGCTCGGTACCATACGTCGGTGAGGTCTGTCG	314
oAA2072	ATTCTCCGCTACTTCGTTCTGGTTAAAGGTATCTACGGTTGTTTCGG	315
oAA2073	CGACAGACCTCACCGACGTATGGTACCGAGCTCTGCGAATT	316
oAA2071	CCGAAACAACCGTAGATACCTTTAACCAGAACGAAGTAGCGGAGAAT	317
oAA2656	AAGCTTTTAATTAAAGATAATCACAGGGGTAGAGACCTTG	318
oAA2657	GGATCCGCATGCGGCCGGCCGATAGCGTGGTATGAATGAATAAGTGTG	319
oAA2658	GGATCCGAGCTCGCGGCCGCGAGCACTAGGTTTTGATAATTTGGTTCTTAC	320

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1760 of 3077

Table 21		
Oligo-nucleotide No.	**Nucleic acid sequence**	**SEQ ID NO**
oAA2659	GAATTCTTAATTAACGGCGAAGAACATAGTGTGATG	321

EXAMPLE 29: Transformation procedure

5 mL YPD start cultures were inoculated with a single colony of Candida strain ATCC20913 and incubated overnight at 30°C, with shaking at about 200 rpm. The following day, fresh 25 mL YPD cultures were inoculated to an initial OD600 nm of 0.4 and the culture incubated at 30°C, with shaking at about 200 rpm until an OD600 nm of 1.0-2.0 was reached. Cells were pelleted by centrifugation at 1,000 x g, 4°C for 10 minutes. Cells were washed by resuspending in 10 mL sterile water, pelleted, resuspended in 1 mL sterile water and transferred to a 1.5 mL microcentrifuge tube. The cells were then washed in 1 mL sterile TE/LiOAC solution, pH 7.5, pelleted, resuspended in 0.25 mL TE/LiOAC solution and incubated with shaking at 30°C for 30 minutes.

The cell solution was divided into 50 µL aliquots in 1.5 mL tubes to which was added 5-8 µg of linearized DNA and 5 µL of carrier DNA (boiled and cooled salmon sperm DNA, 10 mg/mL). 300 µL of sterile PEG solution (40% PEG 3500, 1X TE, 1X LiOAC) was added, mixed thoroughly and incubated at 30°C for 60 minutes with gentle mixing every 15 minutes. 40 µL of DMSO was added, mixed thoroughly and the cell solution was incubated at 42°C for 15 minutes. Cells were then pelleted by centrifugation at 1,000 x g 30 seconds, resuspended in 500 µL of YPD media and incubated at 30°C with shaking at about 200 rpm for 2 hours. Cells were then pelleted by centrifugation and resuspended in 1mL 1X TE, cells were pelleted again, resuspended in 0.2 mL 1X TE and plated on selective media. Plates were incubated at 30°C for growth of transformants.

EXAMPLE 30: Construction of strain sAA7445 for the production of lycopene from glucose

In order to integrate genes into the genome of *C. viswanathii*, genes were cloned into a plasmid that allows for single crossover recombination at the URA3 loci. In addition, since glucose was the primary carbon source, all genes were expressed under a GPD promoter with a POX4

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1761 of 3077

terminator. The backbone for all plasmids was PCR amplified from pAA2534 with primers oAA07511 and oAA07512. The CrtE gene from *Cronobacter sakazakii* (CsCrtE) was codon optimized for *C. viswanathii*, synthesized, and then PCR amplified using primers oAA9745 and oAA9746. The CsCrtE PCR product and backbone contained overlapping sequence at their termini allowing directional ligation, generating plasmid pAA2698. The CrtB gene from *Cronobacter sakazakii* (CsCrtB) was codon optimized for *C. viswanathii*, synthesized, and then PCR amplified using primers oAA9749 and oAA9750. The CsCrtB PCR product and backbone contained overlapping sequence at their termini allowing directional ligation, generating plasmid pAA2699. The CrtI gene from *Cronobacter sakazakii* (CsCrtI) was codon optimized for *C. viswanathii*, synthesized, and then PCR amplified using primers oAA9747 and oAA9748. The CsCrtI PCR product and backbone contained overlapping sequence at their termini allowing directional ligation, generating plasmid pAA2700. Integration cassettes containing glucose inducible CsCrtE, CsCrtB, CsCrtI genes were then generated by PCR amplification of plasmids pAA2698, pAA2699, and pAA2700 with primers oAA2206 and oAA2209. All three cassettes were purified and chemically transformed into strain sAA002, which was then plated onto SCD-ura plates. Yellow/red colored colonies were streaked onto YPD for isolation and characterization. Colony PCR was performed to confirm the presence of all three genes, and one verified isolate was saved as strain sAA7445.

EXAMPLE 31: Shake flask characterization of sAA7445 on glucose

Starter cultures (5 mL) of sAA7445 in YPD were incubated overnight at 30°C, with shaking at approximately 250 rpm. The overnight cultures were used to inoculate 25 mL fresh SP92-glucose media (6.7 g/L yeast nitrogen base, 3.0 g/L yeast extract, 3.0 g/L $(NH_4)_2SO_4$, 1.0 g/L K_2HPO_4, 1.0 g/L KH_2PO_4, 75 g/L glucose) to an initial OD600 nm of 0.4 and incubated approximately 24 hours at 30°C, and 300 rpm shaking. Cells were pelleted by centrifugation for 10 minutes at 3,000xg, at 4°C, and then resuspended in 12.5 mL HiP-TAB media (yeast nitrogen base without amino acids and without ammonium sulfate, 1.7 g/L; yeast extract, 3.0 g/L; potassium phosphate monobasic, 10.0 g/L; potassium phosphate dibasic, 10.0 g/L) and added to 250 mL baffled shake flasks. 500 µL of glucose was added to the shake flasks, which were shaken at approximately 300 rpm, at 30°C. Cultures were incubated for 48 hours and samples were taken at 48 hours for analysis of carotenoid production by HPLC, and results are presented in Table 22 below.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1762 of 3077

Table 22				
Strain	**Substrate**	**Lycopene (g/L)**	**B-carotene (g/L)**	**Astaxanthin (g/L)**
sAA7445	Glucose	0.02		
sAA7882	Glucose	0.003	0.007	
sAA7870	Glucose	0.0015	0.001	0.0015
sAA7446	Oleic Acid	0.001		
sAA7565	Oleic Acid	0.011		
sAA8283	Oleic Acid	0.061		
sAA8283	C10-C14 alkanes	0.008		
sAA8283	Methyl pentadecanoate	0.012		
sAA8283	Ethyl laurate	0.022		
sAA8283	CPO	0.030		
sAA8283	Ethyl PFAD	0.051		
sAA8283	Nonane	0.002		
sAA8283	Octadecane	0.0125		
sAA8519	Oleic Acid	0.002	0.007	
sAA8503	Glucose	0.035		
sAA8932	Glucose	0.064		
sAA8932	Oleic Acid	0.027		
sAA9432	Oleic acid	0.061		
sAA9670	Oleic acid	0.167		
sAA9811	Oleic acid	0.156		
sAA9812	Oleic acid	0.214		
sAA9814	Oleic acid	0.161		
sAA9817	Oleic acid	0.277		
sAA9819	Oleic acid	0.382		
sAA9821	Oleic acid	0.447		

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1763 of 3077

EXAMPLE 32: Construction of strain sAA7446 for the production of lycopene from oleic acid

In order to integrate genes into the genome of *C. viswanathii*, genes were cloned into a plasmid that allows for single crossover recombination at the URA3 loci. In addition, since oleic acid was the primary carbon source, all genes were expressed under a HDE1 promoter with a POX4 terminator. The backbone for all plasmids was PCR amplified from pAA1164 with primers oAA5788 and oAA5789. The CrtE gene from *Cronobacter sakazakii* (CsCrtE) was codon optimized for *C. viswanathii*, synthesized, and then PCR amplified using primers oAA9753 and oAA9754. The CsCrtE PCR product and backbone contained overlapping sequence at their termini allowing directional ligation, generating plasmid pAA2702. The CrtB gene from *Cronobacter sakazakii* (CsCrtB) was codon optimized for *C. viswanathii*, synthesized, and then PCR amplified using primers oAA9757 and oAA9758. The CsCrtB PCR product and backbone contained overlapping sequence at their termini allowing directional ligation, generating plasmid pAA2703. The CrtI gene from *Cronobacter sakazakii* (CsCrtI) was codon optimized for *C. viswanathii*, synthesized, and then PCR amplified using primers oAA9755 and oAA9756. The CsCrtI PCR product and backbone contained overlapping sequence at their termini allowing directional ligation, generating plasmid pAA2704. Integration cassettes containing oleic acid inducible CsCrtE, CsCrtB, CsCrtI genes were then generated by PCR amplification of plasmids pAA2702, pAA2703, and pAA2704 with primers oAA2206 and oAA2209. All three cassettes were purified and chemically transformed into strain sAA002; the cells were plated onto SCD-ura plates. Yellow colored colonies were streaked onto YPD for isolation and characterization. Colony PCR was performed to confirm the presence of all three genes, and one verified isolate was saved as strain sAA7446.

EXAMPLE 33: Shake flask characterization of sAA7446 on oleic acid

Starter cultures (5 mL) of sAA7446 in YPD were incubated overnight at 30°C, with shaking at approximately 250 rpm. The overnight cultures were used to inoculate 25 mL fresh SP92-glycerol media (6.7g/L yeast nitrogen base, 3.0 g/L yeast extract, 3.0 g/L $(NH_4)_2SO_4$, 1.0 g/L K_2HPO_4, 1.0 g/L KH_2PO_4, 75 g/L glycerol) to an initial OD600 nm of 0.4 and incubated approximately 24 hours at 30°C, and 300 rpm shaking. Cells were pelleted by centrifugation for 10 minutes at 3,000xg, at 4°C, and then resuspended in 12.5 mL HiP-TAB media (yeast nitrogen base without amino acids and without ammonium sulfate, 1.7 g/L; yeast extract, 3.0 g/L; potassium phosphate monobasic, 10.0 g/L; potassium phosphate dibasic, 10.0 g/L) and

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1764 of 3077

added to 250 mL baffled shake flasks. 500 µL of oleic acid was added to the shake flasks, which were shaken at approximately 300 rpm, at 30°C. Cultures were incubated for 48 hours and samples were taken at 48 hours for analysis of carotenoid production by HPLC, and results are presented in Table 22 above.

EXAMPLE 34: Construction of strain sAA7449 for the production of valencene from glucose

In order to integrate genes into the genome of *C. viswanathii*, genes were cloned into a plasmid that allows for single crossover recombination at the URA3 loci. In addition, since glucose was the primary carbon source, all genes will be expressed under a GPD promoter with a POX4 terminator. The TPS1 Gene from *Callitropsis nootkatensis* (CnTPS1) was codon optimized for *C. viswanathii*, synthesized, and then PCR amplified using primers oAA9751 and oAA9752. In addition, the plasmid backbone was PCR amplified from pAA2534 using primers oAA07511 and oAA07512. The CnTPS1 PCR product and backbone contained overlapping sequence at their termini allowing directional ligation, generating plasmid pAA2701. An integration cassette for CnTPS1 was then generated by PCR amplification of the pAA2701 plasmid with primers oAA2206 and oAA2209. The PCR product was purified and chemically transformed into strain sAA002, and plated onto SCD-ura plates. Colonies were streaked onto YPD for isolation and characterization. Colony PCR was performed to confirm the presence of the CnTPS1 gene, and one verified isolate was saved as strain sAA7449.

EXAMPLE 35: Shake flask characterization of sAA7449 on glucose

Starter cultures (5 mL) of sAA7449 in YPD were incubated overnight at 30°C, with shaking at approximately 250 rpm. The overnight cultures were used to inoculate 25 mL fresh SP92-glucose media (6.7 g/L yeast nitrogen base, 3.0 g/L yeast extract, 3.0 g/L $(NH_4)_2SO_4$, 1.0 g/L K_2HPO_4, 1.0 g/L KH_2PO_4, 75 g/L glucose) to an initial OD600 nm of 0.4 and incubated approximately 24 hours at 30°C, and 300 rpm shaking. Cells were pelleted by centrifugation for 10 minutes at 3,000xg, at 4°C, and then resuspended in 12.5 mL HiP-TAB media (yeast nitrogen base without amino acids and without ammonium sulfate, 1.7 g/L; yeast extract, 3.0 g/L; potassium phosphate monobasic, 10.0 g/L; potassium phosphate dibasic, 10.0 g/L) and added to 250 mL baffled shake flasks. 500 µL of glucose was added to the shake flasks, in addition to 1.25 mL of n-Decane. The flasks were then incubated at 30°C and shaken at approximately 300 rpm. Incubation of the cultures continued for 48 hours and samples were

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1765 of 3077

taken from the organic layer at 48 hours for analysis of valencene production by GCMS. A Shimadzu GCMS-QP2010SE equipped with AOC-20i auto injector with a 30 m x 0.32 mm x 0.25 µm Phenomenex ZB-5HT column was used. Retention time and identity of valencene was determined by injection of a standard from Sigma-Aldrich and confirmed by NIST MS Search 2.0 library match. Positive valencene containing samples were matched by their retention times and their mass spectra (Fig 48A, Fig. 48B).

EXAMPLE 36: Construction of strain sAA7453 for the production of valencene from oleic acid

In order to integrate genes into the genome of *C. viswanathii*, genes were cloned into a plasmid that allows for single crossover recombination at the URA3 loci. In addition, since oleic acid was the primary carbon source, all genes will be expressed under a HDE1 promoter with a POX4 terminator. The TPS1 Gene from *Callitropsis nootkatensis* (CnTPS1) was codon optimized for *C. viswanathii*, synthesized, and then PCR amplified using primers oAA9759 and oAA9760. In addition, the plasmid backbone was PCR amplified from pAA1164 using primers oAA05788 and oAA05789. The CnTPS1 PCR product and backbone contained overlapping sequence at their termini allowing directional ligation, generating plasmid pAA2705. An integration cassette for CnTPS1 was then generated by PCR amplification of the pAA2705 plasmid with primers oAA2206 and oAA2209. The PCR product was purified and chemically transformed into strain sAA002, and plated onto SCD-ura plates. Colonies were streaked onto YPD for isolation and characterization. Colony PCR was performed to confirm the presence of the CnTPS1 gene, and one verified isolate was saved as strain sAA7453.

EXAMPLE 37: Shake flask characterization of sAA7453 on oleic acid

Starter cultures (5 mL) of sAA7453 in YPD were incubated overnight at 30°C, with shaking at approximately 250 rpm. The overnight cultures were used to inoculate 25 mL fresh SP92-glucose media (6.7 g/L yeast nitrogen base, 3.0 g/L yeast extract, 3.0 g/L $(NH_4)_2SO_4$, 1.0 g/L K_2HPO_4, 1.0 g/L KH_2PO_4, 75 g/L glucose) to an initial OD600 nm of 0.4 and incubated approximately 24 hours at 30°C, and 300 rpm shaking. Cells were pelleted by centrifugation for 10 minutes at 3,000xg, at 4°C, and then resuspended in 12.5 mL HiP-TAB media (yeast nitrogen base without amino acids and without ammonium sulfate, 1.7 g/L; yeast extract, 3.0 g/L; potassium phosphate monobasic, 10.0 g/L; potassium phosphate dibasic, 10.0 g/L) and added to 250 mL baffled shake flasks. 500 µL of oleic acid was added to the shake flasks, in

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1766 of 3077

addition to 1.25 mL of n-Decane. The flasks were then incubated at 30°C and shaken at approximately 300 rpm. Incubation of the cultures continued for 48 hours and samples were taken from the organic layer at 48 hours for analysis of valencene production by GCMS. A Shimadzu GCMS-QP2010SE equipped with AOC-20i auto injector with a 30 m x 0.32 mm x 0.25 µm Phenomenex ZB-5HT column was used. Retention time and identity of valencene was determined by injection of a standard from Sigma-Aldrich and confirmed by NIST MS Search 2.0 library match. Positive valencene containing samples were matched by their retention times and their mass spectra (Fig 49A, Fig. 49B).

EXAMPLE 38: Construction of strain sAA6234

Strain sAA5733 (described in International Patent Application Publication No. WO2016/154046) was placed on 5'FOA for the loop-out of URA3, leaving a URA3 terminator scar. One correct transformant was identified and saved as strain sAA5761. In order to knock out the final copy of CRC1 from the genome, plasmid pAA1701 with 253 bp and 375 bp homology region to the 5' and 3' respectively within the open reading frame of CRC1 in the genome was constructed with a URA3 selection cassette. Plasmid pAA1701 construction details are provided in Example 6 and Fig. 32. CRC1 knockout cassette was amplified from pAA1701 using primers oAA05511 and oAA5512 and the resulting fragment was integrated into the sAA5761 genome with URA selection. Correct crc1/crc1::URA3 transformants were identified and plated onto 5'FOA for the loop-out of URA3, leaving a URA3 promoter scar. One correct transformant was identified and saved as strain sAA6234.

EXAMPLE 39: Construction of strain sAA7565 for the production of lycopene from oleic acid

Plasmid pAA2311 was constructed by PCR amplification of pG6PI-CRC1 fragment from pAA2214 using primers oAA7624 and oAA7625, and PCR amplification of the plasmid backbone from pAA1116 using primers oAA7512 and oAA7265 (in certain instances, PCR amplification of the plasmid backbone was from pAA1164). The two PCR products encoded overlapping sequence at their termini allowing directional ligation, generating plasmid pAA2311. Integration cassettes for CsCrtE, CsCrtB, CsCrtI, and pG6PI-CRC1 were then generated by PCR amplification of plasmids pAA2702, pAA2703, pAA2704, and pAA2311 with primers oAA2206 and oAA2209. All four cassettes were purified and chemically transformed into strain sAA6234, which had both copies of CRC1 deleted. The transformed cells were then plated onto

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1767 of 3077

SCD-ura plates. Colonies were streaked onto YPD for isolation and characterization. Colony PCR was performed to confirm the presence of all four genes, and one verified isolate was saved as strain sAA7565.

EXAMPLE 40: Shake flask characterization of sAA7565 on oleic acid

Starter cultures (5 mL) of sAA7565 in YPD were incubated overnight at 30°C, with shaking at approximately 250 rpm. The overnight cultures were used to inoculate 25 mL fresh SP92-glycerol media (6.7 g/L yeast nitrogen base, 3.0 g/L yeast extract, 3.0 g/L $(NH_4)_2SO_4$, 1.0 g/L K_2HPO_4, 1.0 g/L KH_2PO_4, 75 g/L glycerol) to an initial OD600 nm of 0.4 and incubated approximately 24 hours at 30°C, and 300 rpm shaking. Cells were pelleted by centrifugation for 10 minutes at 3,000xg, at 4°C, and then resuspended in 12.5 mL HiP-TAB media (yeast nitrogen base without amino acids and without ammonium sulfate, 1.7 g/L; yeast extract, 3.0 g/L; potassium phosphate monobasic, 10.0 g/L; potassium phosphate dibasic, 10.0 g/L) and added to 250 mL baffled shake flasks. 500 µL of oleic acid was added to the shake flasks, which were shaken at approximately 300 rpm, at 30°C. Cultures were incubated for 48 hours and samples were taken at 48 hours for analysis of carotenoid production by HPLC, and results are presented in Table 22 above.

EXAMPLE 41: Construction of strain sAA7882 for the production of β-carotene from glucose

In order to integrate genes into the genome of *C. viswanathii*, genes were cloned into a plasmid that allows for single crossover recombination at the URA3 loci. In addition, since glucose was the primary carbon source, all genes were expressed under a GPD promoter with a POX4 terminator. The CrtYB Gene from *Xanthophyllomyces dendrorho* (XdCrtYB) was codon optimized for *C. viswanathii*, and then synthesized with additional flanking regions. The XdCrtYB gene and flanking regions were PCR amplified using primers oAA9958 and oAA9959, while the plasmid backbone was PCR amplified from pAA2534 with primers oAA07511 and oAA07512. The two PCR products encoded overlapping sequence at their termini allowing directional ligation, generating plasmid pAA2866. Integration cassettes for CsCrtE, CsCrtI, and XdCrtYB were then generated by PCR amplification of plasmids pAA2698, pAA2700, and pAA2866 with primers oAA2206 and oAA2209. All three cassettes were purified and chemically transformed into strain sAA002, which was then plated onto SCD-ura plates. Yellow colored colonies were streaked onto YPD for isolation and characterization. Colony PCR was

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1768 of 3077

performed to confirm the presence of all three genes, and one verified isolate was saved as strain sAA7882.

EXAMPLE 42: Shake flask characterization of sAA7882 on glucose

Starter cultures (5 mL) of sAA7882 in YPD were incubated overnight at 30°C, with shaking at approximately 250 rpm. The overnight cultures were used to inoculate 25 mL fresh SP92-glucose media (6.7 g/L yeast nitrogen base, 3.0 g/L yeast extract, 3.0 g/L $(NH_4)_2SO_4$, 1.0 g/L K_2HPO_4, 1.0 g/L KH_2PO_4, 75 g/L glucose) to an initial OD600 nm of 0.4 and incubated approximately 24 hours at 30°C, and 300 rpm shaking. Cells were pelleted by centrifugation for 10 minutes at 3,000xg, at 4°C, and then resuspended in 12.5 mL HiP-TAB media (yeast nitrogen base without amino acids and without ammonium sulfate, 1.7 g/L; yeast extract, 3.0 g/L; potassium phosphate monobasic, 10.0 g/L; potassium phosphate dibasic, 10.0 g/L) and added to 250 mL baffled shake flasks. 500 µL of glucose was added to the shake flasks, which were shaken at approximately 300 rpm, at 30°C. Incubation of the cultures continued for 48 hours and samples were taken at 48 hours for analysis of carotenoid production by HPLC, and results are presented in Table 22 above.

EXAMPLE 43: Construction of strain sAA7870 for the production of astaxanthin from glucose

In order to integrate genes into the genome of *C. viswanathii*, genes were cloned into a plasmid that allows for single crossover recombination at the URA3 loci. In addition, since glucose was the primary carbon source, all genes were expressed under a GPD promoter with a POX4 terminator. The backbone for all plasmids was amplified from pAA2534 with primers oAA07511 and oAA07512. The CrtR Gene from *Xanthophyllomyces dendrorhous* (XdCrtR) was codon optimized for *C. viswanathii*, and then synthesized with additional flanking regions. The XdCrtR gene and flanking regions were amplified using primers oAA9960 and oAA9961. The XdCrtR PCR product and the pAA2534 backbone PCR encoded overlapping sequence at their termini allowing directional ligation, generating plasmid pAA2993. The CrtS Gene from *Xanthophyllomyces dendrorho* (XdCrtS) was codon optimized for *C. viswanathii*, and then synthesized with additional flanking regions. The XdCrtS gene and flanking regions were amplified using primers oAA9962 and oAA9963. The XdCrtS PCR product and the pAA2534 backbone PCR encoded overlapping sequence at their termini allowing directional ligation, generating plasmid pAA2994. The CrtW Gene from *Agrobacterium aurantiacum* (AaCrtW) was

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1769 of 3077

codon optimized for *C. viswanathii*, and then synthesized with additional flanking regions. The AaCrtW gene and flanking regions were amplified using primers oAA9964 and oAA9965. The AaCrtW PCR product and the pAA2534 backbone PCR encoded overlapping sequence at their termini allowing directional ligation, generating plasmid pAA2995. The CrtZ Gene from *Agrobacterium aurantiacum* (AaCrtZ) was codon optimized for *C. viswanathii*, and then synthesized with additional flanking regions. The AaCrtZ gene and flanking regions were amplified using primers oAA9966 and oAA9967. The AaCrtZ PCR product and the pAA2534 backbone PCR encoded overlapping sequence at their termini allowing directional ligation, generating plasmid pAA2996. Integration cassettes for CsCrtE, CsCrtI, XdCrtYB, XdCrtR, XdCrtS, AaCrtW, and AaCrtZ were then generated by PCR amplification of plasmids pAA2698, pAA2700, pAA2866, pAA2993, pAA2994, pAA2995, and pAA2996 with primers oAA2206 and oAA2209. All seven cassettes were purified and chemically transformed into strain sAA002, which was then plated onto SCD-ura plates. Yellow colored colonies were streaked onto YPD for isolation and characterization. Colony PCR was performed to confirm the presence of all seven genes, and one verified isolate was saved as strain sAA7870.

EXAMPLE 44: Shake flask characterization of sAA7870 on glucose

Starter cultures (5 mL) of sAA7870 in YPD were incubated overnight at 30°C, with shaking at approximately 250 rpm. The overnight cultures were used to inoculate 25 mL fresh SP92-glucose media (6.7 g/L yeast nitrogen base, 3.0 g/L yeast extract, 3.0 g/L $(NH_4)_2SO_4$, 1.0 g/L K_2HPO_4, 1.0 g/L KH_2PO_4, 75 g/L glucose) to an initial OD600 nm of 0.4 and incubated approximately 24 hours at 30°C, and 300 rpm shaking. Cells were pelleted by centrifugation for 10 minutes at 3,000xg, at 4°C, and then resuspended in 12.5 mL HiP-TAB media (yeast nitrogen base without amino acids and without ammonium sulfate, 1.7 g/L; yeast extract, 3.0 g/L; potassium phosphate monobasic, 10.0 g/L; potassium phosphate dibasic, 10.0 g/L) and added to 250 mL baffled shake flasks. 500 µL of glucose was added to the shake flasks, which were shaken at approximately 300 rpm, at 30°C. Cultures were incubated for 48 hours and samples were taken at 48 hours for analysis of carotenoid production by HPLC, and results are presented in Table 22 above.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1770 of 3077

EXAMPLE 45: Construction of strain sAA8283 for the production of lycopene from oleic acid

Integration cassettes for CsCrtE, CsCrtB, CsCrtI, and pG6PI-CRC1 were generated by PCR amplification of plasmids pAA2702, pAA2703, pAA2704, and pAA2311 with primers oAA2206 and oAA2209. All four cassettes were purified and chemically transformed into strain sAA4377, which had all copies of FAA1 (Long chain fatty acyl-CoA synthetase), FAT1 (Very long chain fatty acyl-CoA synthetase), CRC1 (Mitochondrial carnitine transporter), and POX4 (Acyl-coenzyme A oxidase) removed. The transformed cells were then plated onto SCD-ura plates. Colonies were streaked onto YPD for isolation and characterization. Colony PCR was performed to confirm the presence of all four genes, and one verified isolate was saved as strain sAA8283.

EXAMPLE 46: Shake flask characterization of sAA8283 on oleic acid

Starter cultures (5 mL) of sAA8283 in YPD were incubated overnight at 30°C, with shaking at approximately 250 rpm. The overnight cultures were used to inoculate 25 mL fresh SP92-glycerol media (6.7 g/L yeast nitrogen base, 3.0 g/L yeast extract, 3.0 g/L $(NH_4)_2SO_4$, 1.0 g/L K_2HPO_4, 1.0 g/L KH_2PO_4, 75 g/L glycerol) to an initial OD600 nm of 0.4 and incubated approximately 24 hours at 30°C, and 300 rpm shaking. Cells were pelleted by centrifugation for 10 minutes at 3,000xg, at 4°C, and then resuspended in 12.5 mL HiP-TAB media (yeast nitrogen base without amino acids and without ammonium sulfate, 1.7 g/L; yeast extract, 3.0 g/L; potassium phosphate monobasic, 10.0 g/L; potassium phosphate dibasic, 10.0 g/L) and added to 250 mL baffled shake flasks. 500 µL of oleic acid was added to the shake flasks, which were shaken at approximately 300 rpm, at 30°C. Cultures were incubated for 48 hours and samples were taken at 48 hours for analysis of carotenoid production by HPLC, and results are presented in Table 22 above.

EXAMPLE 47: Construction of strain sAA8519 for the production of β-carotene from oleic acid

In order to integrate genes into the genome of *C. viswanathii*, genes were cloned into a plasmid that allows for single crossover recombination at the URA3 loci. In addition, since oleic acid was the primary carbon source, all genes were expressed under a HDE promoter with a POX4 terminator. The CrtYB Gene from *Xanthophyllomyces dendrorho* (XdCrtYB) was amplified from pAA2866 using primers oAA9968 and oAA2091, while the plasmid backbone was PCR

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1771 of 3077

amplified from pAA2534 with primers oAA07511 and oAA07512. The two PCR products encoded overlapping sequence at their termini allowing directional ligation, generating plasmid pAA3373. Integration cassettes for CsCrtE, XdCrtYB, CsCrtI, and pG6PI-CRC1 were then generated by PCR amplification of plasmids pAA2702, pAA3373, pAA2704, and pAA2311 with primers oAA2206 and oAA2209. All four cassettes were purified and chemically transformed into strain sAA4377. The transformed cells were then plated onto SCD-ura plates. Colonies were then streaked onto YPD for isolation and characterization. Colony PCR was performed to confirm the presence of all four genes, and one verified isolate was saved as strain sAA8519.

EXAMPLE 48: Shake flask characterization of sAA8519 on oleic acid

Starter cultures (5 mL) of sAA8519 in YPD were incubated overnight at 30°C, with shaking at approximately 250 rpm. The overnight cultures were used to inoculate 25mL fresh SP92-glycerol media (6.7 g/L yeast nitrogen base, 3.0 g/L yeast extract, 3.0 g/L $(NH_4)_2SO_4$, 1.0 g/L K_2HPO_4, 1.0 g/L KH_2PO_4, 75 g/L glycerol) to an initial OD600 nm of 0.4 and incubated approximately 24 hours at 30°C, and 300 rpm shaking. Cells were pelleted by centrifugation for 10 minutes at 3,000xg, at 4°C, and then resuspended in 12.5 mL HiP-TAB media (yeast nitrogen base without amino acids and without ammonium sulfate, 1.7 g/L; yeast extract, 3.0 g/L; potassium phosphate monobasic, 10.0 g/L; potassium phosphate dibasic, 10.0 g/L) and added to 250 mL baffled shake flasks. 500 µL of oleic acid was added to the shake flasks, which were shaken at approximately 300 rpm, at 30°C. Cultures were incubated for 48 hours and samples were taken at 48 hours for analysis of carotenoid production by HPLC, and results are presented in Table 22 above.

EXAMPLE 49: Shake flask characterization of sAA8283 on various carbon sources

Starter cultures (5 mL) of sAA8283 in YPD were incubated overnight at 30°C, with shaking at approximately 250 rpm. The overnight cultures were used to inoculate 25 mL fresh SP92-glycerol media (6.7 g/L yeast nitrogen base, 3.0 g/L yeast extract, 3.0 g/L $(NH_4)_2SO_4$, 1.0 g/L K_2HPO_4, 1.0 g/L KH_2PO_4, 75 g/L glycerol) to an initial OD600 nm of 0.4 and incubated approximately 24 hours at 30°C, and 300 rpm shaking. Cells were pelleted by centrifugation for 10 minutes at 3,000xg, at 4°C, and then resuspended in 12.5 mL HiP-TAB media (yeast nitrogen base without amino acids and without ammonium sulfate, 1.7 g/L; yeast extract, 3.0 g/L; potassium phosphate monobasic, 10.0 g/L; potassium phosphate dibasic, 10.0 g/L) and

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1772 of 3077

added to 250 mL baffled shake flasks. 500 µL of either C10-C14 mixed alkanes, methyl pentadecanoate, ethyl laurate, crude palm oil (CPO), Ethyl PFAD, nonane, or octadecane was added to the shake flasks, which were shaken at approximately 300 rpm, at 30°C. Cultures with CPO and Ethyl PFAD also contained 0.1% Triton X-100. Cultures were incubated for 48 hours and samples were taken at 48 hours for analysis of carotenoid production by HPLC, and results are presented in Table 22 above.

EXAMPLE 50: Construction of strain sAA8503 for production of lycopene

In order to integrate genes into the genome of *C. viswanathii*, genes were cloned into a plasmid that allows for single crossover recombination at the URA3 loci. In addition, since glucose was the primary carbon source, all genes were expressed under a GPD promoter with a POX4 terminator. Integration cassettes for XdCrtE, CsCrtB, and CsCrtI, were generated by PCR amplification of plasmids pAA3189, pAA2699, and pAA2700 with primers oAA2206 and oAA2209. All cassettes were purified and chemically transformed into strain sAA002. The transformed cells were then plated onto SCD-ura plates. Colonies were streaked onto YPD for isolation and characterization. Colony PCR was performed to confirm the presence of all three genes, and one verified isolate was saved as strain sAA8503.

EXAMPLE 51: Shake flask characterization of sAA8503 on glucose

Starter cultures (5 mL) of sAA8503 in YPD were incubated overnight at 30°C, with shaking at approximately 250 rpm. The overnight cultures were used to inoculate 25 mL fresh SP92-glucose media (6.7 g/L yeast nitrogen base, 3.0 g/L yeast extract, 3.0 g/L $(NH_4)_2SO_4$, 1.0 g/L K_2HPO_4, 1.0 g/L KH_2PO_4, 75 g/L glucose) to an initial OD600 nm of 0.4 and incubated approximately 24 hours at 30°C, and 300 rpm shaking. Cells were pelleted by centrifugation for 10 minutes at 3,000xg, at 4°C, and then resuspended in 12.5 mL HiP-TAB media (yeast nitrogen base without amino acids and without ammonium sulfate, 1.7 g/L; yeast extract, 3.0 g/L; potassium phosphate monobasic, 10.0 g/L; potassium phosphate dibasic, 10.0 g/L) and added to 250 mL baffled shake flasks. 500 µL of glucose was added to the shake flasks, which were shaken at approximately 300 rpm, at 30°C. Cultures were incubated for 48 hours and samples were taken at 48 hours for analysis of carotenoid production by HPLC, and results are presented in Table 22 above.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1773 of 3077

EXAMPLE 52: Construction of strain sAA8932 for production of lycopene

In order to integrate genes into the genome of *C. viswanathii*, genes were cloned into a plasmid that allows for single crossover recombination at the URA3 loci. In addition, since glucose was the primary carbon source, all genes were expressed under a GPD promoter with a POX4 terminator. Integration cassettes for XdCrtE, CsCrtB, and XdCrtI, were generated by PCR amplification of plasmids pAA3189, pAA2699, and pAA3490 with primers oAA2206 and oAA2209. All cassettes were purified and chemically transformed into strain sAA002. The transformed cells were then plated onto SCD-ura plates. Colonies were streaked onto YPD for isolation and characterization. Colony PCR was performed to confirm the presence of all three genes, and one verified isolate was saved as strain sAA8932.

EXAMPLE 53: Shake flask characterization of sAA8932 on glucose

Starter cultures (5 mL) of sAA8932 in YPD were incubated overnight at 30°C, with shaking at approximately 250 rpm. The overnight cultures were used to inoculate 25 mL fresh SP92-glucose media (6.7 g/L yeast nitrogen base, 3.0 g/L yeast extract, 3.0 g/L $(NH_4)_2SO_4$, 1.0 g/L K_2HPO_4, 1.0 g/L KH_2PO_4, 75 g/L glucose) to an initial OD600 nm of 0.4 and incubated approximately 24 hours at 30°C, and 300 rpm shaking. Cells were pelleted by centrifugation for 10 minutes at 3,000xg, at 4°C, and then resuspended in 12.5 mL HiP-TAB media (yeast nitrogen base without amino acids and without ammonium sulfate, 1.7 g/L; yeast extract, 3.0 g/L; potassium phosphate monobasic, 10.0 g/L; potassium phosphate dibasic, 10.0 g/L) and added to 250 mL baffled shake flasks. 500 µL of glucose was added to the shake flasks, which were shaken at approximately 300rpm, at 30°C. Cultures were incubated for 48 hours and samples were taken at 48 hours for analysis of carotenoid production by HPLC, and results are presented in Table 22 above.

EXAMPLE 54: Shake flask characterization of sAA8932 on oleic acid

Starter cultures (5 mL) of sAA8932 in YPD were incubated overnight at 30°C, with shaking at approximately 250 rpm. The overnight cultures were used to inoculate 25 mL fresh SP92-glycerol media (6.7g/L yeast nitrogen base, 3.0 g/L yeast extract, 3.0 g/L $(NH_4)_2SO_4$, 1.0 g/L K_2HPO_4, 1.0 g/L KH_2PO_4, 75 g/L glucose) to an initial OD600 nm of 0.4 and incubated approximately 24 hours at 30°C, and 300 rpm shaking. Cells were pelleted by centrifugation for

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1774 of 3077

10 minutes at 3,000xg, at 4°C, and then resuspended in 12.5 mL HiP-TAB media (yeast nitrogen base without amino acids and without ammonium sulfate, 1.7 g/L; yeast extract, 3.0 g/L; potassium phosphate monobasic, 10.0 g/L; potassium phosphate dibasic, 10.0 g/L) and added to 250 mL baffled shake flasks. 500 µL of oleic acid was added to the shake flasks, which were shaken at approximately 300 rpm, at 30°C. Cultures were incubated for 48 hours and samples were taken at 48 hours for analysis of carotenoid production by HPLC and results are presented in Table 22 above.

EXAMPLE 55: Carotenoid extraction

A 0.5 ml sample of fermentation broth was sampled. Yeast cells were isolated from media by centrifugation and supernatant removal. 0.25 ml of zirconia beads and 1.0 ml of tert-Butyl methyl ether (MTBE) were added to each cell pellet. Pellets were bead-beated with a Biospec Mini-Beadbeater-16 for 1.5 minutes. Cell debris and beads were pelleted by centrifugation and the carotenoid containing MTBE extract was saved. Multiple rounds of bead-beating were conducted until remaining cell pellet was white.

EXAMPLE 56: Carotenoid detection

A Thermo Scientific Dionex UltiMate 3000 Liquid Chromatography System with an Acclaim C30 column was used to detect and quantify carotenoids. Column temperature was 20°C, flow rate of 1.0-1.4 mL/min, injection volume of 10 µl, and UV absorbance measured at 474 nm, 450 nm, and 285 nm. Metabolite standards were purchased from Sigma-Aldrich. Samples were diluted in MTBE.

EXAMPLE 57: Construction of strain sAA9432 for production of lycopene

In order to integrate genes into the genome of *C. viswanathii*, genes were cloned into a plasmid that allows for single crossover recombination at the URA3 loci. In addition, since oleic acid was the primary carbon source, all genes were expressed under a HDE promoter with a POX4 terminator. Integration cassettes for XdCrtE, CsCrtB, and XdCrtI, were generated by PCR amplification of plasmids pAA3633, pAA2703, and pAA3634 with primers oAA2206 and oAA2209. All cassettes were purified and chemically transformed into strain sAA002. The transformed cells were then plated onto SCD-ura plates. Colonies were streaked onto YPD for

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1775 of 3077

isolation and characterization. Colony PCR was performed to confirm the presence of all three genes, and one verified isolate was saved as strain sAA9432.

EXAMPLE 58: Shake flask characterization of sAA9432 on oleic acid

Starter cultures (5 mL) of sAA9432 in YPD were incubated overnight at 30°C, with shaking at approximately 250 rpm. The overnight cultures were used to inoculate 25 mL fresh SP92-glycerol media (6.7 g/L yeast nitrogen base, 3.0 g/L yeast extract, 3.0 g/L $(NH_4)_2SO_4$, 1.0 g/L K_2HPO_4, 1.0 g/L KH_2PO_4, 75 g/L glycerol) to an initial OD600 nm of 0.4 and incubated approximately 24 hours at 30°C, and 300 rpm shaking. Cells were pelleted by centrifugation for 10 minutes at 3,000xg, at 4°C, and then resuspended in 12.5 mL HiP-TAB media (yeast nitrogen base without amino acids and without ammonium sulfate, 1.7 g/L; yeast extract, 3.0 g/L; potassium phosphate monobasic, 10.0 g/L; potassium phosphate dibasic, 10.0 g/L) and added to 250 mL baffled shake flasks. 500 µL of oleic acid was added to the shake flasks, which were shaken at approximately 300rpm, at 30°C. Cultures were incubated for 48 hours and samples were taken at 48 hours for analysis of carotenoid production by HPLC, and results are presented in Table 22 above.

EXAMPLE 59: Construction of sAA9539, single copy of RAS2 knocked-out

In order to knock out a single copy of RAS2 from the genome, plasmid pVZ3930 with 400 bp homology region to both the 5' and 3' region surrounding RAS2 in the genome was constructed with a URA3 selection cassette. 5' region was amplified with oAA10284 & oVZ0041, 3' region was amplified with oAA10286 & oAA10287, the URA3 cassette digested from pAA408 using restriction enzymes BamHI/PstI, and all pieces were ligated into a pUC19 vector generating pVZ3930. pVZ3930 was then digested with PacI and the resulting fragment was integrated into the sAA002 genome with URA selection. Correct ras2::URA3/RAS2 transformants were identified and plated onto 5'FOA for the loop-out of URA3, leaving a URA3 terminator scar. One correct transformant was identified and saved as strain sAA9539.

EXAMPLE 60: Construction of strain sAA9670 for production of lycopene

In order to integrate genes into the genome of *C. viswanathii*, genes were cloned into a plasmid that allows for single crossover recombination at the URA3 loci. In addition, since oleic acid

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1776 of 3077

was the primary carbon source, all genes were expressed under a HDE promoter with a POX4 terminator. Integration cassettes for XdCrtE, CsCrtB, and XdCrtI, were generated by PCR amplification of plasmids pAA3633, pAA2703, and pAA3634 with primers oAA2206 and oAA2209. All cassettes were purified and chemically transformed into strain sAA9539. The transformed cells were then plated onto SCD-ura plates. Colonies were streaked onto YPD for isolation and characterization. Colony PCR was performed to confirm the presence of all three genes, and one verified isolate was saved as strain sAA9670.

EXAMPLE 61: Shake flask characterization of sAA9670 on oleic acid

Starter cultures (5 mL) of sAA9670 in YPD were incubated overnight at 30°C, with shaking at approximately 250 rpm. The overnight cultures were used to inoculate 25 mL fresh SP92-glycerol media (6.7 g/L yeast nitrogen base, 3.0 g/L yeast extract, 3.0 g/L $(NH_4)_2SO_4$, 1.0 g/L K_2HPO_4, 1.0 g/L KH_2PO_4, 75 g/L glycerol) to an initial OD600 nm of 0.4 and incubated approximately 24 hours at 30°C, and 300 rpm shaking. Cells were pelleted by centrifugation for 10 minutes at 3,000xg, at 4°C, and then resuspended in 12.5 mL HiP-TAB media (yeast nitrogen base without amino acids and without ammonium sulfate, 1.7 g/L; yeast extract, 3.0 g/L; potassium phosphate monobasic, 10.0 g/L; potassium phosphate dibasic, 10.0 g/L) and added to 250 mL baffled shake flasks. 500 µL of oleic acid was added to the shake flasks, which were shaken at approximately 300rpm, at 30°C. Cultures were incubated for 48 hours and samples were taken at 48 hours for analysis of carotenoid production by HPLC, and results are presented in Table 22 above.

EXAMPLE 62: Construction of sAA9684, leucine auxotrophy in sAA002 background

In order to knock out the first copy of Leu2 from the genome, plasmid pAA3060 with a 259 bp and a 256 bp homology region to both the 5' and 3' region, respectively, of the Leu2 gene was constructed with a URA3 selection cassette. The 5' region was amplified with oAA7682 and oAA7683, the 3' region was amplified with oAA7686 and oAA7687, the URA3 cassette was amplified from pAA408 using oAA7684 and oAA7685, and all pieces were cloned by overlap PCR and topo cloned, generating plasmid pAA3060. pAA3060 was then digested with BamHI and PstI and the resulting fragment was integrated into the sAA002 genome with URA selection. Correct leu2::URA3/LEU2 transformants were identified and plated onto 5'FOA for the loop-out of URA3, leaving a URA3 terminator scar. One correct transformant was identified and saved

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1777 of 3077

as strain sAA9658. In order to knock out the second copy of Leu2 from the genome, plasmid pAA2417 with a 204 bp and a 283 bp homology region to both the 5' and 3' region nested respectively of the Leu2 gene was constructed with a URA3 selection cassette. The 5' region was amplified with oAA7941 and oAA7942, the 3' region was amplified with oAA7945 and oAA7946, the URA3 cassette was amplified from pAA408 using oAA7943 and oAA7944, and all pieces were cloned by overlap PCR and topo cloned, generating plasmid pAA2417. pAA2417 was then digested with BamHI and PstI and the resulting fragment was integrated into the sAA9658 genome with URA selection. Correct leu2/leu2::URA3 transformants were identified and plated onto 5'FOA for the loop-out of URA3, leaving a URA3 terminator scar. One correct transformant was identified and saved as strain sAA9684.

EXAMPLE 63: Construction of sAA9703, leucine auxotrophy in a ras2/RAS2 background

In order to knock out the first copy of Leu2 from the genome, plasmid pAA3060 was digested with BamHI and PstI and the resulting fragment was integrated into the sAA9539 genome with URA selection. Correct leu2::URA3/LEU2 transformants were identified and plated onto 5'FOA for the loop-out of URA3, leaving a URA3 terminator scar. One correct transformant was identified and saved as strain sAA9682. In order to knock out the second copy of Leu2 from the genome, plasmid pAA2417 was digested with BamHI and PstI and the resulting fragment was integrated into the sAA9682 genome with URA selection. Correct leu2/leu2::URA3 transformants were identified and plated onto 5'FOA for the loop-out of URA3, leaving a URA3 terminator scar. One correct transformant was identified and saved as strain sAA9703.

EXAMPLE 64: Construction of strain sAA9748 for production of lycopene with leucine auxotrophy

In order to integrate genes into the genome of *C. viswanathii*, genes were cloned into a plasmid that allows for single crossover recombination at the URA3 loci. In addition, since oleic acid was the primary carbon source, all genes were expressed under a HDE promoter with a POX4 terminator. Integration cassettes for XdCrtE, CsCrtB, and XdCrtI, were generated by PCR amplification of plasmids pAA3633, pAA2703, and pAA3634 with primers oAA2206 and oAA2209. All cassettes were purified and chemically transformed into strain sAA9684. The transformed cells were then plated onto SCD-ura plates with 1.60 g/L leucine supplementation. Colonies were streaked onto YPD for isolation and characterization. Colony PCR was

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1778 of 3077

performed to confirm the presence of all three genes, and one verified isolate was saved as strain sAA9748.

EXAMPLE 65: Construction of strain sAA9750 for production of lycopene with leucine auxotrophy

In order to integrate genes into the genome of *C. viswanathii*, genes were cloned into a plasmid that allows for single crossover recombination at the URA3 loci. In addition, since oleic acid was the primary carbon source, all genes were expressed under a HDE promoter with a POX4 terminator. Integration cassettes for XdCrtE, CsCrtB, and XdCrtI, were generated by PCR amplification of plasmids pAA3633, pAA2703, and pAA3634 with primers oAA2206 and oAA2209. All cassettes were purified and chemically transformed into strain sAA9703. The transformed cells were then plated onto SCD-ura plates with leucine supplementation. Colonies were streaked onto YPD for isolation and characterization. Colony PCR was performed to confirm the presence of all three genes, and one verified isolate was saved as strain sAA9750.

EXAMPLE 66: Construction of single crossover cassettes with leucine selection

In order to integrate genes at the leu2 loci within the genome of C. viswanathii, single crossover plasmid pVZ4045 was generated by amplifying the native LEU2 promoter, open reading frame, and terminator in two pieces split in the open reading frame. The front portion was amplified using primers oVZ339 and oVZ340 using genomic DNA from sAA001, and back portion was amplified using primers oVZ341 and oVZ342 using genomic DNA from sAA001, and the base vector was pUC19 amplified using primers oVZ337 and oVZ338. Fragments were assembled such that the back fragment was 5' of the front fragment using overlap PCR and the resulting correct plasmid was designated pVZ4045. DNA constructs for integration with the single crossover cassettes were cloned in between the back and front portions. Fragments for integration were amplified using primers oVZ373 and oVZ374.

In order to integrate carotenoid genes at the leu2 loci within the genome of *C. viswanathii*, XdCrtE, CsCrtB, and XdCrtI were cloned using overlapping PCR along with the HDE promoter and POX4 terminator from plasmid pAA3633, pAA2703, and pAA3634 respectively using primers oVZ371 and oVZ372 into pVZ4045 amplified with primers oVZ369 and oVZ370, generating plasmids pVZ4056, pVZ4057, and pVZ4058 respectively.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1779 of 3077

In order to integrate mevalonate genes at the leu2 loci with the genome of *C. viswanathii*, each gene was cloned under the POX18 promoter and terminator. First, plasmid pVZ4083 was generated by cloning the POX18 promoter and terminator from genomic DNA isolated from sAA001 using primers oVZ376 and oVZ377, and cloned using overlap PCR into plasmid pVZ4045 amplified with primers oVZ378 and oVZ379. Each gene from the mevalonate pathway was then cloned into pVZ4083. Erg12 was amplified from genomic DNA isolated from sAA001 using primers oVZ394 and oVZ395, and pVZ4083 was amplified using primers oVZ392 and oVZ393, and overlap PCR was used to generate plasmid pVZ4101. Erg13 was amplified from genomic DNA isolated from sAA001 using primers oVZ386 and oVZ387, and pVZ4083 was amplified using primers oVZ384 and oVZ385, and overlap PCR was used to generate plasmid pVZ4099. Erg20 was amplified from genomic DNA isolated from sAA001 using primers oVZ410 and oVZ411, and pVZ4083 was amplified using primers oVZ408 and oVZ409, and overlap PCR was used to generate plasmid pVZ4105. HMG1 was amplified from genomic DNA isolated from sAA001 using primers oVZ390 and oVZ391, and pVZ4083 was amplified using primers oVZ388 and oVZ389, and overlap PCR was used to generate plasmid pVZ4100. IDI1 was amplified from genomic DNA isolated from sAA001 using primers oVZ406 and oVZ407, and pVZ4083 was amplified using primers oVZ404 and oVZ405, and overlap PCR was used to generate plasmid pVZ4104. Erg8 was amplified from genomic DNA isolated from sAA001 using primers oVZ481 and oVZ399, and pVZ4083 was amplified using primers oVZ396 and oVZ480, and overlap PCR was used to generate plasmid pVZ4122. MVD1 was amplified from genomic DNA isolated from sAA001 using primers oVZ402 and oVZ479, and pVZ4083 was amplified using primers oVZ478 and oVZ401, and overlap PCR was used to generate plasmid pVZ4123.

EXAMPLE 67: Construction of strains sAA9811, sAA9812, and sAA9814 for production of lycopene

In order to integrate genes into the genome of *C. viswanathii*, genes were cloned into a plasmid that allows for single crossover recombination at the leu2 loci. Three libraries of genes were integrated into sAA9748 using leucine selection. Integration cassettes for XdCrtE, CsCrtB, and XdCrtI were generated by PCR amplification of plasmids pVZ4056, pVZ4057, and pVZ4058 respectively using primers oVZ373 and oVZ374. All cassettes were purified and chemically transformed into strain sAA9748. The transformed cells were then plated onto SCD-leu plates. Colonies were streaked onto YPD for isolation and characterization. Colony PCR was

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1780 of 3077

performed to confirm the presence of all three genes, and one verified isolate was saved as strain sAA9811.

Integration cassettes for Erg10, Erg13, HMG1, Erg12, Erg8, MVD1, IDI1, and Erg20 were generated by PCR amplification of plasmids pVZ4098, pVZ4099, pVZ4100, pVZ4101, pVZ4122, pVZ4123, pVZ4104, and pVZ4105 using primers oVZ373 and oVZ374. All cassettes were purified and chemically transformed into strain sAA9748. The transformed cells were then plated onto SCD-leu plates. Colonies were streaked onto YPD for isolation and characterization. Colony PCR was performed to confirm the presence of all eight genes, and one verified isolate was saved as strain sAA9812.

Integration cassettes for XdCrtE, CsCrtB, XdCrtI, Erg10, Erg13, HMG1, Erg12, Erg8, MVD1, IDI1, and Erg20 were generated by PCR amplification of plasmids pVZ4056, pVZ4057, pVZ4058, pVZ4098, pVZ4099, pVZ4100, pVZ4101, pVZ4122, pVZ4123, pVZ4104, and pVZ4105 respectively using primers oVZ373 and oVZ374. All cassettes were purified and chemically transformed into strain sAA9748. The transformed cells were then plated onto SCD-leu plates. Colonies were streaked onto YPD for isolation and characterization. Colony PCR was performed to confirm the presence of all 11 genes. One verified isolate contained all genes except overexpression of Erg8, and was saved as strain sAA9814.

EXAMPLE 68: Construction of strains sAA9817, sAA9819, and sAA9821 for production of lycopene

In order to integrate genes into the genome of *C. viswanathii*, genes were cloned into a plasmid that allows for single crossover recombination at the leu2 loci. Three libraries of genes were integrated into sAA9750 using leucine selection. Integration cassettes for XdCrtE, CsCrtB, and XdCrtI were generated by PCR amplification of plasmids pVZ4056, pVZ4057, and pVZ4058 respectively using primers oVZ373 and oVZ374. All cassettes were purified and chemically transformed into strain sAA9750. The transformed cells were then plated onto SCD-leu plates. Colonies were streaked onto YPD for isolation and characterization. Colony PCR was performed to confirm the presence of all three genes, and one verified isolate was saved as strain sAA9819.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1781 of 3077

Integration cassettes for Erg10, Erg13, HMG1, Erg12, Erg8, MVD1, IDI1, and Erg20 were generated by PCR amplification of plasmids pVZ4098, pVZ4099, pVZ4100, pVZ4101, pVZ4122, pVZ4123, pVZ4104, and pVZ4105 using primers oVZ373 and oVZ374. All cassettes were purified and chemically transformed into strain sAA9750. The transformed cells were then plated onto SCD-leu plates. Colonies were streaked onto YPD for isolation and characterization. Colony PCR was performed to confirm the presence of all eight genes. One verified isolate contained all genes except overexpression of Erg8, and was saved as strain sAA9817.

Integration cassettes for XdCrtE, CsCrtB, XdCrtI, Erg10, Erg13, HMG1, Erg12, Erg8, MVD1, IDI1, and Erg20 were generated by PCR amplification of plasmids pVZ4056, pVZ4057, pVZ4058, pVZ4098, pVZ4099, pVZ4100, pVZ4101, pVZ4122, pVZ4123, pVZ4104, and pVZ4105 respectively using primers oVZ373 and oVZ374. All cassettes were purified and chemically transformed into strain sAA9750. The transformed cells were then plated onto SCD-leu plates. Colonies were streaked onto YPD for isolation and characterization. Colony PCR was performed to confirm the presence of all eleven genes. One verified isolate contained all genes except overexpression of Erg8 and IDI1, and was saved as strain sAA9821.

EXAMPLE 69: Shake flask characterization of sAA9811, sAA9812, sAA9814, sAA9817, sAA9819, and sAA9821 on oleic acid

Starter cultures (5 mL) of sAA9811, sAA9812, sAA9814, sAA9817, sAA9819, and sAA9821 in YPD were incubated overnight at 30°C, with shaking at approximately 250 rpm. The overnight cultures were used to inoculate 25 mL fresh SP92-glycerol media (6.7 g/L yeast nitrogen base, 3.0 g/L yeast extract, 3.0 g/L $(NH_4)_2SO_4$, 1.0 g/L K_2HPO_4, 1.0 g/L KH_2PO_4, 75 g/L glycerol) to an initial OD600 nm of 0.4 and incubated approximately 24 hours at 30°C, and 300 rpm shaking. Cells were pelleted by centrifugation for 10 minutes at 3,000xg, at 4°C, and then resuspended in 12.5 mL HiP-TAB media (yeast nitrogen base without amino acids and without ammonium sulfate, 1.7 g/L; yeast extract, 3.0 g/L; potassium phosphate monobasic, 10.0 g/L; potassium phosphate dibasic, 10.0 g/L) and added to 250 mL baffled shake flasks. 500 µL of oleic acid was added to the shake flasks, which were shaken at approximately 300rpm, at 30°C. Cultures were incubated for 48 hours and samples were taken at 48 hours for analysis of carotenoid production by HPLC, and results are presented in Table 22 above.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1782 of 3077

EXAMPLE 70: Genes

Details for certain genes described herein are provided in Tables 23, 24 and 25 below.

Table 23

Gene	Organism	Protein sequence	SEQ ID NO
CsCrtE	Chronobacter sakazakii	MNANAVKSSGQEIELQALRDALQTRLDELLPPGQERDLVCAAMREGALTPGKRVRPLLLILAARD LGCDASQPALMDLACAVEMVHAASLMLDDIPCMDNALLRRGKPTIHRQYGESVAILAAVALLSRA FGVVAQANPLSDSCKTQAVSELSSAVGLQGLVQGQFRDLSEGNQARSAEAILATNDLKTSVLFD ATLQIAAIAAGASASVRHKLREFSRHLGQAFQLLDDLADGLNHTGKDINKDAGKSTLVAMLGPEA VHQRLRDHLLRADEHLTGACSRGASTRRFMYAWFDKQLAMFG	326
CsCrtI	Chronobacter sakazakii	MTKTVVIGSGFGGLALAIRLQAAGVPTLLLEQRDKPGGRAYVYEDKGFTFDAGPTVITDPSAIEEL FTLAGKNIADYVDLLPVTPFYRLCWENGQVFNYDNDQASLEAQIARFNPRDVEGYRQFLAYSQA VFKEGYLKLGAVPFLSFRDMLRAGPQLARLQAWRSVYGMVSKFIENDHLRQAFSFHSLLVGGNP FATSSIYTLIHALERQWGVWFARGGTGALVQGLVKLFTDLGGEIELNAKVTRLDTQGDKISGVTLA DGRRIPARAVASNADVVHTYNNLLGHHPRGVSQAASLRRKRMSNSLFVLYFGLNHHHSQLAHHT VCFGPRYKGLIEDIFKRDSLADDFSLYLHAPCVTDPSLAPPGCGSYYVLAPVPHLGTANLNWDVE GPRLRDRIFEYLEQHYMPGLRDQLVTHRMFTPFDFRDQLGAYHGSAFSVEPILTQSAWFRPHNR DSRIDNLYLVGAGTHPGAGIPGVIGSAKATAGLMLEGHA	327
CsCrtB	Chronobacter sakazakii	MSDKPLLTHATETIEAGSKSFATASKLFDAKTRRSALMLYAWCRHCDDVTDGQALGFRAADAPT DTPQARIALLRALTLEAYAGKPMREPNFAAFQEVALAHQIPPALALDHLEGFAMDVREERYHTFD DTLRYCYHVAGVVGLMMARVMGVRDEAVLDRACDLGLAFQLTNIARDIVEDAAIGRCYLPEAWL QEEGLCADTLTDRAHRPALARLAARLVDEAEPYYASARAGLAGLPLRSAWAIATAHGVYREIGVK VKRAGVNAWETRQGTSKAEKLALLAKGAVMAVSSRGASSSPRPSALWQRPRAQDDRYAHAAP PAA	328
XdCrtE	Xanthophyllomyces dendrorhous	MDYANILTAIPLEFTPQDDIVLLEPYHYLGKNPGKEIRSQLIEAFNYWLDVKKEDLEVIQNVVGMLH TASLLMDDVEDSSVLRRGSPVAHLIYGIPQTINTANYVYFLAYQEIFKLRPTPIPMPVIPPSSASLQS SVSSASSSSSASSENGGTSTPNSQIPFSKDTYLDKVITDEMLSLHRGQGLELFWRDSLTCPSEEE YVKMVLGKTGGLFRIAVRLMMAKSECDIDFVQLVNLISIYFQIRDDYMNLQSSEYAHNKNFAEDLT EGKFSFPTIHSIHANPSSRLVINTLQKKSTSPEILHHCVNYMRTETHSFEYTQEVLNTLSGALEREL GRLQGEFAEANSKIDLGDVESEGRTGKNVKLEAILKKLADIPL	329

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1783 of 3077

Table 23

Gene	Organism	Protein sequence	SEQ ID NO
XdCrtI	Xanthophyllomyces dendrorhous	MGKEQDQDKPTAIIVGCGIGGIATAARLAKEGFQVTVFEKNDYSGGRCSLIERDGYRFDQGPSLL LLPDLFKQTFEDLGEKMEDWVDLIKCEPNYVCHFHDEETFTFSTDMALLKREVERFEGKDGFDR FLSFIQEAHRHYELAVVHVLQKNFPGFAAFLRLQFIGQILALHPFESIWTRVCRYFKTDRLRRVFSF AVMYMGQSPYSAPGTYSLLQYTELTEGIWYPRGGFWQVPNTLLQIVKRNNPSAKFNFNAPVSQV LLSPAKDRATGVRLESGEEHHADVVIVNADLVYASEHLIPDDARNKIGQLGEVKRSWWADLVGG KKLKGSCSSLSFYWSMDRIVDGLGGHNIFLAEDFKGSFDTIFEELGLPADPSFYVNVPSRIDPSAA PEGKDAIVILVPCGHIDASNPQDYNKLVARARKFVIQTLSAKLGLPDFEKMIVAEKVHDAPSWEKE FNLKDGSILGLAHNFMQVLGFRPSTRHPKYDKLFFVGASTHPGTGVPIVLAGAKLTANQVLESFD RSPAPDPNMSLSVPYGKPLKSNGTGIDSQVQLKFMDLERWVYLLVLLIGAVIARSVGVLAF	330
XdCrtY B	Xanthophyllomyces dendrorhous	MTALAYYQIHLIYTLPILGLLGLLTSPILTKFDIYKISILVFIAFSATTPWDSWIIRNGAWTYPSAESGQ GVFGTFLDVPYEEYAFFVIQTVITGLVYVLATRHLLPSLALPKTRSSALSLALKALIPLPIIYLFTAHP SPSPDPLVTDHYFYMRALSLLITPPTMLLAALSGEYAFDWKSGRAKSTIAAIMPTVYLIWVDYVAV GQDSWSINDEKIVGWRLGGVLPIEEAMFFLLTNLMIVLGLSACDHTQALYLLHGRTIYGNKKMPSS FPLITPPVLSLFFSSRPYSSQPKRDLELAVKLLEEKSRSFFVASAGFPSEVRERLVGLYAFCRVTD DLIDSPEVSSNPHATIDMVSDFLTLLFGPPLHPSQPDKILSSPLLPPSHPSRPTGMYPLPPPPSLSP AELVQFLTERVPVQYHFAFRLLAKLQGLIPRYPLDELLRGYTTDLIFPLSTEAVQARKTPIETTADLL DYGLCVAGSVAELLVYVSWASAPSQVPATIEEREAVLVASREMGTALQLVNIARDIKGDATEGRF YLPLSFFGLRDESKLAIPTDWTEPRPQDFDKLLSLSPSSTLPSSNASESFRFEWKTYSLPLVAYAE DLAKHSYKGIDRLPTEVQAGMRAACASYLLIGREIKVVWKGDVGERRTVAGWRRVRKVLSVVMS GWEGQ	331
PaCrtE	Pantoea ananatis	MTVCAKKHVHLTRDAAEQLLADIDRRLDQLLPVEGERDVVGAAMREGALAPGKRIRPMLLLLTAR DLGCAVSHDGLLDLACAVEMVHAASLILDDMPCMDDAKLRRGRPTIHSHYGEHVAILAAVALLSK AFGVIADADGLTPLAKNRAVSELSNAIGMQGLVQGQFKDLSEGDKPRSAEAILMTNHFKTSTLFC ASMQMASIVANASSEARDCLHRFSLDLGQAFQLLDDLTDGMTDTGKDSNQDAGKSTLVNLLGPR AVEERLRQHLQLASEHLSAACQHGHATQHFIQAWFDKKLAAVS	332
PaCrtI	Pantoea ananatis	MKPTTVIGAGFGGLALAIRLQAAGIPVLLLEQRDKPGGRAYVYEDQGFTFDAGPTVITDPSAIEELF ALAGKQLKEYVELLPVTPFYRLCWESGKVFNYDNDQTRLEAQIQQFNPRDVEGYRQFLDYSRAV FKEGYLKLGTVPFLSFRDMLRAAPQLAKLQAWRSVYSKVASYIEDEHLRQAFSFHSLLVGGNPFA TSSIYTLIHALEREWGVWFPRGGTGALVQGMIKLFQDLGGEVVLNARVSHMETTGNKIEAVHLED GRRFLTQAVASNADVVHTYRDLLSQHPAAVKQSNKLQTKRMSNSLFVLYFGLNHHHDQLAHHTV CFGPRYRELIDEIFNHDGLAEDFSLYLHAPCVTDSSLAPEGCGSYYVLAPVPHLGTANLDWTVEG PKLRDRIFAYLEQHYMPGLRSQLVTHRMFTPFDFRDQLNAYHGSAFSVEPVLTQSAWFRPHNRD KTITNLYLVGAGTHPGAGIPGVIGSAKATAGLMLEDLI	333

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1784 of 3077

Table 23

Gene	Organism	Protein sequence	SEQ ID NO
PaCrtB	Pantoea ananatis	MNNPSLLNHAVETMAVGSKSFATASKLFDAKTRRSVLMLYAWCRHCDDVIDDQTLGFQARQPAL QTPEQRLMQLEMKTRQAYAGSQMHEPAFAAFQEVAMAHDIAPAYAFDHLEGFAMDVREAQYSQ LDDTLRYCYHVAGVVGLMMAQIMGVRDNATLDRACDLGLAFQLTNIARDIVDDAHAGRCYLPAS WLEHEGLNKENYAAPENRQALSRIARRLVQEAEPYYLSATAGLAGLPLRSAWAIATAKQVYRKIG VKVEQAGQQAWDQRQSTTTPEKLTLLLAASGQALTSRMRAHPPRPAHLWQRPL	334
XdCrtR	Xanthophyllomyces dendrorhous	MATLSDLVILLLGALLALGFYNKDRLLGSSSSSASTTSGSSAATANGSKPTYSNGNGNAFKGDPR DFVARMKDQKKRLAVFYGSQTGTAEEYATRIAKEAKSRFGVSSLVCDIEEYDFEKLDQVPEDCAI VFCMATYGEGEPTDNAVQFIEMISQDDPEFSEGSTLDGLKYVVFGLGNKTYEQYNVVGRQLDAR LTALGATRVGERGEGDDDKSMEEDYLAWKDDMFAALATTLSFEEGASGETPDFVVTEVPNHPIE KVFQGELSSRALLGSKGVHDAKNPYASPVLACRELFTGGDRNCIHLEFDITGSGITYQTGDHVAV WPSNPDVEVERLLAVLGLTSPEKRRMIIQVVSLDPTLAKVPFPTPTTYDAVFRHYLDISAVASRQT LAVLAKYAPSEQAAEFLTRLGTDKQAYHTEVVGGHLRLAEVLQLAAGNDITVMPTAENTTVWNIP FDHVVSDVSRLQPRFYSISSSPKLHPNSIHVTAVILKYESQATDRHPARWVFGLGTNYLLNVKQA ANNETTPMISDGQDDVPEHVSAPKYTLEGPRGSYKHDDQLFKVPIHVRRSTFRLPTSPKIPVIMIG PGTGVAPFRGFIQERIALARRSIAKNGPDALADWAPIYLFYGSRDEQDFLYAEEWPAYEAELQGK FKIHVAFSRSGPRKPDGSKIYVQDLLWDQKEVIKSAIVEKRASVYICGDGRNMSKDVEQKLAAML AESKNGSAAVEGAAEVKSLKERSRLLMDVWS	335
XdCrtS	Xanthophyllomyces dendrorhous	MFILVLLTGALGLAAFSWASIAFFSLYLAPRRSSLYNLQGPNHTNYFTGNFLDILSARTGEEHAKY REKYGSTLRFAGIAGAPVLNSTDPKVFNHVMKEAYDYPKPGMAARVLRIATGDGVVTAEGEAHK RHRRIMIPSLSAQAVKSMVPIFLEKGMELVDKMMEDAAEKDMAVGESAGEKKATRLETEGVDVK DWVGRATLDVMALAGFDYKSDSLQNKTNELYVAFVGLTDGFAPTLDSFKAIMWDFVPYFRTMKR RHEIPLTQGLAVSRRVGIELMEQKKQAAVLGSASDQAVDKKDVQGRDILSLLVRANIAANLPESQKL SDEEVLAQISNLLFAGYETSSTVLTWMFHRLSEDKAVQDKLREEICQIDTDMPTLDELNALPYLEA FVKESLRLDPPSPYANRECLKDEDFIPLAEPVIGRDGSVINEVRITKGTMVMLPLFNINRSKFIYGE DAEEFRPERWLEDVTDSLNSIEAPYGHQASFISGPRACFGWRFAVAEMKAFLFVTLRRVQFEPII SHPEYEHITLIISRPRIVGREKEGYQMRLQVKPVE	336
ERG10	Candida viswanathii	MTLPPVYIVSTARTPIGSFQGSLSSLTYSDLGAHAVKAALAKVPQIKPQDVDEIVFGGVLQANVGQ APARQVALKAGLPDSIIASTINKVCASGMKAVIIGAQNIICGTSDIVVVGGAESMSNTPYYLPSARS GARYGDAVMVDGVQKDGLLDVYEEKLMGVAAEKCAKDHGFSREDQDNFAINSYKKAGKALSEG KFKSEIAPVTIKGFRGKPDTVIENDEEIGKFNEDRLKSARTVFQKENGTVTAPNASKLNDGGAALV LVSEAKLKQLGLKPLAKISGWGEAARTPFDFTIAPALAVPKAVKHAGLTVDRVDFFELNEAFSVVG LANAELVKIPLEKLNVYGGAVAMGHPLGCSGARIIVTLLSVLTQEGGRFGAAGVCNGGGGASAIVI EKIDSDAKL	337

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1785 of 3077

Table 23

Gene	Organism	Protein sequence	SEQ ID NO
ERG13	Candida viswanathii	MTNAPQNIGIKGIEVYIPGQAVNQSDLEKFDGIPAGKYTIGLGQTNMAFVNDREDIYSIALTVVSRLI KHYNVDTNNVGRLEVGTETLLDKSKSVKSVLMQLFPDNNDIEGIDTVNACYGGTSAVINAINWIES SSWDGRDAIVVAGDIAIYDKGAARPTGGAGSIAMLIGPDAPIVFDSVRGSFMEHAYDFYKPDFTSE YPVVDGHFSLSCYVKAVDNCYKNYSKKVTGNVDKTVGVYDHFDYNAFHVPTCKLVTKSYARLLY NDYKSDPSKFADLIDESTRQHIDSLSYEASLTDKVLEKTFVTLAKEETKKRVQPALQVPTNTGNMY TASSWVSLASLLYYVGAENLKEKRIGLFSYGSGLASTLLSVTVVGDVSPITKVLDFDYKLGEGRKI QSPEEYLAAIELREKAHLQKSFKPQGSLDNLSQGTYYLTEVDDKFRRSYAIKE	338
HMG1	Candida viswanathii	MLSFITEASGRIAQTAAHRPIHFMVIPALLASIAYLSIIDDYVPEHIRAQYMSGVSYFHPQGSASDLD KWEIHDTTQYANANQISVIPLRFRRFHDSIPQIANAIKISNNEQILIVPSDKAESTVEGLSEITENGIT WRARNNDKLSKYYDYARYGLLRIQDAIHNADNFDILLVFVAYLGMWYSLIKVFIDMRRIGSKFWLA FGTLTSSTFSFLFALVISNKFLDAKVSLRSLSECIPFLVAIIGFKHKVAITTSVAQSSTSSPEDVPHVV GKAVSDQCLFILRDHLVVIIGFLACAAYGNELKGLRNFCILGALILSFDIVSVYTFFSAILALKVEINRA RRTQDLQHVLEEDGISSLVAARVAERNATIEHPNETNFFSSNNSSIVYFKVIMSLGFFAFHAFWLG SSWLYNTSDGGSHGSFSFLSNIPILLTQDISNSIPIGRRGTVVTILPTTFYMPSGIIIQFEDMVYLALS KISSAIRDSLISKCIVFALTISIVTNIYFLNAARFQVSATRKLIDQEMSRPKQAAAAAAAPSAAKSVA PEEDEDETSSEELEIKAPVKPLSLEECTRILKEGKVKTLSNAEVSSLVVGGKLPLYALEKQLGDNK RRTQDLQHVLEEDGISSLVAARVAERNATIEHPNETNFFSSNNSSIVYFKVIMSLGFFAFHAFWLG RAVAVRRKAIAKLANAPVLETNRLPYSHYDYDRVFGACCENVIGYMPIPVGVAGPLVIDGKPYHIP MATTEGCLVASTMRGCKAINAGGGVETVLTKDGMTRGPCVRFPTLKRAGAAKLWIDSEEGQITIK KAFNSTSRFARLQHIQTALAGTSLFIRFRTTTGDAMGMNMISKGVEYSLKHMVEECGWDDMEVIS VSGNYCTDKKPAAINWIEGRGKSIVAAATIPADVVTKVLKSDVDALVELNVSKNLVGSAMAGSVG GFNAHAANLVTAVYLACGQDPAQNVESSNCITLMEKDKQTGDLVISVSMPSIEVGTIGGGTILEPQ AAMLDLLGVRGPHPTNPGDNARQLAKIVASAVLAAELSLCSALAAGHLVQSHMQHNRKGATPAA APAISNGSAKGTKTNGSINGKDLKRLKDGSVTCIKS	339
ERG12	Candida viswanathii	MSVSPFVVSAPGKVIIFGEHSAVYGKPAIAAALSLRCYLLVSPSVDDANTIRLQFPDIQLDHSWDIN DIPWDEIKPFVKYDANNKPLTSSELVPEIVDKLSPLLTDFDNKMHYYACFCFLYLYVNLCTAETPG TTFIVRSTLPIGAGLGSSASTSVCLSSALALLGGWISEPSISATDKILNEDIPDLEFVDKWSLIGEKC FHGNPSGIDNAVATFGGAVMFQRTSAPEQPSIRTNMRNFPAIKLLLTNTKVPKSTADLVAGVGKL NAEFNPITTSILTAMEHLSQEAYKVMITPGFGKDETNTLRKLVNINHGLLVALGVSHPSLETVKIIGD RHKIGATKLTGAGGGGCAITLVNDNVEESAIQEAIKEFAGEGYESFETSLGGKGVGVLFSGDVSV DGKFSQSVFCNYPDRASIEDALGMINVKEWKFW	340
ERG8	Candida viswanathii	MSKAFSAPGKALLAGGYLVLEPTYDAYVTALSSRMHAIITPQKPASISKIKISSPQFANGEWEYHV TSNEKPKDIKSRSNPFLEATIFIVLSYIQPTEPFDLDLVIYSDPGYHSQEHTTQKVSSNGKKKFLYH SRAINDVEKTGLGSSAGLVSVVTTSLLSYFIPGIEESNKDMLHNVAQIAHCFAQKKIGSGFDVATAI	341

Table 23

Gene	Organism	Protein sequence	SEQ ID NO
		YGSIVYRRFQPSLINDVFEILEETPGRFPGALKSLVESNWEFKHERCVLPPKIKLLMGDIKGGSET PKLVSKILQWKKDKPEESGLVYDQLNSANVAFMKKISTLNESSQVQEIDELSDYISAVRKGLQELT EKSKVPVEPPVQTELLDRIAKLPGCLGGVVPGAGGYDAIAVLVLEKEVENFRNKTLENPEYYHNV YWVDLEEETEGVVVENWEDYIGL	
MVD1	Candida viswanathii	MYSASVTAPVNIATLKYWGKRDKTLNLPTNSSISVTLSQDDLRTLTTAAASTTFDKDQLWLNGKLE SLDTPRTQACLVDLRKLRADVEQANADLPKLSTMKLHIVSENNFPTAAGLASSAAGFAALITAIAKL YELPQDMSELSKIARKGSGSACRSLFGGFVAWEMGDAADGQDSKAVEVAPLDHWPSMRAVILV VSDDKKDTPSTTGMQATVQTSDLFAHRVTKVVPQCFEEMKKAIVAKDFPKFTELTMKDSNSFHA VCLDSYPPIFYLNDTSKKIIKLVEAINKHATIAAYTFDAGPNAVIYYDAANEDEVLSQLYKSFGHVQG WKKAYTAETAVAGVSRIIQTSIGQGPQVTNESLINESGLPN	342
IDI1	Candida viswanathii	MSSEYAKLVASFSPNDILAKWPEVTPLKKISGIPRSAESDSSNGSHNNTELFNGHDEEQIRLMEEL CIVLDYDDKPVGAGTKKLCHIMDNINEGLLHRAFSVFLFNEDGKLLLQQRADEKITFPAMWTNTCC SHPLCVPSELGVDADAKDVNNLDNAVHGAKVAAQRKLDHELGIPFSDTPLDQFTYLTRIHYKSAS GAEDSKWGEHEIDYILILKTKNDITINANYNEVKDYKYVDAKELQEMFEDKDLVFTPWFKLICQSFL FKWWNNLSDLEKYQDTEIHRLL	343
ERG20	Candida viswanathii	MSDKAAARERFLSVFECAVEELKEVLVSHKMPQEAIDWFVKNLNYNTPGGKLNRGLSVVDTYAIL NNTTADKLNDEQYKKVALLGWSIELLQAYFLVADDMMDQSKTRRGQKCWYLVEGVGNIAINDSF MLEGAIYVLLKKHFRQDPYYVDLLDLFHEVTFQTELGQLLDLVTADEEVVDLDKFSLDKHSFIVIFK TAYYSFYLPVALAMYMSGISSEEDLKQVRDILIPLGEYFQIQDDFLDCFGTPEQIGKIGTDIKDNKC SWVVNQALLHATPEQRKLLDDNYGKKDDESEQRCKDLFKSMGIEKIYHDYEESIVAKLREQIDKV DESRGLKKDVLTAFLGKVYKRSK	344
BTS1	Candida viswanathii	MSFNIDSLIQPGAAYDPSMTDAIMKPYRYISEVPSNNHNVRTRFLLAFNELFYGMKNEDLLHRISHI ISVFHNSSLLIDDIEDDSQLRRGMPVAHVKYGVPLTINCGNMMYFVAVQKAIDLAGEAGSAELKFE TSQILVDEMMNVHHGQGLDIYWRDYLKDLEHLPEIEDYLGMVKDKTGSLFRLAIKLLSLHSDVGE DNGLVAIANLLGIIYQIRDDYLNLVDIKYSAMKGVTCEDLIEGKLSLPILHCLRTTTNSPVHEILYNYN TSAERAKQNALIEECLTYMKNKSRSLQYTLDLIKTLERKIKAMMTKYPNSENSGLIKIIDRLCDL	345
CRC1	Candida viswanathii	MDDVDSALADNVKSFAAGGFGGICAVLTGHPFDLVKVRLQTGLYKSSVQCVKETIAKDGLFGLYR GVLPPLLGVTPMFAVSFWGYDVGKKLVSSFTGKSVDKFEIKDISTAGFISAIPTTLVAAPFERVKV MMQIQEGAKSKSMGAVVAEMYRTGGIRSIFKGTVATLARDGPGSALYFATYEWVKKELTAPGED LSLFAITTAGGFAGIAMMLGVFPIDTIKSTQQSSNVKVSIVQATKNIYAKGGIKAFFPGVGPALARA FPANAATFLGVELARKFLDKVI	346

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1787 of 3077

Table 23

Gene	Organism	Protein sequence	SEQ ID NO
FAA1	Candida viswanathii	MGAPLTVAVGEAKPGETAPRRKAAQKMASVERPTDSKATTLPDFIEECFARNGTRDAMAWRDL VEIHVETKQVTKIIDGEQKKVDKDWIYYEMGPYNYISYPKLLTLVKNYSKGLLELGLAPDQESKLMI FASTSHKWMQTFLASSFQGIPVVTAYDTLGESGLTHSLVQTESDAVFTDNQLLSSLIRPLEKATSV KYVIHGEKIDPNDKRQGGKIYQDAEKAKEKILQIRPDIKFISFDEVVALGEQSSKELHFPKPEDPICI MYTSGSTGAPKGVVITNANIVAAVGGISTNATRDLVRTVDRVIAFLPLAHIFELAFELVTFWWGAPL GYANVKTLTEASCRNCQPDLIEFKPTIMVGVAAVWESVRKGVLSKLKQASPIQQKIFWAAFNAKS TLNRYGLPGGGLFDAVFKKVKAATGGQLRYVLNGGSPISVDAQVFISTLLAPMLLGYGLTETCAN TTIVEHTRFQIGTLGTLVGSVTAKLVDVADAGYYAKNNQGEIWLKGGPVVKEYYKNEEETKAAFT EDGWFKTGDIGEWTADGGLNIIDRKKNLVKTLNGEYIALEKLESIYRSNHLILNLCVYADQTKVKPI AIVLPIEANLKSMLKDEKIIPDADSQELSSLVHNKKVAQAVLRHLLQTGKQQGLKGIELLQNVVLLD DEWTPQNGFVTSAQKLQRKKILESCKKEVEEAYKSS	347
FAT1	Candida viswanathii	MSGLEIAAAAILGSQLLEAKYLIADDVSLAKTVAVNALPYLWKASRGKASYWYFFEQSVFKNPNN KALAFPRPRKNAPTPKTDAEGFQIYDDQFDLEEYTYKELYDMVLKYSYILKNEYGVTANDTIGVSC MNKPLFIVLWLALWNIGALPAFLNFNTKDKPLIHCLKIVNASQVFVDPDCDSPIRDTEAQIREELPH VQINYIDEFALFDRLRLKSTPKHRAEDKTRRPTDTDSSACALIYTSGTTGLPKAGIMSWRKAFMAS VFFGHIMKIDSKSNVLTAMPLYHSTAAMLGLCPTLIVGGCVSVSQKFSATSFWTQARLCGATHVQ YVGEVCRYLLNSKPHPDQDRHNVRIAYGNGLRPDIWSEFKRRFHIEGIGEFYAATESPIATTNLQY GEYGVGACRKYGSLISLLLSTQQKLAKMDPEDESEIYKDPKTGFCTEAAYNEPGELLMRILNPND VQKSFQGYYGNKSATNSKILTNVFKKGDAWYRSGDLLKMDEDKLLYFVDRLGDTFRWKSENVS ATEVENELMGSKALKQSVVVGVKVPNHEGRACFAVCEAKDELSHEEILKLIHSHVTKSLPVYAQP AFIKIGTIEASHNHKVPKNQFKNQKLPKGEDGKDLIYWLNGDKYQELTEDDWSLICTGKAKL	348
LEU2	Candida viswanathii	MSVKTKTITILPGDHVGTEIVAEAIKSLHAIESSTPYQKVHFEFKHHLIGGAAIDATGVPLPDDALAA AKSSDAVLLGAVGGPKWGTGAVRPEQGLLKIRKELNLYANIRPCNFASDSLLELSPLRPEVVKGT NLIIVRELVGGIYFGERQEQEESEDGKSAWDTEKYTVDEVARITRMAAFMALQHTPPLPIWSLDKA NVLASSRLWRKTVDKIISEEFPTLAVQHQLIDSAAMILIQNPTKLNGIIITSNMFGDIISDEASVIPGSL GLLPSASLASLPDTNTAFGLYEPCHGSAPDLPENKVNPIATILSVASMLRLSLDCVKEAEALEQAV KEVLDKGIRTADLRGSSTTTEVGDAVAETVSRILKEAKA	349
AaCrtZ	Agrobacterium aurantiacum	MTNFLIVVATVLVMELTAYSVHRWIMHGPLGWGWHKSHHEEHDHALEKNDLYGLVFAVIATVLFT VGWIWAPVLWWIALGMTVYGLIYFVLHDGLVHQRWPFRYIPRKGYARRLYQAHRLHHAVEGRD HCVSFGFIYAPPVDKLKQDLKMSGVLRAEAQERT	350
AaCrtW	Agrobacterium aurantiacum	MSAHALPKADLTATSLIVSGGIIAAWLALHVHALWFLDAAAHPILAIANFLGLTWLSVGLFIIAHDAM HGSVVPGRPRANAAMGQLVLWLYAGFSWRKMIVKHMAHHRHAGTDDDPDFDHGGPVRWYAR	351

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1788 of 3077

Table 23

Gene	Organism	Protein sequence	SEQ ID NO
		FIGTYFGWREGLLLPVIVTVYALILGDRWMYVVFWPLPSILASIQLFVFGTWLPHRPGHDAFPDRH NARSSRISDPVSLLTCFHFGGYHHEHHLHPTVPWWRLPSTRTKGDTA	
CnTPS 1	Callitropsis nootkatensis	MAEMFNGNSSNDGSSCMPVKDALRRTGNHHPNLWTDDFIQSLNSPYSDSSYHKHREILIDEIRD MFSNGEGDEFGVLENIWFVDVVQRLGIDRHFQEEIKTALDYIYKFWNHDSIFGDLNMVALGFRILR LNRYVASSDVFKKFKGEEGQFSGFESSDQDAKLEMMLNLYKASELDFPDEDILKEARAFASMYL KHVIKEYGDIQESKNPLLMEIEYTFKYPWRCRLPRLEAWNFIHIMRQQDCNISLANNLYKIPKIYMK KILELAILDFNILQSQHQHEMKLISTWWMKNSSAIQLDFFRHRHIESYFWWASPLFEPEFSTCRINCT KLSTKMFLLDDIYDTYGTVEELKPFTTTLTRWDVSTVDNHPDYMKIAFNFSYEIYKEIASEAERKHG PFVYKYLQSCWKSYIEAYMQEAEWIASNHIPGFDEYLMNGVKSSGMRILMIHALILMDTPLSDEIL EQLDIPSSKSQALLSLITRLVDDVKDFEDEQAHGEMASSIECYMKDNHGSTREDALNYLKIRIESC VQELNKELLEPSNMHGSFRNLYLNVGMRVIFFMLNDGDLFTHSNRKEIQDAITKFFVEPIIP	352
tHMG1	Candida viswanathii	MDQEMSRPKQAAAAAVAPSAAKSVAPEEDEDETSSEELEIKAPVKPLPLEECTRILKEGKVKTLS NAEVSSLVVGGKLPLYALEKQLGDHKRAVAVRRKAIAKLANAPVLETNRLPYSHYDYDRVFGACC ENVIGYMPIPVGVAGPLVIDGKPYHIPMATTEGCLVASTMRGCKAINAGGGVETVLTKDGMTRGP CVRFPTLKRAGAAKLWIDSEEGQITIKKAFNSTSRFARLQHIQTALAGTSLFIRFRTTTGDAMGMN MISKGVEYSLKYMVEECGWDDMEVISVSGNYCTDKKPAAINWIEGRGKSIVAAATIPADVVTKVLK SDVDALVELNVSKNLVGSAMAGSVGGFNAHAANLVTAVYLACGQDPAQNVESSNCITLMEKDK QTGDLVISVSMPSIEVGTIGGGTILEPQAAMLDLLGVRGPHPTNPGDNARQLAKIVASAVLAAELS LCSALAAGHLVQSHMQHNRKGATAAAAAAAAAPAIANGRANGTKTNGSINGKDLKRLKDGSVTCIK S	353
POX4	Candida viswanathii	MTFTKKNVSVSQGPDPRSSIQKERDSSKWNPQQMNYFLEGSVERSELMKALAQQMERDPILFT DGSYYDLTKDQQRELTAVKINRIARYREQESIDTFNKRLSLIGIFDPQVGTRIGVNLGLFLSCIRGN GTTSQLNYWANEKETADVKGIYGCFGMTELAHGSNVAGLETTATFDKESDEFVINTPHIGATKW WIGGAAHSATHCSVYARLIVDGQDYGVKTFVVPLRDSNHDLMPGVTVGDIGAKMGRDGIDNGWI QFSNVRIPRFFMLQKFCKVSAEGEVTLPPLEQLSYSALLGGRVMMVLDSYRMLARMSTIALRYAI GRRQFKGDNVDPKDPNALETQLIDYPLHQKRLFPYLAAAYVISAGALKVEDTIHNTLAELDAAVEK NDTKAIFKSIDDMKSLFVDSGSLKSTATWLGAEAIDQCRQACGGHGYSSYNGFGKAYNDWVVQ CTWEGDNNVLAMSVGKPIVKQVISIEDAGKTVRGSTAFLNQLKDYTGSNSSKVVLNTVADLDDIK TVIKAIEVAIIRLSQEAASIVKKESFDYVGAELVQLSKLKAHHYLLTEYIRRIDTFDQKDLVPYLITLGK LYAATIVLDRFAGVFLTFNVASTEAITALASVQIPKLCAEVRPNVVAYTDSFQQSDMIVNSAIGRYD GDIYENYFDLVKLQNPPSKTKAPYSDALEAMLNRPTLDERERFEKSDETAAILSK	354
POX5	Candida viswanathii	MPTELQKERELTKFNPKELNYFLEGSQERSEIISNMVEQMQKDPILKVDASYYNLTKDQQREVTA	355

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1789 of 3077

Table 23

Gene	Organism	Protein sequence	SEQ ID NO
		KKIARLSRYFEHEYPDQQAQRLSILGVFDPQVFTRIGVNLGLFVSCVRGNGTNSQFFYWTINKGID KLRGIYGCFGMTELAHGSNVQGIETTATFDEDTDEFVINTPHIGATKWWIGGAAHSATHCSVYAR LKVKGKDYGVKTFVVPLRDSNHDLEPGVTVGDIGAKMGRDGIDNGWIQFSNVRIPRFFMLQKYC KVSRLGEVTMPPSEQLSYSALIGGRVTMMMDSYRMTSRFITIALRYAIHRRQFKKKDTDTIETKLID YPLHQKRLFPFLAAAYLFSQGALYLEQTMNATNDKLDEAVSAGEKEAIDAAIVESKKLFVASGCLK STCTWLTAEAIDEARQACGGHGYSSYNGFGKAYSDWVVQCTWEGDNNILAMNVAKPMVRDLLK EPEQKGLVLSSVADLDDPAKLVKAFDHALSGLARDIGAVAEDKGFDITGPSLVLVSKLNAHRFLID GFFKRITPEWSEVLRPLGFLYADWILTNFGATFLQYGIITPDVSRKISSEHFPALCAKVRPNVVGLT DGFNLTDMMTNAAIGRYDGNVYEHYFETVKALNPPENTKAPYSKALEDMLNRPDLEVRERGEKS EEAAEILSS	
URA3	Candida viswanathii	MVSTKTYTERASAHPSKVAQRLFRLMESKKTNLCASIDVTTTAEFLSLIDKLGPHICLVKTHIDIISD FSYEGTIEPLLVLAERHGFLIFEDRKFADIGNTVMLQYTSGVYRIAAWSDITNAHGVTGKGVVEGL KRGAEGVEKERGVLMLAELSSKGSLAHGEYTRETIEIAKSDREFVIGFIAQRDMGGREEGFDWIIM TPGVGLDDKGDALGQQYRTVDEVVLTGTDVIIVGRGLFGKGRDPEVEGKRYRDAGWKAYLKRT GQLE	356

Table 24

Gene	Organism	C.V. Codon Optimized Sequence	SEQ ID NO
CsCrtE	Chronobacter sakazakii	ATGAATGCTAATGCCGTTAAGTCCTCCGGTCAGGAAATCGAATTGCAAGCTTTGAGGGATGC CTTGCAAACACGCTTGGACGAATTGCTCCCACCCGGGCAGGAGAGAGACTTGGTTTGCGCA GCCATGCGCGAAGGGGCATTGACGCCAGGGAAGAGAGTTCGTCCATTACTTTTGATCTTGG CAGCGAGGGACTTGGGTTGCGACGCATCCCAGCCAGCCTTGATGGATTTAGCATGTGCAGT AGAGATGGTTCATGCGGCCTCACTCATGTTGGATGATATTCCCTGTATGGACAACGCCCTTTT GAGACGAGGTAAACCTACAATCCATCGACAGTACGGTGAGTCGGTGGCAATATTGGCGGCT GTTGCGTTGCTTAGCAGAGCATTCGGCGTTGTGGCACAGGCGAACCCTCTATCTGACTCATG TAAGACCCAAGCAGTGTCGGAGCTTTCCTCGGCCGTTGGTTTGCAGGGGTTGGTGCAGGGT CAATTTCGTGATTTGTCCGAAGGGAACCAGGCAAGGAGTGCAGAAGCAATTTTGGCCACCAA CGATTTGAAAACCTCAGTGTTGTTCGACGCTACCTTACAGATTGCGGCCATTGCGGCAGGCG CAAGTGCGTCAGTTAGACACAAGTTGAGAGAGTTCTCTCGACACTTGGGGCAAGCCTTCCAG TTGTTGGACGATTTAGCCGATGGATTAAACCACACGGGAAAGGATATCAACAAAGACGCTGG CAAGAGCACCCTTGTCGCAATGTTGGGTCCGGAAGCTGTGCACCAACGCTTGCGAGACCAT	357

Table 24

Gene	Organism	C.V. Codon Optimized Sequence	SEQ ID NO
		CTTCTTCGCGCAGATGAACACTTGACTGGTGCATGTTCAAGAGGTGCCTCCACCCGTCGCTT CATGTATGCATGGTTTGACAAACAGCTTGCAATGTTTGGGTAA	
CsCrtI	Chronobacter sakazakii	ATGACCAAAACGGTTGTAATAGGATCAGGATTTGGGGGTCTCGCACTAGCCATTAGACTTCA AGCCGCTGGTGTACCAACCTTGTTGCTTGAGCAGCGAGACAAGCCGGGAGGCCGTGCATAT GTGTACGAGGACAAAGGATTTACCTTTGACGCCGGGCCCACTGTCATTACGGACCCATCCG CTATCGAAGAGTTATTTACCCTTGCCGGTAAAAATATAGCTGATTACGTGGATTTGTTGCCAG TAACCCCCTTCTACCGTTTGTGTTGGGAAAATGGTCAAGTGTTTAATTACGATAACGACCAAG CCAGTTTGGAAGCACAAATCGCACGTTTTAACCCAAGGGATGTGGAGGGTTACCGACAATTC CTCGCGTATTCACAAGCAGTATTTAAAGAGGGGTACTTGAAACTAGGCGCGGTTCCATTCTT GAGCTTCAGAGACATGTTGCGTGCGGGACCCCAATTGGCCAGACTCCAAGCCTGGAGGAGT GTCTATGGTATGGTTAGCAAGTTTATCGAGAACGATCATTTGCGACAAGCTTTCAGCTTCCAC AGTTTGTTGGTGGGTGGCAACCCATTCGCAACAAGCTCCATCTATACACTCATCCATGCCTT GGAAAGGCAGTGGGGCGTCTGGTTTGCCAGAGGAGGCACCGGAGCTTTGGTTCAGGGCTT AGTAAAGTTGTTTACTGATTTGGGGGGAGAAATTGAACTCAACGCAAAGGTCACCAGGTTGG ATACCCAAGGAGACAAAATAAGTGGCGTGACTTTGGCAGATGGTAGGCGGATTCCAGCTAG GGCCGTCGCATCCAACGCAGATGTGGTCCATACCTACAACAATTTGTTGGGTCATCATCCGA GAGGCGTTTCGCAAGCTGCCTCTCTTAGAAGAAAGCGCATGAGCAATTCCCTCTTTGTGTTG TATTTCGGGCTCAACCATCACCACTCCCAATTAGCCCACCATACGGTCTGTTTTGGACCACGT TATAAGGGGTTGATAGAAGACATTTTCAAGAGAGACAGCTTGGCAGACGATTTTTCCTTGTAT CTTCACGCTCCGTGTGTGACAGACCCTTCATTGGCTCCACCAGGATGTGGATCATACTACGT ATTGGCTCCTGTGCCACACTTGGGTACTGCAAACTTGAACTGGGATGTAGAGGGGCCACGAT TAAGAGACAGGATCTTTGAATACTTAGAACAACACTATGCCGGGTCTTAGAGATCAACTTG TTACACACCGTATGTTCACTCCGTTTGACTTTAGAGACCAATTAGGTGCCTACCACGGTTCCG CTTTCTCCGTTGAGCCAATCCTAACTCAATCTGCTTGGTTTCGCCCACACAACAGAGACTCTC GAATTGATAACCTTTATTTGGTTGGAGCTGGAACGCATCCCGGTGCAGGAATTCCGGGTGTC ATCGGGAGCGCGAAAGCAACGGCCGGCTTAATGTTGGAGGGCCACGCGTAA	358
CsCrtB	Chronobacter sakazakii	ATGAGTGACAAGCCTTTGTTAACACATGCTACAGAGACCATTGAGGCCGGATCAAAGAGTTT TGCGACCGCCAGTAAGCTCTTTGACGCCAAGACACGTAGAAGTGCTCTTATGTTGTACGCTT GGTGTCGTCATTGCGATGACGTCACTGACGGTCAAGCCTTGGGATTTAGGGCTGCAGACGC TCCTACAGATACACCACAGGCTCGTATCGCATTGTTGCGTGCTCTAACCTTGGAAGCGTATG CTGGTAAACCAATGCGTGAACCAAACTTTGCAGCTTTCCAGGAGGTTGCTCTAGCACATCAG ATCCCACCCGCATTGGCACTCGATCACTTGGAAGGGTTTGCGATGGATGTACGTGAAGAACG CTACCATACCTTTGACGATACCCTAAGATACTGTTATCATGTTGCAGGCGTTGTTGGGTTGAT GATGGCTAGAGTTATGGGTGTCCGAGACGAGGCTGTCTTAGATAGAGCCTGCGATTTAGGTT	359

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1791 of 3077

Table 24

Gene	Organism	C.V. Codon Optimized Sequence	SEQ ID NO
		TGGCTTTTCAGTTAACAAATATTGCCCGAGACATTGTCGAAGATGCGGCAATCGGACGATGC TATCTCCCAGAGGCCTGGTTGCAAGAAGAAGGACTTTGTGCGGATACATTGACTGACAGAGC ACACAGACCAGCTTTGGCTAGGCTAGCCGCACGATTGGTGGATGAGGCAGAGCCATACTAT GCCTCAGCGAGAGCTGGATTGGCCGGACTCCCATTAAGGTCAGCCTGGGCCATCGCTACTG CCCACGGGGTTTACAGAGAAATCGGGCGTAAAGGTGAAGAGGGCTGGTGTGAACGCGTGGG AAACGCGACAGGGAACGTCGAAGGCCGAGAAACTCGCCTTGTTGGCCAAGGGGGCGGTGA TGGCAGTCTCCTCCAGAGGAGCCAGTTCAAGTCCACGACCTTCCGCGTTGTGGCAAAGACC ACGCGCTCAGGATGATCGGTACGCGCATGCCGCCCTCCTGCCGCTTAA	
XdCrtE	Xanthophyllomyces dendrorhous	ATGGACTACGCCAACATCCTCACTGCCATTCCATTGGAATTTACGCCACAAGATGATATAGTT TTGTTGGAACCCTACCACTACCTCGGTAAAAACCCCGGGAAGGAAATCCGATCACATTGAT TGAGGCATTCAACTACTGGTTAGATGTTAAAAAGGAGGATCTCGAGGTTATTCAGAATGTAGT CGGAATGTTGCACACAGCATCCCTTTTGATGGATGATCGTCGAAGACTCTTCTGTACTACGAC GTGGATCACCGGTCGCGCATTTGATTTACGGAATCCCACAAACCATTAACACCGCCAACTAC GTTTATTTCCTAGCTTATCAGGAAATATTCAAATTGAGGCCTACCCCAATCCCTATGCCAGTC ATCCCCCCCATCGTCGGCCTCATTGCAGTCCAGCGTAAGCTCAGCTAGCTCCTCGTCATCTGC GTCATCGGAAAACGGGGGAACATCAACACCCAACAGTCAAATTCCATTCAGCAAGGACACCT ACCTTGACAAAGTAATCACCGATGAAATGCTTTCCTTGCACAGGGGCCAGGGCTTAGAACTT TTCTGGAGAGACTCGCTAACATGTCCCAGTGAGGAAGAGTACGTTAAGATGGTCCTTGGCAA AACAGGGGGGGCTATTCCGAATTGCTGTCAGGTTGATGATGGCTAAAAGTGAATGCGATATTG ACTTCGTTCAATTGGTCAATTTAATTTCGATATATTTCCAGATTAGAGATGACTACATGAACTT GCAATCGTCCGAGTATGCTCACAACAAGAACTTCGCGGAGGATCTAACGGAGGGCAAGTTTT CATTCCCTACGATTCACTCGATTCACGCAAATCCATCAAGTCGTTTGGTCATCAACACTTTGC AAAAGAAGTCAACAAGCCCTGAAATCCTTCATCACTGTGTGAATTACATGAGAACTGAAACTC ACTCCTTTGAGTACACCCAAGAGGTATTGAACACACTCAGTGGAGCCTTAGAGAGAGAACTT GGTCGGTTGCAAGGAGAGTTTGCCGAAGCCAACAGCAAAATCGACTTGGGGGATGTGGAAA GTGAAGGTCGCACAGGTAAAAACGTAAAACTTGAGGCAATCTTGAAAAAGCTTGCCGACATC CCATTATAA	360
XdCrtI	Xanthophyllomyces dendrorhous	ATGGGCAAGGAACAAGATCAAGACAAACCTACTGCCATTATCGTCGGTTGTGGTATAGGCGG CATCGCTACCGCCGCTCGTCTTGCTAAGGAGGGCTTTCAAGTCACGGTGTTCGAGAAGAAC GATTATAGCGGAGGCAGGTGTTCGCTAATCGAACGCGATGGTTACCGATTTGACCAGGGTC CGAGTTTGTTATTGCTCCCGGACTTGTTTAAACAAACCTTTGAAGACCTTGGAGAAAAGATGG AGGATTGGGTAGACTTAATAAAATGCGAACCGAACTACGTTTGTCATTTTCATGACGAAGAGA CTTTCACCTTTAGTACGGACATGGCATTGTTGAAGAGAGAAGTCGAGAGATTTGAAGGAAAG GATGGTTTCGACAGATTTCTATCCTTTATTCAAGAAGCCCACCGTCATTATGAGTTGGCTGTC	361

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1792 of 3077

Table 24

Gene	Organism	C.V. Codon Optimized Sequence	SEQ ID NO
		GTGCACGTCTTACAAAAAAATTTCCCAGGTTTCGCTGCATTCCTTCGCCTTCAATTTATAGGC CAGATCTTAGCATTGCACCCATTTGAATCAATATGGACCCGCGTGTGCAGATACTTCAAGAC TGACCGCTTGAGAAGAGTTTTTTCGTTCGCAGTCATGTACATGGGTCAAAGCCCTTACTCAG CGCCGGGAACATATAGTTTGTTGCAGTACACCGAGTTAACGGAGGGAATATGGTACCCAAGG GGCGGTTTCTGGCAGGTTCCAAATACCTTGTTGCAAATTGTTAAACGAAACAACCCTTCAGCT AAATTCAACTTCAATGCACCAGTGAGTCAGGTCCTCTTGAGTCCTGCCAAAGACAGAGCAAC CGGAGTGAGGTTGGAGAGTGGCGAAGAGCATCACGCGGACGTTGTCATCGTCAACGCAGAC CTTGTTTATGCTAGCGAGCATCTTATTCCTGATGACGCTAGGAACAAGATTGGGCAATTAGGA GAGGTGAAAAGGTCTTGGTGGGCGGATTTGGTGGGAGGTAAGAAACTCAAAGGTTCGTGTT CTTCATTGTCATTCTACTGGTCCATGGACCGAATAGTGGACGGTTTGGGGGGTCACAATATC TTTTTGGCCGAGGACTTCAAGGGTAGTTTTGACACCATTTTCGAGGAACTAGGATTGCCGGC TGATCCTTCTTTTTACGTAAATGTTCCATCTCGCATCGATCCGTCAGCTGCCCCGGAAGGTAA GGATGCCATTGTGATACTAGTGCCATGTGGTCATATTGATGCCAGCAACCCACAAGATTACA ATAAGTTGGTTGCTAGGGCTAGGAAGTTCGTGATTCAGACATTGTCGGCAAAGTTGGGACTT CCAGACTTCGAAAAAATGATTGTTGCGGAAAAGGTGCACGATGCTCCGTCGTGGGAAAAGGA GTTCAACCTCAAGGATGGTTCAATCTTGGGTTTGGCGCACAACTTCATGCAAGTGCTCGGGT TCAGGCCGTCGACTAGACACCCTAAGTATGACAAACTCTTTTTTGTTGGAGCTAGCACACATC CTGGTACAGGGGTTCCAATCGTCCTTGCTGGTGCCAAATTGACAGCTAATCAAGTGTTGGAG TCCTTCGACAGATCTCCAGCCCCCGATCCAAATATGAGCCTATCTGTTCCATACGGCAAGCC TTTAAAGTCGAACGGGACGGGCATCGATAGTCAAGTCCAGTTGAAGTTCATGGATTTGGAAA GATGGGTTTACTTGTTGGTTTTGTTAATCGGAGCCGTTATTGCCAGAAGCGTTGGAGTTCTCG CATTTTAA	362
XdCrtY B	Xanthophyllomyces dendrorhous	ATGACGGCATTGGCGTACTACCAAATCCACTTGATTTATACCTTGCCCATACTAGGGTTGTTG GGTTTGTTAACTTCTCCAATCCTTACCAAATTTGATATATATAAAATCTCTATCTTGGTTTTTAT TGCATTTTCAGCAACAACTCCATGGGACAGTTGGATAATTAGAAATGGGGCATGGACTTACC CTTCGGCCGAGTCGGGTCAAGGCGTCTTTGGGACTTTTCTAGATGTTCCATATGAAGAATAC GCTTTCTTCGTTATCCAAACGGTTATAACAGGTTTGGTCTACGTCCTCGCCACAAGGCATTTG TTACCATCGCTCGCCTTGCCAAAGACACGATCCTCGGCATTAAGCCTCGCGTTGAAGGCATT GATTCCGCTCCCAATTATCTATTTGTTCACAGCCCACCCAAGTCCTTCACCAGATCCCTTGGT TACCGATCATTACTTCTATATGAGGGCGTTGTCACTCTTGATCACGCCTCCCACCATGTTACT TGCTGCCTTGAGCGGGGAATACGCGTTCGACTGGAAGTCCGGCCGAGCAAAAAGCACGATT GCAGCTATCATGATTCCAACGGTATACCTAATTTGGGTCGATTACGTCGCAGTAGGACAGGA CAGCTGGAGTATCAACGACGAAAAAATCGTGGGCTGGCGCCTTGGGGGGGTCCTTCCTATA GAAGAGGCAATGTTCTTTTTGTTGACGAATCTCATGATCGTGTTAGGGCTTTCAGCATGCGAC	

Table 24

Gene	Organism	C.V. Codon Optimized Sequence	SEQ ID NO
		CATACTCAAGCCTTGTATCTTCTCCATGGTAGAACTATTTATGGTAATAAAAAAATGCCGTCGT	
		CGTTCCCATTGATTACCCCGCCAGTGCTTTCTTTGTTCTTCTCAAGTAGACCGTATTCATCGC	
		AACCCAAGCGTGATTTGGAATTGGCGGTTAAGTTGTTGGAAGAGAAATCCCGTAGCTTCTTC	
		GTGGCAAGTGCCGGTTTTCCAAGTGAAGTGAGAGAGAGGTTGGTGGGCTTATACGCGTTTTT	
		GCAGAGTCACGGACGACCTTATAGACAGTCCAGAAGTGTCCTCTAACCCACACGCGACGAT	
		CGATATGGTTAGCGATTTTTTTGACCTTGCTTTTTGGACCACCCTTGCACCCATCCCAACCAGA	
		TAAAATTTTGAGCTCACCTCTACTCCCGCCCAGTCACCCAAGCAGGCCCACGGGGATGTACC	
		CACTCCCCCCCGCCACCGAGTTTGTCTCCAGCCGAATTGGTTCAGTTCTTGACGGAACGCGTT	
		CCAGTGCAGTATCACTTTGCGTTCAGGTTGTTGGCGAAATTACAAGGGTTGATCCCAAGATAT	
		CCGTTAGACGAGCTATTGAGAGGTTACACCACCGATTTGATATTCCCTTTGTCCACCGAGGC	
		TGTGCAAGCCAGAAAAACCCCAATCGAGACTACGGCAGATCTTTTGGATTACGGTTTGTGCG	
		TAGCGGGGAGCGTCGCCGAGTTATTAGTCTACGTTTCGTGGGCATCCGCTCCATCACAGGTT	
		CCAGCCACTATTGAGGAAAGAGAGGCCGTGTTGGTCGCTAGCAGAGAGATGGGTACAGCGT	
		TGCAATTAGTGAATATTGCTCGAGATATTAAAGGTGATGCCACTGAGGGCCGTTTCTACTTGC	
		CATTGAGTTTTTTCGGTTTGCGTGACGAATCAAAACTCGCAATACCAACAGATTGGACTGAAC	
		CTAGGCCACAAGACTTCGATAAATTGTTGTCGCTCTCCCCAAGCTCCACTCTTCCATCATCAA	
		ATGCGTCCGAATCCTTCAGATTTGAGTGGAAGACGTACAGTCTTCCCCTAGTGGCCTACGCC	
		GAAGACTTGGCGAAACACAGTTACAAGGGAATTGACCGTCTACCCACCGAGGTGCAGGCAG	
		GTATGAGGGCGGCGTGCGCCAGTTACCTTTTGATAGGGCGCGAAATCAAGGTTGTGTGGAA	
		GGGTGACGTAGGTGAGCGACGGACGGGTGGCCGGTTGGCGAAGAGTCAGAAAAGTTCTTTCT	
		GTGGTTATGTCGGGTTGGGAGGGACAATAA	
PaCrtE	Pantoea ananatis	ATGACGGTGTGTGCTAAGAAGCACGTCCACTTGACTAGAGACGCCGCCGAACAATTATTGGC	363
		CGACATTGACAGGAGACTTGACCAACTTTTGCCGGTTGAAGGTGAGAGGGATGTTGTGGGA	
		GCTGCAATGCGAGAGGGGGCTTTAGCCCCAGGTAAGAGGATAAGACCTATGTTACTTCTATT	
		GACAGCGAGGGATCTAGGTTGTGCTGTCTCACATGATGGTTTGCTAGATTTGGCCTGTGCGG	
		TCGAGATGGTCCATGCCGCCAGCCTCATCCTCGACGATATGCCCTGTATGGATGACGCTAAA	
		TTGAGACGTGGTCGGCCAACGATTCACTCGCACTACGGGGAACATGTTGCAATCCTAGCCG	
		CAGTCGCCTTGTTGTCAAGGCCTTCGGCGTTATCGCCGACGCCGACGGCTTGACGCCATT	
		GGCAAAAAACAGAGCGGTGAGTGAGTTGAGTAACGCTATAGGTATGCAAGGGTTGGTTCAA	
		GGACAATTCAAAGATTTGTCCGAGGGGGACAAGCCAAGAAGTGCTGAAGCAATTTTGATGAC	
		CAATCACTTCAAAACATCCACATTGTTCTGCGCATCTATGCAAATGGCATCCATTGTTGCAAA	
		CGCCTCATCGGAGGCCAGAGATTGTCTACATCGATTTTCTCTTGATTTGGGCCAAGCGTTCC	
		AGTTGTTGGACGACCTTACAGACGGTATGACCGATACTGGCAAAGATTCGAACCAAGACGCA	
		GGTAAAAGTACGTTGGTAAATCTCTTGGGTCCTAGAGCTGTCGAGGAGAGGCTTAGACAACA	

Table 24

Gene	Organism	C.V. Codon Optimized Sequence	SEQ ID NO
		TCTCCAGTTAGCATCCGAACATCTATCAGCTGCCTGCCAACACGGTCACGCCACTCAACACT TTATTCAAGCGTGGTTTGACAAGAAGTTGGCGGCTGTAAGCTAA	
PaCrtI	Pantoea ananatis	ATGAAACCCACAACCGTTATTGGTGCTGGCTTTGGCGGTTTGGCATTGGCCATAAGATTGCA AGCTGCCGGCATACCTGTGTTATTGCTTGAACAGCGGGACAAGCCAGGAGGACGGGCGTAC GTCTATGAAGATCAGGGTTTCACTTTTGATGCTGGGCCCACGGTTATCACAGACCCGTCGGC GATCGAGGAACTTTTTGCCTTGGCTGGAAAGCAATTGAAAGAATATGTCGAGTTACTTCCAGT TACACCATTTTATCGATTGTGCTGGGAATCCGGTAAGGTGTTTAATTACGACAACGACCAAAC TAGGCTTGAGGCCCAAATCCAACAATTCAACCCCCGGGATGTTGAGGGTTATCGCCAATTCC TCGACTATAGCAGAGCTGTCTTCAAGGAAGGGTACTTGAAATTGGGCACTGTGCCATTCTTG TCTTTCCGAGATATGTTACGAGCAGCCCCACAGTTGGCGAAATTGCAAGCATGGAGAAGCGT TTACAGTAAAGTTGCTTCTTATATTGAGGACGAACATTTGAGACAGGCCTTTTCATTTCACAG CTTGTTGGTGGGAGGTAACCCATTCGCTACTAGCTCTATTTACACCTTGATTCACGCCCTAGA AAGGGAATGGGGTGTTTGGTTCCCAAGAGGTGGGACAGGCGCTTTAGTGCAAGGAATGATC AAATTGTTCCAAGACTTAGGAGGGGAAGTGGTGTTGAACGCGCGTGTGTCTCACATGGAAAC TACAGGGAACAAGATAGAAGCAGTGCATCTCGAAGACGGCCGTCGATTCTTGACACAAGCA GTTGCATCGAATGCCGACGTCGTTCACACCTACCGCGACTTGCTATCTCAACATCCAGCCGC TGTGAAGCAATCCAATAAATTACAAACAAAAAGAATGTCAAACTCCTTATTTGTTTTGTACTTC GGTCTAAATCACCACCATGATCAACTCGCTCATCACACAGTCTGCTTCGGTCCAAGATATAGA GAATTGATCGACGAGATATTCAATCATGACGGGTTGGCCGAAGATTTTTCTTTGTATCTACAC GCTCCATGCGTCACTGATAGCTCTCTCGCTCCTGAGGGGTGCGGTTCTTACTACGTTTTGGC CCCTGTACCTCACTTGGGTACGGCGAACTTGGATTGGACCGTAGAGGGTCCCAAGCTACGT GACAGAATCTTTGCATACTTAGAACAACACTACATGCCGGGGTTACGCAGCCAATTGGTTAC ACATAGAATGTTCACGCCGTTCGACTTCCGAGACCAACTAAACGCGTATCACGGCTCTGCTT TTTCAGTTGAGCCAGTGCTCACGCAAAGCGCATGGTTTAGGCCTCACAACAGGGACAAAACG ATCACTAACTTATACCTCGTGGGGGCCGGGACCCACCCAGGAGCAGGAATACCAGGGGTCA TTGGTTCGGCTAAGGCGACTGCCGGCTTGATGCTTGAGGACTTGATTTAA	364
PaCrtB	Pantoea ananatis	ATGAACAACCCCTCATTGTTGAACCATGCCGTGGAGACAATGGCAGTCGGTTCTAAGAGCTT TGCTACCGCATCGAAATTGTTTGATGCCAAAACTCGGAGGTCAGTTTTGATGCTCTACGCGT GGTGTCGTCATTGTGACGATGTTATCGATGACCAAACCCTCGGCTTCCAAGCTCGGCAGCCG GCATTGCAAACCCCAGAACAAAGGTTGATGCAATTGGAGATGAAGACCAGGCAAGCTTATGC CGGCAGCCAGATGCACGAACCCGCCTTTGCGGCTTTTCAAGAGGTTGCAATGGCTCATGATA TAGCACCTGCTTATGCCTTTGACCACTTGGAGGGTTTCGCTATGGACGTCAGGGAAGCCCAG TATAGCCAATTGGATGATACCTTAAGGTACTGTTATCATGTCGCAGGAGTTGTGGGATTGATG ATGGCGCAGATCATGGGGGTTCGAGACAACGCTACCTTGGATAGAGCATGTGATTTGGGGT	365

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1795 of 3077

Table 24

Gene	Organism	C.V. Codon Optimized Sequence	SEQ ID NO
		TGGCGTTTCAATTGACTAATATCGCGCGAGATATTGTCGACGATGCACACGCAGGTAGATGC TACTTGCCAGCCTCTTGGCTAGAACACGAAGGCCTCAACAAGGAGAATTACGCCGCCCCTGA AAATCGACAAGCTTTGTCAAGAATAGCGAGAAGATTGGTGCAAGAGGCCGAGCCATATTATT TGAGTGCGACAGCTGGTTTAGCCGGATTGCCATTGAGGAGTGCATGGGCCATAGCCACGGC GAAACAAGTATACCGTAAAATAGGGGTCAAAGTCGAGCAAGCTGGTCAGCAAGCATGGGATC AGCGGCAATCAACCACACAACGCCAGAGAAGTTGACTTTGCTCTTGGCCGCCAGTGGACAGGC TCTAACCTCCCGTATGAGAGCCCATCCCCCAAGGCCAGCTCACTTGTGGCAGAGGCCATTGT AA	
XdCrtR	Xanthophyllomyces dendrorhous	ATGGCAACCCTATCTGACTTGGTTATCTTGTTGTTGGGTGCCTTGTTAGCATTGGGGTTCTAT AACAAGGACAGATTGCTCGGTTCAAGTTCCAGCAGTGCCTCTACCACATCAGGTTCAAGCGC TGCAACTGCTAACGGTAGTAAGCCCACATATTCTAATGGCAACGGGAACGCCTTTAAAGGCG ACCCACGAGATTTCGTCGCGAGAATGAAGGATCAAAAGAAAAGGTTGGCAGTTTTTTATGGG TCCCAAACTGGAACTGCTGAGGAGTACGCCACCCGAATCGCCAAGGAAGCGAAATCTCGTTT CGGCGTTTCGAGTTTGGTGTGTGACATTGAGGAATATGATTTTGAAAAATTGGATCAAGTCCC CGAGGATTGCGCCATTGTTTTTTGCATGGCTACCTATGGTGAAGGTGAACCAACAGACAATG CTGTCCAATTCATCGAAATGATCAGCCAGGATGACCCAGAGTTTAGTGAAGGTAGTACTTTG GATGGTTTGAAGTACGTTGTCTTTGGATTGGGGAATAAGACATACGAGCAATACAATGTTGTC GGCCGGCAGTTGGACGCTAGATTTAACCGCACTTGGTGCGACCAGAGTAGGTGAGCGAGGT GAAGGAGATGATGACAAAAGTATGGAGGAAGATTATCTCGCATGGAAGGATGATATGTTCGC CGCTTTGGCCACCACATTAAGTTTTGAAGAGGGAGCTTCGGGGGAGACTCCAGATTTCGTGG TCACTGAAGTCCCGAACCACCCCATCGAAAAGGTCTTCCAGGGCGAATTGTCCAGCAGAGC CTTGTTGGGCTCTAAAGGCGTTCACGACGCCAAGAATCCATACGCTTCGCCAGTTTTGGCCT GCCGGGAATTATTCACAGGCGGAGATCGAAACTGTATCCACCTTGAGTTTGACATCACCGGG TCCGGAATCACCTACCAAACAGGTGATCACGTCGCAGTTTGGCCATCCAATCCTGACGTTGA AGTGGAAAGGCTTTTGGCAGTGTTGGGTCTCACATCTCCAGAGAAGACGGATGATTATCC AAGTGGTGAGCCTTGATCCAACTTTGGCCAAGGTTCCATTCCCTACACCAACCACCTATGAT GCCGTTTTCAGACATTACCTTGACATTAGTGCAGTGGCTTCACGCCAGACCTTGGCTGTCCT TGCCAAATACGCCCCTTCCGAACAGGCTGCCGAATTCTTAACCAGGTTGGGAACTGACAAGC AAGCGTACCATACTGAAGTGGTTGGAGGTCACCTAAGATTGGCCGAGGTGCTTCAATTGGCT GCAGGGAACGATATTACTGTTATGCCTACAGCTGAAAATACAACCGTCTGGAACATACCATTC GACCACGTTGTGTCTGACGTTTCCGCTTGCAACCCCGATTTTACTCGATAAGCAGTTCTCCT AAGCTTCACCCGAACAGCATTCACGTAACGGCTGTGATACTTAAATACGAAAGTCAAGCCAC TGACCGTCACCCAGCGAGATGGGTGTTTGGTTTGGGGACAAACTATTTGCTTAATGTTAAGC AAGCAGCAAACAATGAAACCACTCCAATGATCTCTGACGGACAAGACGATGTGCCAGAACAC	366

Table 24

Gene	Organism	C.V. Codon Optimized Sequence	SEQ ID NO
		GTGTCAGCGCCAAAATACACGTTGGAGGGACCCCGTGGCTCCTATAAGCACGACGATCAATT GTTTAAAGTGCCAATCCATGTGCGTAGGTCCACTTTCAGACTTCCAACCAGCCCTAAGATCC CTGTGATCATGATTGGGCCAGGTACCGGTGTCGCTCCTTTCAGAGGATTCATCCAAGAAAGA ATTGCCCTTGCCAGAAGGTCCATTGCTAAGAATGGGCCTGACGCCTTGGCGGATTGGGCCC CAATCTATCTTTTCTACGGTAGTCGCGATGAGCAAGACTTCTTGTACGCGGAAGAGTGGCCT GCTTATGAAGCGGAGTTGCAAGGGAAATTCAAGATTCATGTTGCTTTCAGTAGATCCGGGCC AAGAAAGCCAGATGGGAGTAAGATTTACGTGCAAGATCTTTTGTGGGATCAAAAGGAAGTCA TCAAATCTGCCATAGTGGAAAAACGTGCATCTGTTTATATTTGTGGGGACGGTCGAAATATGT CGAAAGATGTCGAGCAGAAGTTGGCCGCGATGTTGGCCGAAAGTAAAAACGGTAGTGCGGC CGTGGAAGGAGCCGCGGAAGTCAAGTCATTGAAAGAGAGATCTCGATTATTGATGGATGTGT GGAGCTAA	
XdCrtS	Xanthophyllomyces dendrorhous	ATGTTCATTTTGGTGTTGTTGACCGGTGCTTTGGGTTTGGCAGCTTTTAGCTGGGCTTCGATT GCCTTCTTTTCTCTTTATTTGGCGCCCCGTCGCAGTAGCTTGTATAATTTGCAAGGTCCAAAC CACACTAACTATTTCACCGGCAATTTTTTAGATATACTCTCGGCTAGGACGGGTGAGGAGCAT GCTAAGTATAGGGAAAAATACGGATCCACGTTGAGATTTGCCGGCATCGCCGGTGCCCCAG TTTTGAACTCAACTGATCCAAAAGTCTTCAACCATGTGATGAAGGAGGCATATGATTACCCAA AACCTGGTATGGCTGCGAGAGTGTTGAGGATTGCAACGGGTGACGGAGTGGTTACGGCTGA GGGTGAGGCACATAAACGACATCGAAGAATCATGATCCCTTCCTTGTCAGCACAAGCCGTTA AGTCCATGGTTCCCATCTTCCTTGAGAAGGGCATGGAGCTCGTTGACAAGATGATGGAGGAC GCAGCCGAGAAAGACATGGCAGTGGGCGAATCCGCAGGTGAGAAGAAGGCTACTCGACTTG AAACGGAAGGCGTTGACGTAAAGGACTGGGTTGGGAGAGCAACTTTGGATGTCATGGCTTT GGCCGGATTTGATTACAAATCAGATTCGCTACAAAATAAGACGAACGAATTGTATGTCGCTTT CGTGGGGCTTACAGACGGATTTGCACACCCACCTTGGACAGTTTAAGGCCATCATGTGGGATT TCGTGCCGTACTTTCGCACAATGAAGAGAAGACATGAGATTCCCTTGACCCAGGGTTTGGCT GTTAGTCGCCGTGTTGGCATTGAACTAATGGAGCAAAAGAAGCAAGCAGTGTTGGGTTCGGC ATCGGATCAAGCCGTGGACAAAAAGGACGTGCAGGGACGCGACATTCTTAGTCTCTTGGTCA GAGCGAACATTGCCGCAAATCTTCCCGAATCTCAAAAGTTGTCGGACGAAGAGGTTCTTGCT CAAATTTCGAATTTGCTATTCGCGGGTTATGAGACTTCAAGTACTGTTCTCACATGGATGTTC CATCGATTGTCGGAGGACAAAGCAGTGCAAGATAAATTAAGGGAGGAAATCTGTCAAATAGA TACAGACATGCCAACACTAGACGAGTTGAATGCATTGCCGTACTTGGAGGCGTTTGTGAAAG AGTCACTAAGACTCGATCCCCCGTCTCCATATGCTAATAGGGAGTGTTTGAAGGATGAAGAT TTCATTCCATTGGCCGAACCTGTCATTGGACGCGATGGGTCCGTGATCAATGAAGTCAGAAT CACAAAGGGGACAATGGTGATGCTTCCATTGTTCAACATCAATAGATCGAAATTCATATACGG TGAGGACGCCGAAGAATTTAGACCAGAAAGGTGGTTGGAGGATGTTACCGACTCCCTTAACA	367

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1797 of 3077

Table 24

Gene	Organism	C.V. Codon Optimized Sequence	SEQ ID NO
		GCATAGAAGCTCCATATGGCCATCAAGCGTCCTTCATATCAGGACCAAGAGCCTGCTTTGGA TGGCGGTTCGCGGTCGCAGAAATGAAAGCATTCTTGTTCGTAACACTAAGGAGAGTGCAATT TGAGCCGATCATAAGCCATCCAGAGTACGAACATATTACTCTCATTATCAGCAGACCCCGGA TCGTTGGCAGAGAAGGAAGGAAGGATATCAGATGCGTTTGCAGGTAAAGCCAGTGGAGTAA	
AaCrtZ	Agrobacterium aurantiacum	ATGACAAACTTCTTGATCGTTGTGGCCACCGTGCTTGTTATGGAATTGACCGCCTACTCGGT GCATAGATGGATAATGCACGGACCATTGGGTTGGGGATGGCACAAATCACACCATGAGGAA CACGATCACGCCTTAGAAAAAAATGACCTCTATGGCTTGGTGTTCGCTGTCATTGCGACCGT TTTGTTCACCGTCGGTTGGATCTGGGCACCAGTGTTGTGGTGGATTGCACTTGGTATGACGG TTTATGGCCTCATATATTTTGTGCTCCACGATGGTCTAGTTCACCAAAGGTGGCCTTTCAGAT ACATTCCCAGGAAGGGTTATGCACGGCGATTATATCAAGCCCATCGTCTTCATCACGCTGTG GAGGGAAGGGATCACTGTGTCAGCTTTGGGTTCATCTACGCACCACCCGTTGACAAGTTGAA ACAAGATTTGAAAATGTCTGGTGTCTTGCGGGCTGAGGCCCAAGAGAGGACGTAA	368
AaCrtW	Agrobacterium aurantiacum	ATGAGTGCCCATGCCTTGCCAAAGGCAGATCTAACTGCCACTAGTTTAATTGTGTCGGGCGG TATTATCGCGGCTTGGTTGGCATTGCATGTACATGCATTGTGGTTCTTGGACGCGGCAGCGC ACCCTATCCTCGCAATCGCCAATTTTCTTGGACTAACGTGGTTGTCAGTAGGTCTATTTATCA TCGCCCACGATGCGATGCACGGATCCGTAGTCCCTGGTAGACCTAGAGCAAATGCTGCAAT GGGCCAGTTGGTGCTTTGGTTGTACGCTGGCTTCAGTTGGAGAAAGATGATTGTTAAGCATA TGGCACATCACCGTCACGCAGGTACCGATGACGACCCAGATTTTGATCATGGAGGTCCAGT GCGATGGTACGCACGCTTCATAGGAACCTACTTTGGCTGGAGAGAAGGTTTGCTTTTGCCAG TCATAGTAACCGTCTACGCTTTAATACTTGGAGATCGATGGATGTACGTCGTGTTCTGGCCAT TGCCATCAATCTTGGCCTCTATCCAATTGTTCGTTTTTGGGACTTGGTTACCACATAGACCCG GTCACGATGCATTCCCAGATAGACACAACGCCCGATCATCCAGAATATCCGATCCAGTTTCTT TGCTAACTTGCTTCCACTTCGGTGGTTATCACCACGAACACCATTTGCATCCGACGGTCCCAT GGTGGAGGTTGCCTAGCACTAGAACTAAAGGTGATACCGCTTAA	369
CnTPS 1	Callitropsis nootkatensis	ATGGCAGAAATGTTCAATGGAAACAGCAGTAATGACGGGAGCTCATGTATGCCGGTAAAGGA TGCATTGAGACGTACGGGCAACCACCACCCAAATTTGTGGACTGACGATTTCATTCAATCGC TTAATAGTCCATACTCGGACTCCTCCTACCATAAACATCGCGAAATATTGATCGATGAGATTC GCGATATGTTTTCAAATGGGGAAGGGGACGAATTTGGCGTGTTGGAAAATATTTGGTTTGTC GATGTTGTTCAACGGTTAGGAATCGACCGTCACTTTCAGGAAGAGATTAAAACAGCGTTGGA CTACATCTACAAGTTTTGGAACCACGACTCGATATTTGGTGACCTTAACATGGTCGCTTTGGG TTTCAGAATCCTAAGATTGAACCGTTACGTGGCAAGCAGCGACGTTTTCAAGAAGTTTAAGG GTGAAGAGGGCCAATTTTCAGGCTTCGAAAGTTCAGACCAAGATGCGAAGTTGGAAATGATG CTAAACTTGTATAAAGCAAGTGAGCTAGATTTTCCAGACGAGGATATCCTTAAGGAAGCTAGG	370

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1798 of 3077

Table 24

Gene	Organism	C.V. Codon Optimized Sequence	SEQ ID NO
		GCCTTTGCGAGTATGTACTTAAAGCACGTAATCAAGGAATACGGTGATATCCAAGAGAGCAA GAACCCATTGCTTATGGAGATAGAATACACATTCAAGTACCCATGGCGTTGTCGTTTGCCTCG ACTTGAGGCTTGGAATTTCATTCACATTATGAGACAACAAGACTGTAATATTTCATTGGCTAAC AACCTTTACAAGATTCCGAAGATATACATGAAAAAGATCTTGGAGTTGGCCATCTTGGACTTT AATATTCTCCAGTCACACAACACCAACATGAAATGAAACTCATCTCCACGTGGTGGAAGAACTCT AGCGCTATCCAATTGGACTTTTTTCAGACACCGCCACATCGAAAGCTACTTCTGGTGGGCTTC ACCGTTGTTCGAGCCCGAGTTTTCTACATGTCGGATCAACTGTACGAAGTTGTCAACTAAAAT GTTTCTACTTGACGATATATACGACACATACGGTACTGTGGAAGAACTAAAGCCGTTTACGAC CACATTAACGAGGTGGGACGTGTCTACAGTCGACAACCATCCAGACTACATGAAGATCGCTT TTAACTTCAGCTATGAAATCTACAAGGAGATCGCCTCCGAGGCGGAACGTAAACACGGTCCA TTCGTGTATAAATATTTGCAATCATGCTGGAAGTCCTATATAGAAGCTTACATGCAAGAAGCA GAGTGGATTGCTTCAAATCATATTCCAGGCTTCGATGAGTACCTCATGAATGGTGTAAAGTCA TCCGGCATGAGAATATTGATGATCCACGCATTAATCCTAATGGATACACCATTGAGCGACGA GATTTTGGAACAATTAGATATACCGAGTTCGAAGTCACAGGCCTTGTTGTCGTTGATTACCAG ACTCGTGGATGATGTCAAGGACTTTGAGGACGAACAAGCCCATGGGGAAATGGCATCCTCC ATTGAATGTTACATGAAAGATAACCACGGTTCAACAAGAGAAGATGCATTAAACTACTTGAAG ATCCGAATCGAGTCGTGTGTCCAGGAGTTGAATAAGGAATTGTTAGAACCTTCGAATATGCA CGGTAGTTTTAGGAATCTTTATTTGAATGTTGGCATGCGTGTCATATTTTTCATGCTCAACGAT GGGGACTTGTTCACTCATTCGAACAGGAAGGAAATTCAGGATGCCATCACCAAGTTCTTCGT CGAACCGATTATACCTTAA	

Table 25

Gene	Organism	Native Cv sequence	SEQ ID NO
ERG10	Candida viswanathii	ATGACCCTCCCACCAGTCTACATCGTCTCCACAGCAAGAACTCCAATTGGTTCCTTCCAAGG TTCGTTATCTTCCTTGACCTACTCCGATTTGGGTGCCCACGCCGTTAAGGCAGCTTTAGCCAA GGTCCCACAGATCAAGCCACAGGACGTCGACGAGATCGTCTTTGGTGGTGTCTTGCAAGCC AACGTTGGTCAGGCACCAGCCAGACAAGTCGCCTTGAAGGCAGGCTTGCCAGACTCTATCA TTGCCTCCACCATTAACAAAGTCTGTGCCTCCGGTATGAAGGCCGTCATTATCGGTGCCCAA AACATCATCTGTGGAACCAGTGACATCGTCGTTGTTGGTGGTGCCGAATCCATGTCTAACAC TCCATACTACTTGCCAAGCGCCAGAAGCGGGGCCAGATACGGTGACGCCGTCATGGTTGAT GGTGTCCAGAAGGATGGTTTGTTGGATGTCTACGAAGAAAAATTGATGGGTGTTGCTGCTGA	371

Table 25

Gene	Organism	Native Cv sequence	SEQ ID NO
		AAAGTGTGCCAAGGACCACGGATTCAGCAGAGAAGACCAGGACAACTTTGCCATCAACTCTT ACAAGAAGGCTGGCAAGGCTTTGAGTGAAGGTAAGTTCAAATCAGAAATCGCCCCCAGTCACC ATTAAAGGATTCAGAGGCAAGCCAGATACTGTTATTGAAAATGATGAAGAGATTGGCAAATTC AACGAAGACAGACTCAAGTCTGCCAGAACTGTCTTCCAAAAGGAAAACGGTACTGTTACTGC TCCAAACGCTTCTAAATTGAACGATGGTGGTGCTGCCTTGGTCTTGGTTTCTGAAGCCAAGTT GAAGCAATTGGGCTTGAAGCCATTGGCCAAGATCTCTGGTTGGGGTGAAGCTGCCAGAACT CCATTCGATTTCACCATTGCCCCAGCTTTGGCTGTTCCAAAGGCTGTCAAGCACGCTGGTTT GACCGTTGACAGAGTCGACTTCTTTGAATTGAACGAAGCCTTCTCCGTTGTTGGTTTGGCCA ACGCTGAATTGGTCAAGATCCCATTGGAAAAATTGAACGTCTACGGTGGTGCTGTCGCCATG GGTCACCCATTGGGTTGCAGTGGTGCTAGAATTATTGTCACCTTGTTGTCCGTCTTGACCCA AGAAGGCGGTAGATTTGGTGCTGCTGGTGTCTGTAACGGTGGTGGTGGTGCCTCTGCCATC GTCATTGAAAAGATTGACTCCGATGCCAAGTTGTAA	
ERG13	Candida viswanathii	ATGACTAACGCACCACACAAAACATTGGTATCAAAGGAATCGAAGTCTACATTCCAGGCCAAGC TGTCAATCAATCAGACTTGGAAAAATTCGACGGCATCCCAGCCGGCAAGTACACCATCGGCT TGGGCCAGACAAACATGGCTTTTGTCAACGACAGAGAGGACATCTACTCCATCGCCTTGACT GTTGTCTCCAGATTGATCAAGCATTACAATGTTGACACCAACAACGTCGGTCGTTTAGAGGTC GGTACCGAGACCTTGTTGGACAAGTCCAAGTCTGTCAAGTCCGTGTTGATGCAATTGTTCCC CGACAACAACGACATTGAAGGTATCGACACTGTCAACGCCTGCTACGGTGGTACCTCCGCTG TCATCAATGCCATCAACTGGATCGAGTCGTCTTCCTGGGACGGCAGAGACGCCATTGTTGTT GCTGGTGACATTGCTATCTACGACAAGGGTGCTGCTAGACCAACTGGTGGTGCTGGTTCCAT TGCTATGTTGATTGGTCCTGACGCTCCTATTGTCTTTGACTCTGTGCGTGGCTCATTCATGGA ACACGCCTATGATTTCTACAAGCCAGACTTCACCAGTGAATACCCAGTTGTCGATGGTCACTT CTCCTTGTCCTGTTACGTTAAGGCTGTTGACAACTGTTACAAGAACTACTCCAAGAAGGTCAC CGGCAACGTCGACAAGACCGTCGGCGTTTACGACCACTTTGACTACAACGCTTTCCACGTGC CAACCTGTAAGTTGGTCACCAAGAGTTACGCCAGATTGTTGTACAACGACTACAAGTCGGAT CCATCCAAGTTTGCTGACTTGATTGACGAAAGCACCAGACAACACATTGACAGTTTGTCCTAC GAAGCATCCTTGACCGACAAGGTCTTGGAAAAAACTTTCGTCACATTGGCCAAAGAAGAGAC CAAAAAGAGAGTGCAACCAGCCTTGCAAGTGCCAACCAACACCGGTAACATGTACACTGCTT CGTCATGGGTCTCCTTGGCCTCGTTGTTGTACTATGTTGGCGCTGAAAACTTGAAAGAAAAG AGAATCGGTCTCTTCTCCTACGGTTCAGGTTTGGCCTCCACCTTGTTGTCCGTCACTGTTGTC GGCGACGTTTCTCCAATCACCAAGGTCTTGGACTTCGACTACAAATTAGGCGAAGGAAGAAA GATCCAGAGCCCAGAAGAATACTTGGCTGCCATTGAATTGAGAGAAAAGGCACACTTGCAAA AGAGCTTCAAGCCACAAGGTTCCCTCGATAACTTGAGCCAAGGCACCTACTACTTGACTGAG GTTGACGACAAGTTTAGAAGAAGTTATGCTATAAAGGAATAA	372

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1800 of 3077

Table 25

Gene	Organism	Native Cv sequence	SEQ ID NO
HMG1	Candida viswanathii	ATGCTAAGTTTCATCACCGAAGCAAGTGGCAGAATAGCCCAGACTGCAGCCCACAGACCAAT TCACTTCATGGTGATACCTGCGTTACTAGCATCCATAGCATACTTGTCCATCATCGATGATTA CGTCCCCGAACACATCAGAGCCCAGTACATGTCAGGCGTCAGCTATTTCCACCCTCAAGGGT CCGCCTCGGATTTGGACAAGTGGATCGAAATCCACGACACTACCCAATACGCCAACGCCAAC CAAATCAGTGTCATCCCTTTGAGATTCAGACGATTCCATGACTCAATCCCACAGATCGCTAAC GCCATCAAGATTTCAAACAATGAACAGATCTTGATTGTTCCTTCGGATAAAGCAGAGTCGACA GTAGAAGGTTTGAGCGAGATCACTGAGAATGGCATCACTTGGAGGGCTAGAAATAACGACAA GTTGTCCAAGTATTATGATTACGCCAGGTACGGGTTGCTTAGAATCCAGGACGCTATACACA ATGCTGATAATTTCGACATCCTCTTGGTGTTTGTTGCCTACCTTGGAATGTGGTACTCACTTAT CAAAGTGTTTATTGACATGAGACGCATTGGCTCCAAGTTTTGGTTAGCCTTTGGCACGTTGAC TTCCTCGACTTTTTCATTCTTGTTTGCCTTGGTCATTTCAAACAAATTTTTGGACGCTAAAGTTT CCTTGAGAAGTCTTTCTGAATGCATCCCATTTTTGGTTGCCATTATTGGGTTCAAACATAAAGT GGCCATCACGACTAGTGTTGCCCAGTCATCCACTTCTTCACCCGAGGATGTTCCTCATGTCG TGGGCAAAGCTGTCTCAGACCAGTGTTTGTTTATCTTGAGAGACCATTTGGTTGTCATTATTG GATTTCTCGCCTGTGCGGCTTATGGCAATGAATTGAAAGGTTTAAGAAACTTCTGTATCTTGG GGGCATTGATTTTGTCCTTTGACATTGTTTCCGTTTATACTTTCTTTTCTGCGATTTTGGCCTT GAAAGTCGAGATTAACCGTGCCCGCCGTACTCAAGATTTACAACACGTATTGGAAGAGGACG GCATTTCATCCTTAGTTGCCGCTAGAGTTGCCGAACGAAATGCAACTATTGAACACCCCAAC GAAACCAACTTCTTCTCATCTAATAATTCTTCCATTGTTTATTTCAAAGTTATTATGAGTCTTGG GTTTTTCGCCTTCCATGCATTTTGGTTGGGCAGCTCGTGGTTATACAACACTTCTGATGGCGG TAGCCATGGAAGCTTTTCGTTCCTCAGCAACATCCCATTATTGACACAAGACATCTCCAACTC CATCCCAATTGGCAGAAGAGGAACCGTGGTTACCATTTTGCCAACCACCTTCTACATGCCAT CCGGTATCATCATCCAATTTGAAGATATGGTCTACTTAGCATTGAGCAAGATCAGCAGTGCTA TCAGAGACAGTTTGATTTCCAAGTGTATTGTGTTTGCATTGACCATTTCCATTGTCACCAACAT CTACTTCTTGAATGCTGCCCGTTTCCAAGTTTCAGCTACCCGTAAGTTGATTGACCAAGAAAT GTCTCGTCCAAAACAAGCCGCCGCCGCTGCTGCTCCATCTGCTGCAAAATCAGTG GCGCCAGAAGAAGACGAAGACCTCCAGTGAAGAACTTGAAATCAAAGCTCCTGTAAA GCCGTTATCACTTGAAGAATGTACCAGAATCCTCAAAGAGGGCAAAGTCAAGACCTTGTCCA ACGCTGAAGTGTCTAGCTTGGTTGTTGGCGGTAAATTGCCATTGTACGCTTTGGAGAAACAA TTGGGCGACAACAAGAGAGCCGTGGCTGTTCGTCGTAAGGCTATTGCAAAATTAGCTAACGC GCCAGTATTAGAAACCAACCGTTTGCCTTACTCCCACTATGATTATGATAGAGTCTTTGGTGC TTGTTGTGAGAATGTTATTGGGTACATGCCAATCCCTGTTGGTGTTGCTGGACCATTAGTGAT TGATGGCAAGCCATACCACATCCCCATGGCCACTACCGAAGGTTGTCTTGTTGCCTCTACTA TGCGTGGCTGTAAAGCTATTAACGCTGGAGGTGGTGTTGAAACCGTCTTAACCAAAGATGGT ATGACTAGAGGTCCTTGTGTTCGTTTCCCAACTTTGAAAAGAGCCGGTGCTGCAAAGTTGTG	373

Table 25

Gene	Organism	Native Cv sequence	SEQ ID NO
		GATTGACTCCGAAGAAGGTCAAATCACCATCAAAAAGGCATTCAACTCGACTTCTAGATTTGC TCGTTTGCAACATATTCAAACTGCCCTCGCTGGTACCTCGTTGTTTATTAGATTTAGAACTACT ACTGGTGATGCTATGGGCATGAACATGATTTCCAAAGGTGTTGAATATTCCTTGAAGCACATG GTTGAAGAATGTGGCTGGGACGATATGGAAGTTATCTCTGTTTCTGGTAACTACTGTACCGAT AAGAAACCAGCTGCTATTAACTGGATTGAAGGTAGAGGTAAGTCTATTGTTGCTGCTGCTACC ATCCCTGCTGATGTTGTCACTAAAGTCTTGAAATCCGATGTCGATGCCTTAGTTGAATTGAAT GTTTCCAAAAACTTGGTTGGCTCAGCCATGGCCGGTTCAGTTGGTGGTTTCAATGCTCATGC AGCCAACTTGGTCACTGCTGTTTACTTGGCCTGTGGTCAAGACCCCGCACAGAATGTCGAAT CCTCAAACTGTATTACCTTGATGGAAAAGGACAAACAGACTGGCGACTTGGTCATTTCTGTTT CTATGCCATCTATTGAAGTTGGTACGATTGGTGGTGGTACCATTTTGGAACCACACAGCAGCC ATGTTGGACTTGTTGGGCGTGCGTGGCCCTCACCCAACCAATCCAGGTGACAATGCAAGAC AATTGGCCAAGATTGTTGCTTCGGCAGTGTTAGCAGCTGAATTGTCACTTTGTTCTGCTTTGG CCGCTGGTCACTTGGTGCAATCCCACATGCAACACAACCGTAAAGGAGCCACCCCAGCTGC GGCCCCTGCAATCAGCAACGGAAGTGCCAAGGGAACCAAAACCAATGGGTCTATAAATGGG AAAGACTTGAAGCGTCTCAAAGACGGATCCGTTACATGTATAAAGTCATAG	
ERG12	Candida viswanathii	ATGTCAGTCTCGCCTTTCGTTGTCAGTGCTCCAGGGAAAGTCATCATCTTTGGAGAACATTCA GCAGTTTATGGGAAACCCGCCATTGCAGCCGCCTTGAGTTTGCGATGTTACCTACTCGTTTC CCCGTCAGTTGATGATGCAAACACAATCAGATTGCAGTTCCCAGACATCCAATTGGACCATTC CTGGAACATCAACGATATCCCCTGGGATGAAATCAAACCGTTCGTTAAATATGACGCCAATAA CAAACCACTCACCCCATCGGAGTTGGTACCAGAGATCCTCGATAAGTTGTCTCCCTTGTTGA CTGATTTCGATAACAAAATGCATTACTACGCTTGTTTTTGTTTCCTTTACTTGTACGTCAATTTA TGTTCCACTGAAACTCCAGGAACCACCTTCATCGTCAGATCCACCTTGCCAATCGGTGCCGG GTTAGGGTCGTCTGCATCCACGTCTGTGTGCTTGTCGTCTGCATTAGCATTGTTAGGCGGCT GTATCAGTGAGCCTTCGATATCAGCCACCGACACAAGATCCTCAACGAAGACATACCCGATTTG GAGTTCGTCGATAAGTGGTCCTTGATAGGTGAGAAGTGCTTCCACGGGAACCCATCCGGTAT CGACAACGCAGTAGCCACGTTTGGCGGTGCCGTCATGTTTCAAAGAACTTCGGCTCCTGAG CAACCATCCATCAGAACCAACATGAGAAACTTCCCAGCTATCAAACTCTTACTTACAAACACC AAAGTCCCAAAAAGTACAGCTGATTTAGTCGCCGGTGTTGGCAAACTCAACGCGGAGTTCAA CCCGATAACCACGTCGATTTTGACCGCTATGGAGCATTTGTCACAGGAAGCTTACAAAGTCA TGATCACTCCGGGGTTCGGCAAGGACGAAACAAACACGTTGCGTAAGTTGGTCAATATCAAC CATGGTCTATTGGTAGCGTTGGGCGTGTCACATCCGTCGCTAGAAACCGTCAAGATCATTGG TGACAGACACAAGATTGGCGCTACTAAGTTGACTGGCGCTGGTGGTGGCGGGTGTGCGATC ACTTTGGTTAATGATAATGTCGAGGAGTCTGTTATCCAAGAAGCTATCAAAGAGTTTGCAACT GAGGGATATGAGAGTTTTGAAACGTCGTTGGGTGGTAAAGGTGTCGGTGTCCTATTTAGTGG	374

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1802 of 3077

Table 25

Gene	Organism	Native Cv sequence	SEQ ID NO
		TGATGTTTCTGATAATGGCAAGTTTTCTCAATCTGTGTTCTGCAACTATCCCGACCGTGCTAG CATTGAAGATGCTTTAGGAATGATAAACGTTAAAGAATGGAAGTTCTGGTGA	
ERG8	Candida viswanathii	ATGTCAAAAGCTTTTAGTGCCCCTGGTAAAGCATTGTTAGCTGGCGGATACTTGGTTCTTGAG CCAACATATGATGCATATGTGACGGCATTGTCCTCGAGAATGCACGCTATAATAACACCCCA GAAGCCAGCATCCATTTCAAAAATCAAAATCAGTTCGCCTCAGTTTGCCAATGGTGAGTGGG AGTATCACGTCACGTCTAATGAGAAACCCAAGGACATCAAGTCCAGAAGCAATCCGTTTTTG GAGGCTACCATTTTCATTGTTTTGTCTTATATCCAACCAACAGAACCATTCGATCTCGATTTGG TCATATATTCAGACCCAGGATACCATTCACAAGAACACACCACGCAGAAAGTGTCCAGCAAC GGCAAGAAAAAGTTCCTTTATCATTCGCGTGCAATTAACGATGTTGAGAAGACAGGATTAGGT TCTTCGGCCGGTCTAGTTTCTGTCGTCACTACCAGTTTGTTATCCTACTTTATTCCGGGAATT GAAGAGTCAAACAAGGACATGTTGCACAATGTAGCACAGATTGCCCATTGTTTTGCCCAGAA GAAGATTGGGTCCGGGTTCGATGTGGCCACTGCCATCTATGGTCTGATTGTCTACCGAAGAT TCCAGCCAAGCTTGATCAACGACGTTTTCGAGATCTTAGAGGAAACCCCGGGTAGATTCCCC GGTGCTTTGAAGTCGTTAGTTGAGTCTAATTGGGAATTCAAACATGAAAGGTGCGTACTACCA CCAAAGATCAAGCTCTTGATGGGAGATATCAAAGGAGGATCAGAGACTCCGAAATTGGTTTC CAAAATACTCCAATGGAAGAAGGACAAACCAGAAGAAAGTGGGTTGGTGTATGATCAACTTA ATAGCGCAAATGTGGCGTTCATGAAGAAGATCAGTACATTGAATCGAGTCAGGTCCAA GAGATTGACGAATTGTCTGATTACATAAGCGCGGTACGCAAAGGGTTGCAAGAATTGACAGA AAAATCCAAAGTCCCAGTTGAACCACCTGTGCAAACTGAATTGTTGGACCGCATTGCAAAACT CCCAGGTTGTCTAGGTGGGGTTGTGCCCGGTGCTGGTGGGTACGATGCTATAGCTGTGCTT GTGTTGGAAAAGGAGGTTGAGAACTTTAGAAACAAGACACTTGAAAACCCAGAGTACTACCA CAACGTGTACTGGGTTGACTTGGAGGAAGAAACCGAGGGTGTAGTGGTGGAAAACTGGGAG GACTATATCGGTTTATAG	375
MVD1	Candida viswanathii	ATGTATTCAGCTTCAGTCACAGCACCTGTCAACATAGCGGTATGTATGTTCCCACCTAGTTGC TCCTCCCCGACTCATCACTAACACCCTTATAGACCCTCAAGTACTGGGGGAAACGCGACAAG ACCCTAAACTTGCCAACCAACTCCTCCATATCAGTCACCCTCTCCCAGGATGATTTGCGAACC TTAACCACAGCAGCAGCGTCCACCACTTTCGACAAGGACCAATTGTGGCTCAATGGCAAGTT GGAGTCGTTGGACACTCCGAGAACGCAAGCGTGCTTGGTCGATTTAAGGAAATTGAGAGCC GATGTTGAGCAAGCCAACGCTGACCTCCCAAAGCTTTCCACGATGAAGCTCCACATTGTTTC GGAAAACAACTTCCCCACCGCCGCTGGTTTGGCTTCCTCGGCTGCTGGGTTTGCTGCGTTG ATCACAGCCATCGCCAAGTTGTACGAATTGCCCCAAGACATGTCGGAGTTGTCCAAGATTGC TCGTAAAGGTTCTGGTTCTGCGTGCAGATCCCTCTTTGGCGGGTTTGTTGCGTGGGAGATGG GCGACGCCGCCGACGGACAAGACTCCAAGGCAGTTGAAGTTGCGCCTTTGGACCACTGGCC TAGCATGAGGGCGGTTATTTTGGTGGTCAGCGACGACAAAAAGGACACCCCAAGCACCACC	376

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1803 of 3077

Table 25

Gene	Organism	Native Cv sequence	SEQ ID NO
		GGTATGCAAGCCACCGTGCAAACCTCCGACTTGTTTGCCCACCGCGTCACCAAAGTCGTCC CCCAGTGTTTCGAAGAGATGAAGAAGGCCATCGTCGCCAAGGACTTTCCTAAATTTACCGAG TTAACCATGAAGGATTCCAACTCGTTCCACGCTGTTTGTTTGGACTCGTACCCTCCAATCTTC TACCTCAACGACACCTCCAAGAAGATCATCAAGTTGGTGGAGGCCATCAACAAGCACGCCAC CATCGCTGCCTACACGTTCGACGCGGGCCCCAACGCCGTCATCTACTACGACGCCGCTAAC GAAGACGAGGTCTTGTCCCAGCTCTACAAGTCGTTTGGCCACGTTCAAGGATGGAAGAAGC CTACACCGCGGAAACAGCCGTTGCCGGTGTTTCACGTATTATTCAAACATCTATTGGTCAAG GACCGCAAGTTACAAACGAATCATTAATCAACGAAAGTGGGTTACCAAACTAA	
IDI1	Candida viswanathii	ATGTCATCAGAATACGCTAAACTAGTGGCAAGTTTCTCCCCAAACGACATATTGGCTAAATGG CCAGAGGTCACTCCATTGAAGAAGATATCAGGCATCCCAAGATCAGCCGAGTCAGACAGCTC CAACGGCTCCCACAACAACACAGAACTTTTCAATGGCCACGACGAGGAGCAAATCAGATTGA TGGAGGAGCTTTGCATTGTTTTGGACTACGACGACAAGCCTGTTGGGGCAGGAACAAAGAA GTTGTGTCATATCATGGACAATATCAACGAAGGATTGTTGCACCGTGCCTTCTCGGTGTTTTT ATTCAATGAGGACGGGAAATTATTATTACAACAACGTGCCGATGAAAAAATCACCTTCCCTGC CATGTGGACAAATACTTGCTGTTCCCATCCTTTGTGTTCCTAGTGAGTTGGGGGTTGATGC CGATGCCAAAGATGTCAACAACTTGGATAACGCCGTGCATGGAGCCAAGGTGGCCGCACAA AGAAAATTGGACCATGAGTTGGGCATCCCATTCAGTGACACTCCCTTGGACCAGTTCACGTA CTTGACTAGAATTCACTATAAATCAGCCAGTGGTGCAGAAGACTCCAAATGGGGTGAGCATG AAATAGACTATATTTTGATATTGAAGACCAAGAATGACATCACCATCAATGCCAACTATAATGA AGTCAAGGATTATAAGTATGTGGACGCCAAGGAGTTGCAGGAGATGTTTGAAGATAAGGACT TGGTGTTCACCCCGTGGTTTAAATTGATCTGTCAATCTTTCTTATTCAAATGGTGGAATAATTT AAGTGATTTAGAAAAGTACCAAGATACTGAGATACATAGATTACTTTAA	377
ERG20	Candida viswanathii	ATGTCTGATAAAGCAGCCGCTAGAGAGAGATTCCTCTCTGTTTTTGAGTGTGCCGTCGAGGA ATTGAAAGAAGTCTTGGTTTCTCACAAGATGCCGCAAGAAGCAATTGACTGGTTTGTCAAGAA CTTGAACTACAACACCCCCGGCGGTAAGTTGAACAGAGGTTTGTCTGTTGTCGACACCTACG CTATCTTGAACAACACCACCGCTGACAAGTTGAACGATGAACAATACAAGAAGGTCGCCTTG TTGGGCTGGTCAATTGAATTGTTGCAAGCTTACTTTTTGGTTGCCGATGATATGATGGACCAA TCCAAGACCAGAAGAGGACAGAAATGTTGGTACTTGGTCGAAGGTGTTGGAAACATTGCAAT TAATGACTCCTTCATGTTGGAAGGTGCCATTTACGTCTTGTTGAAGAAGCACTTCCGTCAAGA TCCATACTATGTCGACTTGTTGGACTTGTTCCACGAAGTCACCTTCCAGACCGAATTGGGTCA ATTATTGGACTTGGTGACTGCTGATGAAGAAGTCGTCGACTTGGACAAGTTCTCCTTGGACA AGCACTCGTTCATTGTCATTTTCAAAACCGCATACTACTCCTTCTACTTGCCTGTTGCTTTGGC CATGTACATGAGCGGTATCAGCAGCGAAGAAGACTTGAAGCAAGTCAGAGATATCTTGATCC CATTGGGTGAGTACTTCCAAATCCAGGACGATTTCTTGGACTGTTTCGGAACCCCAGAACAA	378

Table 25

Gene	Organism	Native Cv sequence	SEQ ID NO
		ATTGGCAAGATCGGTACTGATATCAAAGACAACAAGTGTTCCTGGGTGGTCAACCAAGCTTT GTTGCATGCTACTCCAGAACAACGTAAGTTGTTGGACGACAACTACGGTAAGAAAGACGACG AGTCTGAACAGAGATGCAAGGACTTGTTCAAGTCCATGGGCATTGAAAAGATCTACCACGAC TACGAAGAGTCAATTGTTGCTAAATTAAGAGAACAAATCGATAAAGTTGATGAATCAAGAGGT TTGAAAAAAGATGTCTTGACCGCTTTCTTGGGCAAGGTTTACAAGAGATCCAAATAG	
BTS1	Candida viswanathii	ATGTCATTCAACATCGATAGCTTAATCCAACCAGGCGCCGCCTACGACCCATCTATGACCGA CGCCATCATGAAGCCATACAGGTACATCTCCGAGGTCCCCAGCAACACCACAACGTGCGTA CGCGCTTCTTGCTTGCATTCAACGAGTTGTTCTACGGCATGAAGAACGAGGACTTGCTCCAC CGCATCAGCCACATCATCTCCGTGTTCCACAACTCGTCCTTGTTGATCGACGACATCGAGGA CGACTCGCAGTTGCGCCGCGGCATGCCGGTGGCACACGTCAAGTACGGCGTCCCGTTGAC CATCAACTGCGGCAACATGATGTACTTCGTCGCGGTCCAGAAAGCGATTGACTTGGCGGGC GAGGCTGGGTCCGCGGAGTTGAAGTTCGAGACGTCGCAGATCTTGGTTGACGAAATGATGA ATGTTCACCATGGACAAGGCTTGGATATTTACTGGAGAGACTACTTGAAGGACCTCGAGCAC TTGCCTGAGATCGAGGATTACTTGGGCATGGTCAAGGACAAGACTGGGAGCTTGTTCAGATT GGCCATCAAATTGTTGCTGTTGCATTCGGACGTGGGCGAGGACAATGGTTTGGTCGCGATTG CCAACTTGTTGGGCATAATCTACCAAATCAGAGACGACTACTTGAACTTGGTCGATATCAAGT ACTCCGCCATGAAGGGCGTTACTTGTGAAGACTTGATTGAAGGTAAGTTATCCTTGCCTATTT TGCATTGCTTGAGAACCACGACTAATTCCCCGGTTCACGAGATCCTATACAACTACAACACCA GCGCTGAGCGCGCCAAACAGAACGCCTTGATTGAAGAGTGCTTGACTTACATGAAGAACAAG TCGAGATCCTTGCAATATACTTTGGACTTGATCAAGACCTTGGAGCGCAAGATCAAGGCTAT GATGACCAAGTATCCAAACCTGGAGAACTCCGGCTTGATAAAAATCATTGACAGGTTGTGTG ATCTT	379
CRC1	Candida viswanathii	ATGGACGACGTTGATTCTGCTTTAGCCGATAATGTTAAATCCTTCGCTGCCGGTGGTTTTGGT GGTATTTGTGCCGTGTTGACCGGTCATCCATTCGACTTGGTCAAGGTCAGATTGCAAACTGG GTTGTACAAATCGTCAGTGCAATGTGTCAAAGAAACAATAGCCAAAGACGGGTTGTTTGGGT TATACAGAGGTGTCTTGCCTCCATTGTTGGGTGTCACCCCAATGTTTGCTGTTTCCTTCTGGG GTTACGACGTTGGTAAGAAATTGGTTTCTTCCTTCACTGGCAAATCAGTCGACAAGTTTGAAA TCAAAGACATTTCGACTGCCGGTTTCATCAGTGCCATCCCAACCACCTTGGTTGCCGCTCCA TTCGAAAGAGTCAAGGTCATGATGCAGATCCAAGAAGGTGCCAAGTCCAAGTCCATGGGTGC CGTTGTTGCTGAAATGTACAGAACCGGTGGTATCAGATCCATCTTCAAGGGTACTGTGGCTA CTTTGGCCAGAGATGGTCCAGGTTCTGCCTTGTATTTCGCTACTTATGAATGGGTCAAGAAG GAATTGACTGCCCCTGGTGAAGACTTGTCCTTGTTTGCTATTACCACTGCTGGTGGTTTCGCT GGTATTGCCATGTGGTTGGGTGTTTTCCCAATTGACACCATCAAGTCTACCCAGCAGTCTTCC AACGTCAAGGTCTCCATTGTGCAGGCCACTAAGAACATCTACGCTAAAGGTGGTATCAAGGC	380

Table 25

Gene	Organism	Native Cv sequence	SEQ ID NO
		TTTCTTCCCTGGTGTTGGTCCTGCCTTGGCCAGAGCTTTCCCAGCCAACGCTGCTACCTTCC TAGGGGTTGAATTGGCCAGAAAGTTCTTGGACAAAGTTATATAA	
FAA1	Candida viswanathii	ATGGGTGCCCCTTTAACAGTCGCCGTTGGCGAAGCAAAACCAGGCGAAACCGCTCCAAGAA GAAAAGCCGCTCAAAAAATGGCCTCTGTCGAACGCCCAACAGACTCAAAGGCAACCACTTTG CCAGACTTCATTGAAGAGTGTTTTGCCAGAAACGGCACCAGAGATGCCATGGCCTGGAGAG ACTTGGTCGAAATCCACGTCGAAACCAAACAGGTTACCAAAATCATTGACGGCGAACAGAAA AAGGTCGATAAGGACTGGATCTACTACGAAATGGGTCCTTACAACTACATATCCTACCCCAA GTTGTTGACGTTGGTCAAGAACTACTCCAAGGGTTTGTTGGAGTTGGGCTTGGCCCCAGATC AAGAATCCAAGTTGATGATCTTTGCCAGTACCTCCCACAAGTGGATGCAGACCTTCTTAGCCT CCAGTTTCCAAGGTATCCCCGTTGTCACCGCCTACGACACCTTGGGTGAGTCGGGCTTGAC CCACTCCTTGGTGCAAACCGAATCCGATGCCGTGTTCACCGACAACCAATTGTTGTCCTCCT TGATTCGTCCTTTGGAGAAGGCCACCTCCGTCAAGTATGTCATCCACGGGGAAAAGATTGAC CCTAACGACAAGAGACAGGGCGGCAAAATCTACCAGGATGCGGAAAAGGCCAAGGAGAAGA TTTTACAAATTAGACCAGATATTAAATTTATTTCTTTCGACGAGGTTGTTGCATTGGGTGAACA ATCGTCCAAAGAATTGCATTTCCCAAAACCAGAAGACCCAATCTGTATCATGTACACCTCGGG TTCCACCGGTGCTCCAAAGGGTGTGGTTATCACCAATGCCAACATTGTTGCCGCCGTGGGTG GTATCTCCACCAATGCTACTAGAGACTTGGTTAGAACTGTCGACAGAGTGATTGCATTTTTGC CATTGGCCCACATTTTCGAGTTGGCCTTTGAGTTGGTTACCTTCTGGTGGGGGGCTCCATTG GGTTACGCCAATGTCAAGACTTTGACCGAAGCCTCCTGCAGAAACTGTCAGCCAGACTTGAT TGAATTCAAACCAACCATCATGGTTGGTGTTGCTGCCGTTTGGGAATCGGTCAGAAAGGGTG TCTTGTCTAAATTGAAACAGGCTTCTCCAATCCAACACAAAGATCTTCTGGGCTGCATTCAATG CCAAGTCTACTTTGAACCGTTATGGCTTGCCAGGCGGTGGGTTGTTTGACGCTGTCTTCAAG AAGGTTAAAGCCGCCACTGGTGGCCAATTGCGTTATGTGTTGAATGGTGGGTCCCCAATCTC TGTTGATGCCCAAGTGTTTATCTCCACCTTGCTTGCGCCAATGTTGTTGGGTTACGGTTTGAC TGAAACCTGTGCCAATACCACCATTGTCGAACACACGCGCTTCCAGATTGGTACTTTGGGTA CCTTGGTTGGATCTGTCACTGCCAAGTTGGTTGATGTTGCTGATGCTGGATACTACGCCAAG AACAACCAGGGTGAAATCTGGTTGAAAGGCGGTCCAGTTGTCAAGGAATACTACAAGAACGA AGAAGAAACCAAGGCTGCATTCACCGAAGATGGCTGGTTCAAGACTGGTGATATTGGTGAAT GGACCGCCGACGGTGGTTTGAACATCATTGACCGTAAGAAGAACTTGGTCAAGACTTTGAAT GGTGAATACATTGCTTTGGAGAAATTGGAAAGTATTTACAGATCCAACCACTTGATTTTGAAC TTGTGTGTTTACGCTGACCAAACCAAGGTCAAGCCAATTGCTATTGTCTTGCCAATTGAAGCC AACTTGAAGTCTATGTTGAAGGACGAAAAGATTATCCCAGATGCTGATTCACAAGAATTGAGC AGCTTGGTTCACAACAAGAAGGTTGCCCAAGCTGTCTTGAGACACTTGCTCCAAACCGGTAA ACAACAAGGTTTGAAAGGTATTGAATTGTTGCAGAATGTTGTCTTGTTGGATGACGAGTGGAC	381

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1806 of 3077

Table 25

Gene	Organism	Native Cv sequence	SEQ ID NO
		CCCACAGAATGGTTTTGTTACTTCTGCCCAAAAGTTGCAGAGAAAGAAGATTTTAGAAAGTTG TAAAAAAGAAGTTGAAGAGGCATACAAGTCGTCTTAG	
FAT1	Candida viswanathii	ATGTCAGGATTAGAAATAGCCGCTGCTGCCATCCTTGGTAGTCAGTTATTGGAAGCCAAATAT TTAATTGCCGACGACGTGCTGTTAGCCAAGACAGTCGCTGTCAATGCCCTCCCATACTTGTG GAAAGCCAGCAGAGGTAAGGCATCATACTGGTACTTTTTCGAGCAGTCCGTGTTCAAGAACC CAAACAACAAAGCGTTGGCGTTCCCAAGACCAAGAAAGAATGCCCCCCACCCCAAGACCGA CGCCGAGGGATTCCAGATCTACGACGATCAGTTTGACCTAGAAGAATACACCTACAAGGAAT TGTACGACATGGTTTTGAAGTACTCATACATCTTGAAGAACGAGTACGGCGTCACTGCCAAC GACACCATCGGTGTTTCTTGTATGAACAAGCCGCTTTTCATTGTCTTGTGGTTGGCATTGTGG AACATTGGTGCCTTGCCTGCGTTCTTGAACTTCAACACCAAGGACAAGCCATTGATCCACTGT CTTAAGATTGTCAACGCTTCGCAAGTTTTCGTTGACCCGGACTGTGATTCCCCAATCAGAGAT ACCGAGGCTCAGATCAGAGAGGAATTGCCACATGTGCAAATAAACTACATTGACGAGTTTGC CTTGTTTGACAGATTGAGACTCAAGTCGACTCCAAAACACAGAGCCGAGGACAAGACCAGAA GACCAACCGATACTGACTCCTCCGCTTGTGCATTGATTTACACCTCGGGTACCACCGGTTTG CCAAAAGCCGGTATCATGTCCTGGAGAAAAGCCTTCATGGCCTCGGTTTTCTTTGGCCACAT CATGAAGATTGACTCGAAATCGAACGTCTTGACCGCCATGCCCTTGTACCACTCCACCGCGG CCATGTTGGGGTTGTGTCCTACTTTGATTGTCGGTGGCTGTGTCTCCGTGTCCCAGAAATTC TCCGCTACTTCGTTCTGGACCCAGGCCAGATTATGTGGTGCCACCCACGTGCAATACGTCGG TGAGGTCTGTCGTTACTTGTTGAACTCCAAGCCTCATCCAGACCAAGACAGACACAATGTCA GAATTGCCTACGGTAACGGGTTGCGTCCAGATATATGGTCTGAGTTCAAGCGCAGATTCCAC ATTGAAGGTATCGGTGAGTTCTACGCCGCCACCGAGTCCCCTATCGCCACCACCAACTTGCA GTACGGTGAGTACGGTGTCGGCGCCTGTCGTAAGTACGGGTCCCCTCATCAGCTTGTTATTGT CTACCCAGCAGAAATTGGCCAAGATGGACCCAGAAGACGAGAGTGAAATCTACAAGGACCC CAAGACCGGGTTCTGTACCGAGGCCGCTTACAACGAGCCAGGTGAGTTGTTGATGAGAATC TTGAACCCTAACGACGTGCAGAAATCCTTCCAGGGTTATTATGGTAACAAGTCCGCCACCAA CAGCAAAATCCTCACCAATGTTTTCAAAAAAGGTGACGCGTGGTACAGATCCGGTGACTTGT TGAAGATGGACGAGGACAAATTGTTGTACTTTGTCGACAGATTAGGTGACACTTTCCGTTGG AAGTCCGAAAACGTCTCCGCCACCGAGGTCGAGAACGAATTGATGGGCTCCAAGGCCTTGA AGCAGTCCGTCGTTGTCGGTGTCAAGGTGCCAAACCACGAAGGTAGAGCCTGTTTTGCCGT CTGTGAAGCCAAGGACGAGTTGAGCCATGAAGAAATCTTGAAATTGATTCACTCTCACGTGA CCAAGTCTTTGCCTGTGTATGCTCAACCTGCGTTCATCAAGATTGGCACCATTGAGGCTTCG CACAACCACAAGGTTCCTAAGAACCAATTCAAGAACCAAAAGTTGCCAAAGGGTGAAGACGG CAAGGATTTGATCTACTGGTTGAATGGCGACAAGTACCAGGAGTTGACTGAAGACGATTGGT CTTTGATTTGTACCGGTAAAGCCAAATTGTAG	382

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1807 of 3077

Table 25

Gene	Organism	Native Cv sequence	SEQ ID NO
tHMG1	Candida viswanathii	ATGGACCAAGAAATGTCTCGTCCAAACAAGCCGCTGCTGCTGTTGCTCCATCTGCCGC AAAATCAGTGGCGCCAGAAGAAGACGAAGACGAAACCTCCAGCGAGGAACTTGAAATCAAA GCTCCTGTCAAGCCATTACCACTTGAAGAATGTACCAGAATCCTCAAGAGAGGCAAAGTCAA GACCTTGTCCAATGCTGAAGTGTCTAGTTTTGGTTGTTGGTGGTAAATTGCCATTGTACGCTTT GGAGAAGCAATTGGGCGACCACACAAGAGAGCCGTGGCTGTTCGTCGTAAGGCTATTGCAAAA TTAGCTAATGCGCCAGTATTAGAAACCAACCGTTTACCTTACTCCCACTATGATTATGATAGA GTTTTTGGTGCTTGTTGTGAGAATGTTATTGGATACATGCCAATCCCTGTTGGTGTTGCTGGA CCATTAGTTATTGATGGCAAGCCATACCACATCCCCATGGCCACTACTGAAGGTTGTCTTGTT GCCTCTACTATGCGTGGTTGTAAAGCTATTAACGCTGGAGGTGGTGTTGAAACTGTCTTGAC CAAAGATGGTATGACTAGAGGTCCTTGTGTTCGTTTCCCAACTTTGAAAGAGCCGGCGCTG CAAAGTTGTGGATTGACTCCGAAGAAGGTCAAATCACCATCAAAAAGGCATTCAACTCGACTT CTAGATTTGCTCGTTTGCAACATATTCAAACTGCCCTCGCTGGTACCTCGTTGTTTATTAGATT TAGAACTACTACTGGTGATGCCATGGGTATGAACATGATTTCCAAAGGTGTTGAGTATTCCTT GAAGTACATGGTTGAAGAATGTGGCTGGGACGATATGGAAGTTATCTCTGTTTCTGGTAACTA CTGTACCGATAAGAAACCAGCTGCTATTAACTGGATCGAAGGTAGAGGTAAGTCTATTGTTG CCGCTGCTACCATTCCTGCTGATGTTGTCACTAAAGTCTTGAAATCCGATGTCGATGCATTAG TTGAATTGAATGTTTCTAAGAACTTGGTTGGTTCAGCCATGGCCGGTTCAGTTGGTGGTTTCA ACGCCCATGCAGCCAACTTGGTCACAGCTGTTTATTTGGCCTGTGGTCAGGACCCTGCACAA AACGTCGAATCTTCAAACTGTATTACTTTGATGGAAAAGGACAAACAGACTGGTGACTTGGTG ATTTCTGTTTCTATGCCATCTATCGAAGTTGGTACGATTGGTGGTGGTACCATTTTGGAACCA CAAGCAGCCATGTTGGACTTGTTGGGCGTGCGTGGTCCTCACCCAACCAACCCTGGTGACA ATGCAAGACAATTGGCCAAGATTGTTGCTTCGGCAGTGTTAGCAGCTGAATTGTCACTTTGTT CCGCTTTGGCTGCTGGTCACTTGGTGCAATCCCACATGCAACACAACCGTAAAGGAGCTACC GCCGCCGCCGCCGCCGCAGCAGCCCCCTGCAATCGCCAACGGCCGTGCCAATGGAACCAAA ACCAATGGGTCTATCAATGGGAAAGACTTGAAGCGTCTCAAAGACGGATCAGTTACATGTAT AAAGTCA	383
POX4	Candida viswanathii	ATGACTTTTACAAAGAAAAACGTTAGTGTATCACAAGGTCCTGACCCTAGATCATCCATCCAA AAGGAAAGAGACAGCTCCAAATGGAACCCTCAACAAATGAACTACTTCTTGGAAGGCTCCGT CGAAAGAAGTGAGTTGATGAAGGCTTTGGCCCAACAAATGGAAAGAGACCCAATCTTGTTCA CAGACGGCTCCTACTACGACTTGACCAAGGACCAACAAAGAGAATTGACCGCCGTCAAGATC AACAGAATCGCCAGATACAGAGAACAAGAATCCATCGACACTTTCAACAAGAGATTGTCCTTG ATTGGTATCTTTGACCCACAGGTCGGTACCAGAATTGGTGTCAACCTCGGTTTGTTCCTTTCT TGTATCAGAGGTAACGGTACCACTTCCCAATTGAACTACTGGGCTAACGAAAAGGAAACCGC TGACGTTAAAGGTATCTACGGTTGTTTCGGTATGACCGAATTGGCCCACGGTTCCAACGTTG	384

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1808 of 3077

Table 25

Gene	Organism	Native Cv sequence	SEQ ID NO
		CTGGTTTGGAAACCACCGCCACATTTGACAAGGAATCTGACGAGTTTGTCATCAACACCCCA CACATTGGTGCCACCAAGTGGTGGATTGGTGGTGCTGCTCACTCCGCCACCCACTGTTCTGT CTACGCCAGATTGATTGTTGACGGTCAAGATTACGGTGTCAAGACTTTTGTTGTCCCATTGAG AGACTCCAACCACGACCTCATGCCAGGTGTCACTGTTGGTGACATTGGTGCCAAGATGGGTA GAGATGGTATCGATAACGGTTGGATCCAATTCTCCAACGTCAGAATCCCAAGATTCTTTATGT TGCAAAGTTCTGTAAGGTTTCTGCTGAAGGTGAAGTCACCTTGCCACCTTTGGAACAATTGT CTTACTCCGCCTTGTTGGGTGGTAGAGTCATGATGGTTTTGGACTCCTACAGAATGTTGGCT AGAATGTCCACCATTGCCTTGAGATACGCCATTGGTAGAAGACAATTCAAGGGTGACAATGT CGATCCAAAAGATCCAAACGCTTTGGAAACCCAATTGATAGATTACCCATTGCACCAAAAGAG ATTGTTCCCATACTTGGCTGCTGCCTACGTCATCTCCGCTGGTGCCCTCAAGGTTGAAGACA CCATCCATAACACCTTGGCTGAATTGGACGCTGCCGTTGAAAAGAACGACACCAAGGCTATC TTTAAGTCTATTGACGACATGAAGTCATTGTTTGTTGACTCTGGTTCCTTGAAGTCCACTGCC ACTTGGTTGGGTGCTGAAGCCATTGACCAATGTAGACAAGCCTGTGGTGGTCACGGTTACTC GTCCTACAACGGCTTCGGTAAAGCCTACAACGATTGGGTTGTCCAATGTACTTGGGAAGGTG ACAACAATGTCTTGGCCATGAGTGTTGGTAAGCCAATTGTCAAGCAAGTTATCAGCATTGAAG ATGCCGGCAAGACCGTCAGAGGTTCCACCGCTTTCTTGAACCAATTGAAGGACTACACTGGT TCCAACAGCTCCAAGGTTGTTTTGAACACTGTTGCTGACTTGGACGACATCAAGACTGTCATC AAGGCTATTGAAGTTGCCATCATCAGATTGTCCCAAGAAGCTGCTTCTATTGTCAAGAAGGAA TCTTTCGACTATGTCGGCGCTGAATTGGTTCAACTCTCCAAGTTGAAGGCTCACCACTACTTG TTGACTGAATACATCAGAAGAATTGACACCTTTGACCAAAAGGACTTGGTTCCATACTTGATC ACCCTCGGTAAGTTGTACGCTGCCACTATTGTCTTGGACAGATTTGCCGGTGTCTTCTTGACT TTCAACGTTGCCTCCACCGAAGCCATCACTGCTTTGGCCTCTGTGCAAATTCCAAAGTTGTGT GCTGAAGTCAGACCAAACGTTGTTGCTTACACCGACTCCTTCCAACAATCCGACATGATTGTC AATTCTGCTATTGGTAGATACGATGGTGACATCTATGAGAACTACTTTGACTTGGTCAAGTTG CAGAACCCACCATCCAAGACCAAGGCTCCTTACTCTGATGCTTTGGAAGCCATGTTGAACAG ACCAACCTTGGACGAAAGAGAAAGATTTGAAAAGTCTGATGAAACCGCTGCTATCTTGTCCAA GTAA	
POX5	Candida viswanathii	ATGCCTACCGAACTTCAAAAGAAAGAGAACTCACCAAGTTCAACCCAAAGGAGTTGAACTA CTTCTTGGAAGGTTCCCAAGAAAGATCCGAGATCATCAGCAACATGGTCGAACAAATGCAAA AAGACCCTATCTTGAAGGTCGACGCTTCATACTACAACTTGACCAAAGACCAACAAGAGAA GTCACCGCCAAGAAGATTGCCAGACTCTCCAGATACTTTGAGCACGAGTACCCAGACCAACA GGCCCAGAGATTGTCGATCCTCGGTGTCTTTGACCCACAAGTCTTCACCAGAATCGGTGTCA ACTTGGGTTTGTTTGTTTCCTGTGTCCGTGGTAACGGTACCAACTCCCAGTTCTTCTACTGGA CCATAAATAAGGGTATCGACAAGTTGAGAGGTATCTATGGTTGTTTTGGTATGACTGAGTTGG	385

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1809 of 3077

Table 25

Gene	Organism	Native Cv sequence	SEQ ID NO
		CCCACGGTTCCAACGTCCAAGGTATTGAAACCACCGCCACTTTTGACGAAGACACTGACGAG TTTGTCATCAACACCCCACACATTGGTGCCACCAAGTGGTGGATCGGTGGTGCTGCGCACTC CGCCACCCACTGCTCCGTCTACGCCAGATTGAAGGTCAAGGAAAGGACTACGGTGTCAAG ACCTTTGTTGTCCCATTGAGAGACTCCAACCACGACCTCGAGCCAGGTGTGACTGTTGGTGA CATTGGTGCCAAGATGGGTAGAGACGGTATCGATAACGGTTGGATCCAGTTCTCCAACGTCA GAATCCCAAGATTCTTTATGTTGCAAAAGTACTGTAAGGTTTCCCGTCTGGGTGAAGTCACCA TGCCACCATCTGAACAATTGTCTTACTCGGCTTTGATTGGTGGTAGAGTCACCATGATGATGG ACTCCTACAGAATGACCAGTAGATTCATCACCATTGCCTTGAGATACGCCATCCACAGAAGA CAATTCAAGAAGAAGGACACCGATACCATTGAAACCAAGTTGATTGACTACCCATTGCATCAA AAGAGATTGTTCCCATTCTTGGCTGCCGCTTACTTGTTCTCCCAAGGTGCCTTGTACTTAGAA CAAACCATGAACGCAACCAACGACAAGTTGGACGAAGCTGTCAGTGCTGGTGAAAAGGAAG CCATTGACGCTGCCATTGTCGAATCCAAGAAATTGTTCGTCGCTTCCGGTTGTTTGAAGTCCA CCTGTACCTGGTTGACTGCTGAAGCCATTGACGAAGCTCGTCAAGCTTGTGGTGGTCACGGT TACTCGTCTTACAACGGTTTCGGTAAAGCCTACTCCGACTGGGTTGTCCAATGTACCTGGGA AGGTGACAACAACATCTTGGCCATGAACGTTGCCAAGCCAATGGTTAGAGACTTGTTGAAGG AGCCAGAACAAAAGGGATTGGTTCTCTCCAGCGTTGCCGACTTGGACGACCCAGCCAAGTT GGTTAAGGCTTTCGACCACGCCCTTTCCGGCTTGGCCAGAGACATTGGTGCTGTTGCTGAAG ACAAGGGTTTCGACATTACCGGTCCAAGTTTGGTTTTGGTTTCCAAGTTGAACGCTCACAGAT TCTTGATTGACGGTTTCTTCAAGCGTATCACCCCAGAATGGTCTGAAGTCTTGAGACCTTTGG GTTTCTTGTATGCCGACTGGATCTTGACCAACTTTGGTGCCACCTTCTTGCAGTACGGTATCA TTACCCCAGATGTCAGCAGAAAGATTTCCTCCGAGCACTTCCCAGCCTTGTGTGCCAAGGTT AGACCAAACGTTGTTGGTTTGACTGATGGTTTCAACTTGACTGACATGATGACCAATGCTGCT ATTGGTAGATATGATGGTAACGTCTACGAACACTACTTCGAAACTGTCAAGGCTTTGAACCCA CCAGAAAACACCAAGGCTCCATACTCCAAGGCTTTGGAAGACATGTTGAACCGTCCAGACCT TGAAGTCAGAGAAAGAGGTGAAAAGTCCGAAGAAGCTGCTGAAATCTTGTCCAGTTAA	386
URA3	Candida viswanathii	ATGGTTAGCACAAAAACATACACAGAAAGGGCATCAGCACACCCCTCCAAGGTTGCCCAACG TTTATTCCGCTTAATGGAGTCCAAAAAGACCAACCTCTGCGCCTCGATCGACGTGACCACAA CCGCCGAGTTCCTTTCGCTCATCGACAAGCTCGGTCCCCACATCTGTCTCGTGAAGACGCAC ATCGATATCATCTCAGACTTCAGCTACGAGGGCACGATTGAGCCGTTGCTTGTGCTTGCAGA GCGCCACGGGTTCTTGATATTCGAGGACAGGAAGTTTGCTGATATCGGAAACACCGTGATGT TGCAGTACACCTCGGGGGTATACCGGATCGCCGGCGTGGAGTGACATCACGAACGCGCACG GAGTGACTGGGAAGGGCGTCGTTGAAGGGTTGAAACGCGGTGCGGAGGGGGTAGAAAAGG AAAGGGGCGTGTTGATGTTGGCGGAGTTGTCGAGTAAAGGCTCGTTGGCGCATGGTGAATA TACCCGTGAGACGATCGAGATTGCGAAGAGTGATCGGGAGTTCGTGATTGGGTTCATCGCG	

Table 25

Gene	Organism	Native Cv sequence	SEQ ID NO
		CAGCGGGACATGGGGGGGTAGAGAAGAAGGGTTTGATTGGATCATCATGACGCCTGGTGTGG GGTTGGATGATAAAGGCGATGCGTTGGGCCAGCAGTATAGGACTGTTGATGAGGTGGTTCT GACTGGTACCGATGTGATTATTGTCGGGAGAGGGTTGTTTGGAAAAGGAAGAGACCCTGAG GTGGAGGGAAAGAGATACAGGGATGCTGGATGGAAGGCATACTTGAAGAGAACTGGTCAGT TAGAATAA	
LEU2	Candida viswanathii	ATGTCCGTTAAAACCAAAACCATCACTATTCTCCCAGGTGACCACGTCGGTACCGAGATCGT TGCCGAGGCAATCAAGTCCTTACATGCCATTGAGTCCCTGACCCCATACCAAAAAGTTCATTT TGAGTTCAAGCACCACTTGATAGGCGGTGCCGCCATTGACGCCACTGGTGTGCCGCTCCCA GACGACGCCTTAGCCGCCGCCAAGTCCTCAGACGCCGTGTTGCTCGGGGCCGTCGGTGGA CCCAAATGGGGAACCGGCGCTGTGCGTCCAGAACAGGGATTGTTGAAGATCCGTAAGGAGT TGAACCTTTATGCCAACATCAGACCATGCAATTTCGCTAGCGACTCCTTGCTTGAGCTCTCTC CGTTACGTCCTGAGGTCGTCAAGGGCACCAACTTGATCATTGTGCCGTGAGCTAGTTGGTGG GATTTACTTTGGTGAGAGACAGGAGCAGGAAGAAAGCGAGGATGGGAAATCGGCTTGGGAT ACTGAAAAGTACACCGTTGATGAAGTCGCGAGAATTACACGCATGGCTGCGTTCATGGCATT GCAACACACCCCACCATTGCCAATCTGGTCCTTGGACAAAGCCAACGTGCTTGCCTCGTCAA GATTGTGGAGAAAGACAGTGGATAAAATCATCAGTGAAGAGTTCCCTACCTTGGCTGTCCAG CACCAGTTGATTGACTCCGCGGCTATGATCTTGATCCAGAACCCAACGAAGTTGAACGGGAT CATCATCACGTCGAACATGTTTGGTGATATCATATCCGATGAAGCGTCCGTGATCCCAGGGT CCTTGGGATTGTTGCCATCTGCGTCATTGGCTTCGTTACCGGACACAAACACTGCGTTTGGT CTCTATGAGCCATGCCATGGTTCGGCACCTGATTTGCCAGAGAACAAGTCAACCCTATTGC GACAATCTTGTCCGTTGCTAGTATGTTGAGATTGAGTTTGGATTGTGTTAAGGAAGCCGAGG CTTTGGAACAGGCCGTGAAGGAAGTCTTGGATAAGGGGATCAGAACGGCGGATTTGAGAGG TAGTAGCACACCACTGAAGTCGGTGATGCCGTTGCTGAAACCGTCTCCAGAATCTTGAAAG AAGCTAAGGCTTGA	387

EXAMPLE 71: Plasmids

Certain plasmids described herein are further characterized in Table 26 below.

Table 26

Plasmid No.	Description
pAA1116	POX4 promoter + POX5F98G + POX4 terminator in AFC
pAA1164	HDE1 promoter + POX5F98G + POX4 terminator in AFC
pAA1701	CRC1 knockout cassette 1
pAA2214	G6PI promoter + CRC1 + terminator POX4 @ int5 Integration site
pAA2311	G6PI promoter + CRC1 + terminator POX4 in AFC
pAA2534	GPD promoter + Caenorhabditis elegans FAT5 CvCO + POX4 terminator in AFC
pAA2698	GPD promoter + CsCrtE CvCO + POX4 terminator in AFC
pAA2699	GPD promoter + CsCrtB CvCO + POX4 terminator in AFC
pAA2700	GPD promoter + CsCrtI CvCO + POX4 terminator in AFC
pAA2701	GPD promoter + CnTPS1 CvCO + POX4 terminator in AFC
pAA2702	HDE1 promoter + CsCrtE CvCO + POX4 terminator in AFC
pAA2703	HDE1 promoter + CsCrtB CvCO + POX4 terminator in AFC
pAA2704	HDE1 promoter + CsCrtI CvCO + POX4 terminator in AFC
pAA2705	HDE1 promoter + CnTPS1 CvCO + POX4 terminator in AFC
pAA2866	GPD promoter + XdCrtYB CvCO + POX4 terminator in AFC
pAA2993	GPD promoter + XdCrtR CvCO + POX4 terminator in AFC
pAA2994	GPD promoter + XdCrtS CvCO +

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1812 of 3077

	POX4 terminator in AFC
pAA2995	GPD promoter + AaCrtW CvCO + POX4 terminator in AFC
pAA2996	GPD promoter + AaCrtZ CvCO + POX4 terminator in AFC
pAA2998	Plasmid from Invitrogen containing XdCrtR CvCO
pAA2999	Plasmid from Invitrogen containing XdCrtS CvCO
pAA3000	Plasmid from Invitrogen containing AaCrtW CvCO
pAA3001	Plasmid from Invitrogen containing AaCrtZ CvCO
pAA3373	HDE1 promoter + XdCrtYB CvCO + POX4 terminator in AFC
pAA3189	GPD promoter + XdCrtE + POX4 terminator in AFC
pAA3190	GPD promoter + PaCrtE + POX4 terminator in AFC
pAA3378	GPD promoter + PaCrtB + POX4 terminator in AFC
pAA3490	GPD promoter + XdCrtI + POX4 terminator in AFC
pAA3633	HDE promoter + XdCrtE CvCO + POX4 terminator in AFC
pAA3634	HDE promoter + XdCrtI CvCO + POX4 terminator in AFC
pAA408	Contains URA3 terminator + URA3 promoter + URA3 + URA3 terminator
pAA3060	LEU2 knock-out cassette 1
pAA2417	LEU2 knock-out cassette 2
pVZ3930	RAS2 knock-out cassette
pVZ4045	LEU2 single crossover AFC
pVZ4056	HDE promoter + XdCrtE CvCO + POX4 terminator in AFC, leucine
pVZ4057	HDE promoter + CsCrtB CvCO + POX4 terminator in AFC, leucine
pVZ4058	HDE promoter + XdCrtI CvCO + POX4 terminator in AFC, leucine
pVZ4083	Ppox18 -Tpox18 in AFC, leucine

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1813 of 3077

pVZ4098	POX18 promoter + Erg10 + POX18 terminator in AFC, leucine
pVZ4099	POX18 promoter + Erg13 + POX18 terminator in AFC, leucine
pVZ4100	POX18 promoter + HMG1 + POX18 terminator in AFC, leucine
pVZ4101	POX18 promoter + Erg12 + POX18 terminator in AFC, leucine
pVZ4122	POX18 promoter + Erg8 + POX18 terminator in AFC, leucine
pVZ4123	POX18 promoter + Mvd1 + POX18 terminator in AFC, leucine
pVZ4104	POX18 promoter + Idi1 + POX18 terminator in AFC, leucine
pVZ4105	POX18 promoter + Erg20 + POX18 terminator in AFC, leucine

AFC = antibiotic free cassette, uracil marker unless indicated otherwise
CvCO = Codon optimized for Candida viswanathii

EXAMPLE 72: Oligonucleotides

Certain oligonucleotides described herein are provided in Table 27 below.

Table 27

Oligo No.	Sequence	SEQ ID NO
oAA02206	TTCCGCTTAATGGAGTCCAAA	388
oAA2091	GCATACAACGGAAATTTGCTTT	389
oAA02209	TAAACGTTGGGCAACCTTGG	390
oAA5511	ATGGACGACGTTGATTCTGCTTTAGCCGATAATGTT	391
oAA5512	TTATATAACTTTGTCCAAGAACTTTCTGGCC	392
oAA05788	GAATAGAAGAGAGTGACTCTTTTG	393
oAA05789	GATTGATTGTTATAGTTTCTTTCTTTC	394

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1814 of 3077

Table 27		
Oligo No.	**Sequence**	**SEQ ID NO**
oAA07265	CTGCAGGCATGATGATCTG	395
oAA07511	ATCGATTAAATTCTTTAATTGAGGG	396
oAA07512	GAATAGAAGAGAGTGACTCTTTTG	397
oAA07624	AGATCATCATGCCTGCAGAAAATCAGAGGCTACTCC	398
oAA07625	GTCACTCTCTTCTATTCTTATATAACTTTGTCCAAGAACTTTC	399
oAA09745	CCCTCAATTAAAGAATTTAATCGATATGAATGCTAATGCCGTTAAGTC	400
oAA09746	TCAAAAGAGTCACTCTCTTCTATTCTTACCCAAACATTGCAAGCTGT	401
oAA09747	CCCTCAATTAAAGAATTTAATCGATATGACCAAAACGGTTGTAATAGGATCAGG	402
oAA09748	TCAAAAGAGTCACTCTCTTCTATTCTTACGCGTGGCCCTCCAACATT	403
oAA09749	CCCTCAATTAAAGAATTTAATCGATATGAGTGACAAGCCTTTGTTAACACATGCTACAGA	404
oAA09750	TCAAAAGAGTCACTCTCTTCTATTCTTAAGCGGCAGGAGGGGCGGCA	405
oAA09751	CCCTCAATTAAAGAATTTAATCGATATGGCAGAAATGTTCAATGGAAAC	406
oAA09752	TCAAAAGAGTCACTCTCTTCTATTCTTAAGGTATAATCGGTTCGACG	407
oAA09753	AAGAAAGAAACTATAACAATCAATCATGAATGCTAATGCCGTTAAGTC	408
oAA09754	TCAAAAGAGTCACTCTCTTCTATTCTTACCCAAACATTGCAAGCTGT	409
oAA09755	AAGAAAGAAACTATAACAATCAATCATGACCAAAACGGTTGTAATAGGATCAGG	410
oAA09756	TCAAAAGAGTCACTCTCTTCTATTCTTACGCGTGGCCCTCCAACATT	411
oAA09757	AAGAAAGAAACTATAACAATCAATCATGAGTGACAAGCCTTTGTTAACACATGCTACAGA	412
oAA09758	TCAAAAGAGTCACTCTCTTCTATTCTTAAGCGGCAGGAGGGGCGGCA	413
oAA09759	AAGAAAGAAACTATAACAATCAATCATGGCAGAAATGTTCAATGGAAAC	414
oAA09760	TCAAAAGAGTCACTCTCTTCTATTCTTAAGGTATAATCGGTTCGACG	415
oAA9960	CCCTCAATTAAAGAATTTAATCGATATGGCAACCCTATCTGACTTGG	416
oAA9961	TCAAAAGAGTCACTCTCTTCTATTCTTAGCTCCACACATCCATCAATAATC	417
oAA9962	CCCTCAATTAAAGAATTTAATCGATATGTTCATTTTGGTGTTGTTGACC	418

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1815 of 3077

Table 27		
Oligo No.	**Sequence**	**SEQ ID NO**
oAA9963	TCAAAAGAGTCACTCTCTTCTATTCTTACTCCACTGGCTTTACCTGC	419
oAA9964	CCCTCAATTAAAGAATTTAATCGATATGAGTGCCCATGCCTTGCCAA	420
oAA9965	TCAAAAGAGTCACTCTCTTCTATTCTTAAGCGGTATCACCTTTAGTTCTAGTGC	421
oAA9966	CCCTCAATTAAAGAATTTAATCGATATGACAAACTTCTTGATCGTTGTGG	422
oAA9967	TCAAAAGAGTCACTCTCTTCTATTCTTACGTCCTCTCTTGGGCCTCA	423
oAA9968	GATCCACAAGTTTTCACTAGAATTGG	424
oAA10284	CACTGCTTAATTAAGCGCTGTCACTCTTATATCAGTATGACC	425
oAA10286	CACTGCGGATCCGGTACCCCACCACGACAACAGGATTCCCC	426
oAA10287	CGAGATTTAATTAATTAGCCAACAGGACAATAGAGTTAAAAAGAATGAATTG	427
oVZ0041	CGAGATCTGCAGGCGGCCGCTTCTGTGAGTATTTGTTATTGAAAATGAAGGGATGG	428
oAA7682	CCACGTCGGTACCGAGATCGTTGCCGAGGCAATCAAGTCCTT	429
oAA7683	AACGGCTTCGTCTAAACAACCACGGATCTTCAACAATCCCTGTTCTGGAC	430
oAA7684	GTCCAGAACAGGGATTGTTGAAGATCCGTGGTTGTTTAGACGAAGCCGTT	431
oAA7685	AGTGTTTGTGTCCGGTAACGACCGAAATATTACAATTGGAGCTCC	432
oAA7686	GGAGCTCCAATTGTAATATTTCGGTCGTTACCGGACACAAACACT	433
oAA7687	TTTCAGCAACGGCATCACC	434
oAA7941	ACCTTTATGCCAACATCAGACC	435
oAA7942	AACGGCTTCGTCTAAACAACCCATCAACGGTGTACTTTTCAGTATCC	436
oAA7943	GGATACTGAAAAGTACACCGTTGATGGGTTGTTTAGACGAAGCCGTT	437
oAA7944	TTGCAATGCCATGAACGCCCGAAATATTACAATTGGAGCTCC	438
oAA7945	GGAGCTCCAATTGTAATATTTCGGGCGTTCATGGCATTGCAA	439
oAA7946	CAGATGGCAACAATCCCAAG	440
oVZ337	AATTAACCTATGGTGCAC	441
oVZ338	TTAATTAAAAGCTTGGCGTAATC	442
oVZ339	CTCGTGCTAGTCAGTCTTGCACGCTTTGGGTG	443
oVZ340	ATGATTACGCCAAGCTTTTAATTAACAACACGGCGTCTGAGGAC	444
oVZ341	ACTGAGAGTGCACCATAGGTTAATTAACTCGGGGCCGTCGGTGGA	445

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1816 of 3077

Table 27		
Oligo No.	**Sequence**	**SEQ ID NO**
oVZ342	AAGCGTGCAAGACTGACTAGCACGAGCGAAGATGGGG	446
oVZ373	CTC GGG GCC GTC GGT GGA	447
oVZ374	CAACACGGCGTCTGAGGACTTGG	448
oVZ369	GCTCAACAATTGTCTGACAAGATCTC	449
oVZ370	GAGATTGTTAACTTTATGGGGCAATAAC	450
oVZ371	TCCCCATCTTCGCTCGTGCTAGTCAAAGGGAAGAAGAGTCGTTG	451
oVZ372	CGTCGGCACCCAAAGCGTGCAAGACGTCGACCTAAATTCGCAAC	452
oVZ376	GTGCTAGTCAAGATAAACCGAGGCATGGAAG	453
oVZ377	CGTGCAAGACGCATAGACTGCAAGTAAGCAGC	454
oVZ378	CAGTCTATGCGTCTTGCACGCTTTGGGTGC	455
oVZ379	CGGTTTATCTTGACTAGCACGAGCGAAGATGG	456
oVZ384	AAAGGAATAAGCTGGGTGTTTGCTCATTTTG	457
oVZ385	CGTTAGTCATTTTTGTTAATTGGTGTTTTGTTTAATAATGGGG	458
oVZ386	AATTAACAAAAATGACTAACGCACCACAAAACATTG	459
oVZ387	AACACCCAGCTTATTCCTTTATAGCATAACTTCTTCTAAACTTGTCG	460
oVZ388	AAAGTCATAGGCTGGGTGTTTGCTCATTTTG	461
oVZ389	AACTTAGCATTTTTGTTAATTGGTGTTTTGTTTAATAATGGGG	462
oVZ390	AATTAACAAAAATGCTAAGTTTCATCACCGAAGC	463
oVZ391	AACACCCAGCCTATGACTTTATACATGTAACGGATCCGTC	464
oVZ392	GTTCTGGTGAGCTGGGTGTTTGCTCATTTTG	465
oVZ393	AGACTGACATTTTTGTTAATTGGTGTTTTGTTTAATAATGGGG	466
oVZ394	AATTAACAAAAATGTCAGTCTCGCCTTTCGTTG	467
oVZ395	AACACCCAGCTCACCAGAACTTCCATTCTTTAACGTTTATC	468
oVZ396	CGGTTTATAGGCTGGGTGTTTGCTCATTTTG	469
oVZ399	AACACCCAGCCTATAAACCGATATAGTCCTCCCAGTTTTC	470
oVZ401	CTGAATACATTTTTGTTAATTGGTGTTTTGTTTAATAATGGGG	471
oVZ402	AATTAACAAAAATGTATTCAGCTTCAGTCACAGC	472
oVZ404	GATTACTTTAAGCTGGGTGTTTGCTCATTTTG	473
oVZ405	CTGATGACATTTTTGTTAATTGGTGTTTTGTTTAATAATGGGG	474
oVZ406	AATTAACAAAAATGTCATCAGAATACGCTAAACTAGTGG	475
oVZ407	AACACCCAGCTTAAAGTAATCTATGTATCTCAGTATCTTGGTACTTTTC	476

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1817 of 3077

Table 27		
Oligo No.	**Sequence**	**SEQ ID NO**
oVZ408	ATCCAAATAGGCTGGGTGTTTGCTCATTTTG	477
oVZ409	TATCAGACATTTTTGTTAATTGGTGTTTTGTTTAATAATGGGG	478
oVZ410	AATTAACAAAAATGTCTGATAAAGCAGCCGCTAG	479
oVZ411	AACACCCAGCCTATTTGGATCTCTTGTAAACCTTGCC	480
oVZ478	TACCAAACTAAGCTGGGTGTTTGCTCATTTTG	481
oVZ479	AACACCCAGCTTAGTTTGGTAACCCACTTTCGTTG	482
oVZ480	CTTTTGCACATTTTTGTTAATTGGTGTTTTGTTTAATAATGGGG	483
oVZ481	CCAATTAACAAAAATGTCAAAAGCTTTTAGTGCCCC	484

EXAMPLE 73: Promoters and terminators

Certain promoters and terminators described herein are provided in Tables 28 and 29 below.

Table 28			
Promoter	**Sequence**	**SEQ ID NO**	**Length**
HDE1	AAGGGAAGAAGAGTCGTTGAGTTGATGTAATT AAGCTGGCACGTAGATACCAGAAGGTTCTAG AGTAGAGCTTGGGTGGTGTTTGGCCCTGTTT GGACCACGGATAGAGATGGAGAATCCCTTGG TTAGAGCGGAGAGGAAAAAATTGAAACTTTGC ATATCCCACTTCATTATCCTTGATGTAACCGTT TTATGGGGTAATTAAAGTGTGGAAAAATAATC AGGGAGACATATTCCCGATCAATTGGGTGGT GGTCGCTCAATTTCTGTGAGTAGTAGGCTCAG TGGTGTGTATTGGGATTGGTAGTAGTCTGTAT AAGCAGTGTTATATAACCCATTGCTTGTTGATT CCTATTTTGCTGGCAAAAGTGACAACTGTAGT TGTGAGATAATCCTCGGTTATTACGCCTGGGG GGGCAGACAGCCAAAGTTGTGCCCGTGCGAC AATGGCATCAGAAGAAACAGAAAAAAAAAACA CAGGCATTTTTATCCACATGCACACTACCCCC ACTATTCCTGTCTGCAGTGTGCTTGTGTGTGG CCCCCCGCAGAATCAACAGGGCAAACTCTGG AGCCTGAATCTTTATATAAACTTCAGGCATTG GCCCCCCTTTTCACAATTCTTCACATCCACCA TTTTTTTTCTTCTTTCCTACCATATTAGTTTTTT TTTATTCTTTTCCTACCTATCTGATTATTATCAA ACATCTGGTCATCCTCAAAAGAAAGAAAGAAA CTATAACAATCAATC	485	751
POX4	GAGCTCCAATTGTAATATTTCGGGAGAAATAT	486	531

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1818 of 3077

Table 28			
Promoter	**Sequence**	**SEQ ID NO**	**Length**
	CGTTGGGGTAAAACAACAGAGAGAGAGAGGG AGAGATGGTTCTGGTAGAATTATAATCTGGTT GTTGCAAATGCTACTGATCGACTCTGGCAATG TCTGTAGCTCGCTAGTTGTATGCAACTTAGGT GTTATGCATACACACGGTTATTCGGTTGAATT GTGGAGTAAAAATTGTCTGAGTTGTGTCTTAG CTACTGGCTGGCCCCCCGCGAAAGATAATCA AAATTACACTTGTGAATTTTTGCACACACACC GATTAACATTTCCCTTTTTTTGTCCACCGATACA CGCTTGCCTCTTCTTTTTTTTCTCTGTGCTTCC CCCTCCTGTGACTTTTTCCACCATTGATATAAA ATCAACTCCATTTCCCTAAAATCTCCCCAGATT CTAAAAACAACTTCTTCTCTTCTGCTTTTCCTT TTTTTTTGTTATATTTATTTACCATCCCTTTTTT TTGAATAGTTATTCCCCACTAACATTGTTCAAA TCTTCACGACATA		
URA3	CGACGGGTACAACGAGAATTGTATTGAATTGA TCAAGAACATGATCTTGGTGTTACAGAACATC AAGTTCTTGGACCAGACTGAGAATGCACAGAT ATACAAGGCGTCATGTGATAAAATGGATGAGA TTTATCCACAATTGAAGAAAGAGTTTATGGAA AGTGGTCAACCAGAAGCTAAACAGGAAGAAG CAAACGAAGAGGTGAAACAAGAAGAAGAAGG TAAATAAGTATTTTGTATTATATAACAAACAAA GTAAGGAATACAGATTTATACAATAAATTGCC ATACTAGTCACGTGAGATATCTCATCCATTCC CCAACTCCCAAGAAAAAAAAAAAGTGAAAAAA AAAATCAAACCCAAAGATCAACCTCCCCATCA TCATCGTCATCAAACCCCCAGCTCAATTCGCA	487	415
G6PI	AAAATCAGAGGCTACTCCGGAGATGGCACAT TATCACGTGGGCAGTCCCATCTCTCTGACAAT GGCACTGACAATGCCATGTCATTCTAGAAGTC GCTCGTTCCATTGCTAGTCGTGTAGGTGTCTT ACTCAGACCATATGTTTGGCTGTGTGTGTGGT AAGGGGGTAAGATATCTCTAGGAGGAGGCAA CTCTACTCGAAAGTACGAAATGGCCGACGCC AATCCGACCTGCCATTGTGCAGGACGCGGGT CGAGATCCGGAAGTGCCCTCCTTCTGAGGTT GTTGGTTTGTTCCAGTTGGCATAAGAAATTAA TAATCGTTTAGTCAATGGTACAGTATTAATAAT TCCTGACAAAACCGTGCCCTACAAGCAGTGT GAGTTCCGGGGGGTAGCACTAGTGGTCGTGGT GGTAGGGCCCACAGAGAGTATTTCCGGAACC GACTGATCCTTTGATTCCAATATTCCAAGATTT GGGTTTGTGGTTGTGTGTAATGCATGAGGTTA TTTTGTCAACCTCCTTCCTTCTCTACGGTTGTT GCTCCTCCTCCTCTTCCTTCCCTTTGTTCATTT AAATACACACACACCACTCCCCTTCCCACCCA CAAGAACTTTTTTTTTTTTTTACTTTCTTCTTCTC TTACTTTCTTCTTCTTCTTTCATTACCTGTTAG ATTATATTTATCAACCTCAATTGGTTTATTATCT	488	745

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1819 of 3077

Table 28			
Promoter	**Sequence**	**SEQ ID NO**	**Length**
	ACCACCCATACAACCCCCACCCCTGCCACAT CCACCAACA		
GPD	CGGAAGTTGTTTACCGACCTGACCGTAAATTT GCTGCTGAAAGAAACGTGTCAAACAAGACCA ATTGGCTCAATTGACCCTGTGGAAATGCTTTG TTGACCACCAATGCTTCCACCAAACGTTACTT TTTTTTTGCAATCGGATGGTATGGGTCTGGGG TTCACCTGTTTTGTAAAGCTACAGAAGGTGGC ATATTTCTCTGATCAGGTGTTTTTTTTTTCGGC TGCTGCTGCTCGTGGTGGTGTAGTGGTAGTG GTGTGTGTGTGTGTGTGTGCGTGCGTGTGGA AGGACGCTTTTTGCTCTCTGACTCCTCCCAAT CAGAAGTTGCTATAGTGGTGAAACAACAATGG ATGATAATGCCCCGGGCGGTGCGTGTCCGAC ACAAACCACTACATTTTTTAGCTGGGAGCCTA CTGCCACTACGACCCACCCACCCATGGTCAA CAAAAAAATTCTGACAAATTATAAAATAACCCT TGAATTCCCCCTTGGAAAAATTTTTGGTATTTC TCTCTCTCTTTTCCTTTCCCTCTTCTTTTTCTCT CCATCAATCAATTGACGTTCAGTAACTCAATTA ATTACATCACATCCCTCAATTAAAGAATTTAAT CGAT	489	614
POX18	AGATAAACCGAGGCATGGAAGAAGTCAGAGA CTTCCTAATTGGGAAATATCACATGAAAATGT ATGGATGTGTGCAGGACACGTCTACTGAGTG CCATACTTTGTGCCCACCGCGAAATTAAGGG CAGCCATCAGATTACGTAAGCAGTGGTATGCA GTTGTTACAGCGTCAAGTTTTTCTGTATTACC CCAAAGTTTTCTTCTCTCCCCTGCAAAAGTTTT CATATGCAAAACTGATCCCACAGCAGTATAAA TACCCAATAACTTTCCCCAACCAACAGCAACA ACCTCTCATTCTTTTTTTTCCTATTCTTTTCTTT AATCAACCCCCATTATTAAACAAAACACCAATT AACAAAA	490	360
LEU2	GTCTTGCACGCTTTGGGTGCCGACGACGCCT TGGCCCACTCAAGTATTAGATTCGGTATTGGT AGATTCACCACTGAGGCAGAGGTCGACTATG TCATCAAGGCTATTAACGAAAGAGTTGATTTC TTGAGAAAGATGTCTCCATTGTGGGAGATGGT GCAAGAAGGCATTGACTTGGACTCAATTGAAT GGAGTGGTCATTAGTGTGACCCCCTATCGCT ACGGTCTTCCTCTTCGTTTCATTTAAGTATGAC TTGTTTTGTTTATTTATCTCGTGTATAGAATAG TAATTATTTTTGGTATAAGCCATGTTGGTATTC GTGAGTCAGATGTGAGCTGTATGTTCTCCTGA ACCGGCGTAACTTCCCTTCCGTTGCCGCGGG AATGCGACAACAACGACGGGAAAAAAATCCC AGTATAAAACCACCCCAAATGACGTGTGTGAA ATTTTTCACTCCAGGTTTCTCTCCCTCTTTTTT GTCCCCCCAAACCAATCACCA	491	500

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1820 of 3077

Table 29			
Terminator	**Sequence**	**SEQ ID NO**	**Length**
POX4	GAATAGAAGAGAGTGACTCTTTTGATAAGAGT CGCAAATTTGATTTCATAAGTATATATTCATTA TGTAAAGTAGTAAATGGAAAATTCATTAAAAAA AAAGCAAATTTCCGTTGTATGCATACTCCGAA CACAAAACTAGCCCCGGAAAAACCCTTAGTTG ATAGTTGCGAATTTAGGTCGAC	492	184
URA3	TAAATATTGTAATAAATAGGTCTATATACATAC ACTAAGCTTCTAGGACGTCATTGTAGTCTTCG AAGTTGTCTGCTAGTTTAGTTCTCATGATTTCG AAAACCAATAACGCAATGGATGTAGCAGGGAT GGTGGTTAGTGCGTTCCTGACAAACCCAGAG TACGCCGCCTCAAACCACGTCACATTCGCCC TTTGCTTCATCCGCATCACTTGCTTGAAGGTA TCCACGTACGAGTTGTAATACACCTTGAAGAA CGGCTTCGTCT	493	267
POX18	GCTGGGTGTTTGCTCATTTTGTTTCAATGCAT GGATGCACACTGGTCTTTTTATGATTTGAAAT GAATAGATTTGATCATCATCATTTTTTTTTTGC AAATATTTATCATATAATGTGAGTTTTCTTTGG TGTATATTTGTTTTCTACAGTATTTAGATAATAT TATTTGAAAAGAATATATGCTTAAAAGAAACTA CAAGTTAGAGTTCGACGATGAAGTGGTTGTCA TGTCGGCGACTTTACCCCAGAGTAAGAAGGA ATTGTCGTGGTCGATATTGGTTGAAACATATC GGGACTAGTCTTGCGACCGAGCGCAGGGCAA TTGTGGAGTCACACATACAATGAAAATTGGCT GCTTACTTGCAGTCTATGC	494	374
LEU2	GGATCCTCTACATATGAACATTCTATATTGAAA AATGCATGTATATTATTGTCGTTAACGTTTTTT AAAAATTCATGGTCTTTTCTAAAAATGCAGATA GTTGAGCAACAGAATCCTCTTTATCAATCACC ACAACCACCGACTGTTTGCTACTCTCATTCGT AACAATGAATCTATTCGACGTGACCGTAGCCT CACTAGACTCAGATTGGCTGGTGTTCCTCGCA AGATTGGGTTCTTCATTAGGCGAGTAAAACTC CTCTTCCTCGTCCTCCTCCCCATCTTCGCTCG TGCTAGTCA	495	300

5

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1821 of 3077

EXAMPLE 74: Strains

Certain strains described herein are further characterized in Table 30 below.

Table 30

Strain No.	Description	Construction
sAA001	Candida viswanathii ATCC 20336	Acquired from ATCC
sAA002	Candida viswanathii ATCC 20913, Δura3	Acquired from ATCC
sAA4377	Δpox4, Δfaa1, Δfat1, Δcrc1, Δura3	Described in Example 21 and Table 11
sAA5733	ald6/ald6::URA3, crc1:: ura3 promoter/CRC1	Described in International Patent Application Publication No. WO2016/154046
sAA5761	ald6/ald6, crc1:: ura3 promoter /CRC1	5'FOA of sAA5733
sAA6234	crc1::ura3 promoter/crc1::ura3 promoter, ald6::ura3 promoter/ald6::ura3 promoter	5'FOA of sAA5761 + pAA1701
sAA7443	GPD promoter - CsCrtE CvCO, GPD promoter - CsCrtB CvCO, GPD promoter - CsCrtI CvCO	sAA002 + PCR cassettes amplified from pAA2698, pAA2699, and pAA2700 using primers oAA02206 and oAA02209
sAA7444	GPD promoter - CsCrtE CvCO, GPD promoter - CsCrtB CvCO, GPD promoter - CsCrtI CvCO	sAA002 + PCR cassettes amplified from pAA2698, pAA2699, and pAA2700 using primers oAA02206 and oAA02209
sAA7445	GPD promoter - CsCrtE CvCO, GPD promoter - CsCrtB CvCO, GPD promoter - CsCrtI CvCO	sAA002 + PCR cassettes amplified from pAA2698, pAA2699, and pAA2700 using primers oAA02206 and oAA02209
sAA7446	HDE1 promoter - CsCrtE CvCO, HDE1 promoter - CsCrtB CvCO, HDE1 promoter - CsCrtI CvCO	sAA002 + PCR cassettes amplified from pAA2702, pAA2703 , and pAA2704 using primers oAA02206 and oAA02209
sAA7449	GPD promoter - CnTPS1 CvCO	sAA002 + PCR cassettes amplified from pAA2701 using primers oAA02206 and oAA02209

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1822 of 3077

Table 30		
Strain No.	**Description**	**Construction**
sAA7453	HDE1 promoter - CnTPS1 CvCO	sAA0002 + PCR cassettes amplified from pAA2705 using primers oAA02206 and oAA02209
sAA7565	crc1::ura3 promoter/crc1::ura3 promoter, ald6::ura3 promoter/ald6::ura3 promoter, HDE1 promoter - CsCrtE CvCO, HDE1 promoter - CsCrtB CvCO, HDE1 promoter - CsCrtI CvCO	sAA06234 +PCR cassettes amplified from pAA2702, pAA2703, pAA2704, and pAA2311 using primers oAA02206 and oAA02209
sAA7870	GPD promoter - CsCrtE CvCO, GPD promoter - CsCrtI CvCO, GPD promoter - XdCrtYB CvCO, GPD promoter - XdCrtR CvCO, GPD promoter - AaCrtZ CvCO, GPD promoter - XdCrtS CvCO, GPD promoter - AaCrtW CvCO	sAA002 + PCR cassettes amplified from pAA2698, pAA2866, pAA2700, pAA2993, pAA2994, pAA2995, and pAA2996 using primers oAA02206 and oAA02209
sAA7882	GPD promoter - CsCrtE CvCO, GPD promoter - CsCrtI CvCO, GPD promoter - XdCrtYB CvCO	sAA002 + PCR cassettes amplified from pAA2698, pAA2866, and pAA2700 using primers oAA02206 and oAA02209
sAA8283	Δpox4, Δfaa1, Δfat1, Δcrc1, Δura3, G6PIpromoter - CRC1, HDE1 promoter - CsCrtE CvCO, HDE1 promoter - CsCrtB CvCO, HDE1 promoter - CsCrtI CvCO	sAA4377 + PCR cassettes amplified from pAA2311, pAA2702, pAA2703, sAA2704
sAA8519	Δpox4, Δfaa1, Δfat1, Δcrc1, Δura3, G6PIpromoter -CRC1, HDE1 promoter - CsCrtE CvCO, HDE1 promoter - XdCrtYB CvCO, HDE1 promoter - CsCrtI CvCO	sAA4377 + PCR cassettes amplified from pAA2311, pAA2702, pAA3373, sAA2704
sAA8503	GPD promoter - XdCrtE CvCO, GPD promoter - CsCrtB CvCO, GPD promoter - CsCrtI CvCO	sAA002 + PCR cassettes amplifieid from pAA3189, pAA2699, and pAA2700
sAA8932	GPD promoter - XdCrtE CvCO, GPD promoter - CsCrtB CvCO, GPD promoter - XdCrtI CvCO	sAA002 + PCR cassettes amplifieid from pAA3189, pAA2699, and pAA3490
sAA9432	HDE1 promoter - XdCrtE CvCO, HDE1 promoter - CsCrtB CvCO, HDE1 promoter - XdCrtI CvCO	sAA002 + PCR cassettes amplified from pAA3633, pAA2703, and pAA3634
sAA9539	Δras2/RAS2, Δura3	sAA002 + pVZ3930, after 5'FOA
sAA9670	Δras2/RAS2, HDE1 promoter - XdCrtE CvCO, HDE1 promoter -	sAA9539 + PCR cassettes amplified from pAA3633,

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1823 of 3077

Table 30		
Strain No.	**Description**	**Construction**
	CsCrtB CvCO, HDE1 promoter - XdCrtI CvCO	pAA2703, and pAA3634
sAA9658	Δura3, Δleu2::tURA3/LEU2	sAA002 + pAA3060 (BamHI/PstI) after 5'FOA
sAA9682	Δras2/RAS2, Δura3, Δleu2::tURA3/LEU2	sAA9539 + pAA3060 (BamHI/PstI) after 5'FOA
sAA9684	Δura3, Δleu2	sAA9658 + pAA2417 (BamHI/PstI) after 5'FOA
sAA9703	Δras2/RAS2, Δura3, Δleu2	sAA9682 + pAA3060 (BamHI/PstI) after 5'FOA
sAA9748	Δleu2, HDE1 promoter - XdCrtE CvCO (URA3), HDE1 promoter - CsCrtB CvCO (URA3), HDE1 promoter - XdCrtI CvCO (URA3)	sAA9684 + PCR cassettes amplified from pAA3633, pAA2703, and pAA3634
sAA9750	Δleu2, Δras2/RAS2, HDE1 promoter - XdCrtE CvCO (URA3), HDE1 promoter - CsCrtB CvCO (URA3), HDE1 promoter - XdCrtI CvCO (URA3)	sAA9703 + PCR cassettes amplified from pAA3633, pAA2703, and pAA3634
sAA9812	HDE1 promoter - XdCrtE CvCO (URA3), HDE1 promoter - CsCrtB CvCO (URA3), HDE1 promoter - XdCrtI CvCO (URA3) +leu2 library of (Ppox18Erg10, Ppox18Erg13, Ppox18HMG1, Ppox18Erg12, Ppox18Erg8, Ppox18Mvd1, Ppox18IDI1, Ppox18Erg20	sAA9748 + PCR cassettes amplified from pVZ4098, pVZ4099, pVZ4100, pVZ4101, pVZ4122, pVZ4123, pVZ4104, pVZ4105
sAA9811	HDE1 promoter - XdCrtE CvCO (URA3), HDE1 promoter - CsCrtB CvCO (URA3), HDE1 promoter - XdCrtI CvCO (URA3) +leu2 library of (Phde-XdCrtE, Phde-CsCrtB, Phde-XdCrtI)	sAA9748 + PCR cassettes amplified from pVZ4056, pVZ4057, pVZ4058
sAA9814	HDE1 promoter - XdCrtE CvCO (URA3), HDE1 promoter - CsCrtB CvCO (URA3), HDE1 promoter - XdCrtI CvCO (URA3) +leu2 library of (Phde-XdCrtE, Phde-CsCrtB, Phde-XdCrtI, Ppox18Erg10, Ppox18Erg13, Ppox18HMG1, Ppox18Erg12, Ppox18Mvd1, Ppox18IDI1, Ppox18Erg20)	sAA9748 + PCR cassettes amplified from pVZ4098, pVZ4099, pVZ4100, pVZ4101, pVZ4122, pVZ4123, pVZ4104, pVZ4105, pVZ4056, pVZ4057, pVZ4058
sAA9819	Δras2/RAS2, HDE1 promoter - XdCrtE CvCO (URA3), HDE1	sAA9750 + PCR cassettes amplified from pVZ4056,

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1824 of 3077

Table 30		
Strain No.	**Description**	**Construction**
	promoter - CsCrtB CvCO (URA3), HDE1 promoter - XdCrtI CvCO (URA3) +leu2 library of (Phde-XdCrtE, Phde-CsCrtB, Phde-XdCrtI)	pVZ4057, pVZ4058
sAA9817	Δras2/RAS2, HDE1 promoter - XdCrtE CvCO (URA3), HDE1 promoter - CsCrtB CvCO (URA3), HDE1 promoter - XdCrtI CvCO (URA3) + leu2 library of (Ppox18Erg10, Ppox18Erg13, Ppox18HMG1, Ppox18Erg12, Ppox18Mvd1, Ppox18IDI1, Ppox18Erg20)	sAA9750 + PCR cassettes amplified from pVZ4098, pVZ4099, pVZ4100, pVZ4101, pVZ4122, pVZ4123, pVZ4104, pVZ4105
sAA9821	Δras2/RAS2, HDE1 promoter - XdCrtE CvCO (URA3), HDE1 promoter - CsCrtB CvCO (URA3), HDE1 promoter - XdCrtI CvCO (URA3) + leu2 library of (Ppox18Erg10, Ppox18Erg13, Ppox18HMG1, Ppox18Erg12, Ppox18Mvd1, Ppox18Erg20, Phde-XdCrtE, Phde-CsCrtB, Phde-XdCrtI)	sAA9750 + PCR cassettes amplified from pVZ4098, pVZ4099, pVZ4100, pVZ4101, pVZ4122, pVZ4123, pVZ4104, pVZ4105, pVZ4056, pVZ4057, pVZ4058

EXAMPLE 75: Examples of embodiments

The examples set forth below illustrate certain embodiments and do not limit the technology.

A1. A genetically modified microorganism, comprising:

one or more heterologous nucleic acids encoding one or more terpene biosynthesis polypeptides, wherein expression of at least one of the heterologous nucleic acids is regulated by a nucleic acid that provides for fatty acid or alkane induction of expression of the terpene biosynthesis polypeptide.

A2. The microorganism of embodiment A1, wherein the microorganism is a fungus.

A3. The microorganism of embodiment A2, wherein the fungus is a yeast.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1825 of 3077

A4. The microorganism of embodiment A3, wherein the yeast is chosen from *Candida* spp, *Yarrowia* spp, *Rhodotorula* spp, *Rhodosporidium* spp, *Cryptococcus* spp, *Trichosporon* spp, *Lipomyces* spp, and *Blastobotrys* spp.

A5. The microorganism of any one of embodiments A1 to A4, wherein the one or more heterologous nucleic acids encode one or more terpene biosynthesis polypeptides chosen from terpene synthase, phytoene synthase, geranylgeranyl diphosphate synthase, phytoene desaturase, lycopene cyclase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase, zeaxanthin glucosyltransferase, valencene synthase, and cytochrome p450 reductase.

A6. The microorganism of any one of embodiments A1 to A5, wherein the fatty acid is a saturated fatty acid or an unsaturated fatty acid.

A7. The microorganism of any one of embodiments A1 to A6, wherein the fatty acid is chosen from one or more of oleic acid, palmitoleic acid, erucic acid, linoleic acid, palmitic acid, caproic acid, enanthic acid, caprylic acid pelargonic acid, capric acid, undecylic acid, lauric acid, myristic acid, pentadecanoic acid, margaric acid, stearic acid arachidic acid, behenic acid, tridecylic acid, and linolenic acid.

A7.1 The microorganism of any one of embodiments A1 to A5, wherein the alkane is chosen from one or more of hexane, heptane, nonane, decane, undecane, dodecane, tridecane, tetradecane, pentadecane, hexadecane, heptadecane, and octadecane

A8. The microorganism of any one of embodiments A1 to A7.1, wherein the nucleic acid that provides for fatty acid or alkane induction of expression of a terpene biosynthesis polypeptide comprises a fatty acid response element or an alkane response element.

A9. The microorganism of embodiment A8, wherein the fatty acid response element comprises an oleic acid response element.

A9.1 The microorganism of embodiment A8, wherein the alkane response element comprises an alkane response element 1 (ARE1) sequence.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1826 of 3077

A10. The microorganism of any one of embodiments A1 to A9, wherein the nucleic acid that provides for fatty acid induction of expression of a terpene biosynthesis polypeptide comprises a promoter region chosen from promoter regions of genes encoding hydratase-dehydrogenase-epimerase (HDE), acyl co-A oxidase (POX), acyl co-A thiolase (POT), peroxin (PEX) and peroxisomal adenine nucleotide transporter protein (ANT1).

A11. The microorganism of any one of embodiments A1 to A10, wherein the microorganism is *Candida viswanathii.*

A12. The microorganism of any one of embodiments A1 to A10, wherein the microorganism is *Blastobotrys adeninivorans*.

A13. The microorganism of any one of embodiments A1 to A12, wherein the one or more heterologous nucleic acids encode phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase.

A14. The microorganism of embodiment A13, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a promoter region of the gene encoding a hydratase-dehydrogenase-epimerase (HDE).

A14.1 The microorganism of embodiment A13, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a promoter region of the gene encoding a *Candida* hydratase-dehydrogenase-epimerase (HDE).

A14.2 The microorganism of embodiment A13, A14 or A14.1, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a terminator region of the gene encoding an acyl co-A oxidase 4 (POX4).

A14.3 The microorganism of embodiment A13, A14 or A14.1, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1827 of 3077

and phytoene desaturase is regulated by a terminator region of the gene encoding a *Candida* acyl co-A oxidase 4 (POX4).

A15. The microorganism of any one of embodiments A1 to A12, wherein the one or more heterologous nucleic acids encode valencene synthase.

A16. The microorganism of embodiment A15, wherein expression of the heterologous nucleic acids encoding valencene synthase is regulated by a promoter region of the gene encoding a hydratase-dehydrogenase-epimerase (HDE).

A16.1 The microorganism of embodiment A15, wherein expression of the heterologous nucleic acids encoding valencene synthase is regulated by a promoter region of the gene encoding a *Candida* hydratase-dehydrogenase-epimerase (HDE).

A16.2 The microorganism of embodiment A15, A16 or A16.1, wherein expression of the heterologous nucleic acids encoding valencene synthase is regulated by a terminator region of the gene encoding an acyl co-A oxidase 4 (POX4).

A16.3 The microorganism of embodiment A15, A16 or A16.1, wherein expression of the heterologous nucleic acids encoding valencene synthase is regulated by a terminator region of the gene encoding a *Candida* acyl co-A oxidase 4 (POX4).

A17. The microorganism of any one of embodiments A1 to A12, wherein the one or more heterologous nucleic acids encode geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase.

A18. The microorganism of embodiment A17, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase is regulated by a promoter region of the gene encoding a hydratase-dehydrogenase-epimerase (HDE).

A18.1 The microorganism of embodiment A17, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1828 of 3077

lycopene cyclase/phytoene synthase is regulated by a promoter region of the gene encoding a *Candida* hydratase-dehydrogenase-epimerase (HDE).

A18.2 The microorganism of embodiment A17, A18 or A18.1, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase is regulated by a terminator region of the gene encoding an acyl co-A oxidase 4 (POX4).

A18.3 The microorganism of embodiment A17, A18 or A18.1, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase is regulated by a terminator region of the gene encoding a *Candida* acyl co-A oxidase 4 (POX4).

A19. The microorganism of any one of embodiments A1 to A18.3, wherein the one or more heterologous nucleic acids encoding the one or more terpene biosynthesis polypeptides are endogenously expressed in a microorganism chosen from *Cronobacter* spp, *Callitropsis* spp, *Xanthophyllomyces* spp, *Agrobacterium* spp, and *Pantoea* spp.

A20. The microorganism of any one of embodiments A1 to A19, wherein the amount and/or activity of a Ras2 protein has been decreased.

A21. The microorganism of any one of embodiments A1 to A20, wherein the microorganism has been genetically modified to reduce or eliminate expression of an endogenous *RAS2* gene.

A22. The microorganism of any one of embodiments A1 to A21, wherein the amount and/or activity of an Faa1 protein has been decreased.

A23. The microorganism of any one of embodiments A1 to A22, wherein the microorganism has been genetically modified to reduce or eliminate expression of an endogenous *FAA1* gene.

A24. The microorganism of any one of embodiments A1 to A23, wherein the amounts and/or activities of one or more proteins chosen from acetyl-CoA C-acetyltransferase, HMG-CoA synthase, HMG-CoA reductase, mevalonate kinase, phosphomevalonate kinase,

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1829 of 3077

diphosphomevalonate decarboxylase, and isopentyl diphosphate delta isomerase, have been increased.

A25. The microorganism of any one of embodiments A1 to A24, wherein the microorganism has been genetically modified to increase expression of one or more endogenous genes chosen from *ERG10*, *ERG13*, *HMG1*, *ERG12*, *ERG8*, *MVD1*, and *IDI1*.

A26. The microorganism of any one of embodiments A1 to A25, wherein the amount and/or activity of dimethylallyltranstransferase or farnesyl diphosphate synthetase has been increased.

A27. The microorganism of any one of embodiments A1 to A26, wherein the microorganism has been genetically modified to increase expression of an endogenous *ERG20* gene.

A28. The microorganism of embodiment A25, wherein expression of *ERG10*, *ERG13*, *HMG1*, *ERG12*, *ERG8*, *MVD1*, and/or *IDI1* is regulated by a promoter region of the gene encoding oleate-induced peroxisomal protein (POX18).

A29. The microorganism of embodiment A25 or A28, wherein expression of *ERG10*, *ERG13*, *HMG1*, *ERG12*, *ERG8*, *MVD1*, and/or *IDI1* is regulated by a terminator region of the gene encoding oleate-induced peroxisomal protein (POX18).

A30. The microorganism of embodiment A27, wherein expression of *ERG20* is regulated by a promoter region of the gene encoding oleate-induced peroxisomal protein (POX18).

A31. The microorganism of embodiment A27 or A30, wherein expression of *ERG20* is regulated by a terminator region of the gene encoding oleate-induced peroxisomal protein (POX18).

B1. A genetically modified microorganism, comprising:

one or more heterologous nucleic acids encoding one or more terpene biosynthesis polypeptides, and

a genetic modification that alters the expression of a polypeptide providing for transport of acetyl-carnitine in the microorganism.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1830 of 3077

B1.1 The microorganism of embodiment B1, wherein expression of at least one of the heterologous nucleic acids is regulated by a nucleic acid that provides for fatty acid or alkane induction of expression of the terpene biosynthesis polypeptide.

B1.2 The microorganism of embodiment B1, wherein expression of at least one of the heterologous nucleic acids is regulated by a nucleic acid that provides for glucose induction of expression of the terpene biosynthesis polypeptide.

B2. The microorganism of embodiment B1, B1.1 or B1.2, wherein the microorganism is a fungus.

B3. The microorganism of embodiment B2, wherein the fungus is a yeast.

B4. The microorganism of embodiment B3, wherein the yeast is chosen from *Candida* spp, *Yarrowia* spp, *Rhodotorula* spp, *Rhodosporidium* spp, *Cryptococcus* spp, *Trichosporon* spp, *Lipomyces* spp, and *Blastobotrys* spp.

B5. The microorganism of any one of embodiments B1 to B4, wherein the one or more heterologous nucleic acids encode one or more terpene biosynthesis polypeptides chosen from terpene synthase, phytoene synthase, geranylgeranyl diphosphate synthase, phytoene desaturase, lycopene cyclase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase, zeaxanthin glucosyltransferase, valencene synthase, and cytochrome p450 reductase.

B6. The microorganism of any one of embodiments B1 to B5, wherein the genetic modification reduces expression of the polypeptide providing for transport of acetyl-carnitine in the microorganism relative to a microorganism that does not have the genetic modification.

B6.1 The microorganism of embodiment B6, wherein the genetic modification is a disruption, deletion or knockout of (i) a polynucleotide that encodes a polypeptide providing for transport of acetyl-carnitine, or (ii) a promoter operably linked to a polynucleotide that encodes a polypeptide providing for transport of acetyl-carnitine, whereby endogenous activity of a polypeptide providing for transport of acetyl-carnitine is reduced or abolished.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1831 of 3077

B7. The microorganism of embodiment B6, further comprising a nucleic acid encoding a polypeptide providing for transport of acetyl-carnitine wherein expression of the nucleic acid encoding a polypeptide providing for transport of acetyl-carnitine is regulated by a promoter that provides for reduced expression relative to endogenous expression.

B7.1 The microorganism of embodiment B7, wherein the genetic modification comprises replacing the promoter of an endogenous gene encoding the polypeptide providing for transport of acetyl-carnitine in the microorganism with a promoter that provides for reduced expression of the polypeptide in the microorganism relative to a microorganism that does not have the genetic modification.

B8. The microorganism of embodiment B7 or B7.1, wherein the promoter that provides for reduced expression relative to endogenous expression is a promoter for a glucose-6-phosphate isomerase (G6PI) gene.

B9. The microorganism of embodiment B7 or B7.1, wherein the promoter that provides for reduced expression relative to endogenous expression is a promoter for a *Candida* glucose-6-phosphate isomerase (G6PI) gene.

B10. The microorganism of any one of embodiments B1 to B9, wherein the polypeptide providing for transport of acetyl-carnitine in the microorganism is an acetyl-carnitine translocase (CRC).

B11. The microorganism of embodiment B10, wherein the acetyl-carnitine translocase (CRC) is acetyl-carnitine translocase 1 (CRC1).

B12. The microorganism of any one of embodiments B1 to B11, wherein the microorganism is *Candida viswanathii*.

B13. The microorganism of any one of embodiments B1 to B11, wherein the microorganism is *Blastobotrys adeninivorans*.

B14. The microorganism of any one of embodiments B1 to B13, wherein the one or more heterologous nucleic acids encode phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1832 of 3077

B15. The microorganism of embodiment B14, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a promoter region of the gene encoding a hydratase-dehydrogenase-epimerase (HDE).

B15.1 The microorganism of embodiment B14, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a promoter region of the gene encoding a *Candida* hydratase-dehydrogenase-epimerase (HDE).

B15.2 The microorganism of embodiment B14, B15 or B15.1, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a terminator region of the gene encoding an acyl co-A oxidase 4 (POX4).

B15.3 The microorganism of embodiment B14, B15 or B15.1, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a terminator region of the gene encoding a *Candida* acyl co-A oxidase 4 (POX4).

B16. The microorganism of any one of embodiments B1 to B13, wherein the one or more heterologous nucleic acids encode geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase.

B17. The microorganism of embodiment B16, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase is regulated by a promoter region of the gene encoding a hydratase-dehydrogenase-epimerase (HDE).

B17.1 The microorganism of embodiment B16, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase is regulated by a promoter region of the gene encoding a *Candida* hydratase-dehydrogenase-epimerase (HDE).

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1833 of 3077

B17.2 The microorganism of embodiment B16, B17 or B17.1, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase is regulated by a terminator region of the gene encoding an acyl co-A oxidase 4 (POX4).

B17.3 The microorganism of embodiment B16, B17 or B17.1, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase is regulated by a terminator region of the gene encoding a *Candida* acyl co-A oxidase 4 (POX4).

B18. The microorganism of any one of embodiments B1 to B17.3, wherein the one or more heterologous nucleic acids encoding the one or more terpene biosynthesis polypeptides are endogenously expressed in a microorganism chosen from *Cronobacter* spp, *Callitropsis* spp, *Xanthophyllomyces* spp, *Agrobacterium* spp, and *Pantoea* spp.

B19. The microorganism of any one of embodiments B1 to B18, wherein the amount and/or activity of a Ras2 protein has been decreased.

B20. The microorganism of any one of embodiments B1 to B19, wherein the microorganism has been genetically modified to reduce or eliminate expression of an endogenous *RAS2* gene.

B21. The microorganism of any one of embodiments B1 to B20, wherein the amount and/or activity of an Faa1 protein has been decreased.

B22. The microorganism of any one of embodiments B1 to B21, wherein the microorganism has been genetically modified to reduce or eliminate expression of an endogenous *FAA1* gene.

B23. The microorganism of any one of embodiments B1 to B22, wherein the amounts and/or activities of one or more proteins chosen from acetyl-CoA C-acetyltransferase, HMG-CoA synthase, HMG-CoA reductase, mevalonate kinase, phosphomevalonate kinase, diphosphomevalonate decarboxylase, and isopentyl diphosphate delta isomerase, have been increased.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1834 of 3077

B24. The microorganism of any one of embodiments B1 to B23, wherein the microorganism has been genetically modified to increase expression of one or more endogenous genes chosen from *ERG10*, *ERG13*, *HMG1*, *ERG12*, *ERG8*, *MVD1*, and *IDI1*.

B25. The microorganism of any one of embodiments B1 to B24, wherein the amount and/or activity of dimethylallyltranstransferase or farnesyl diphosphate synthetase has been increased.

B26. The microorganism of any one of embodiments B1 to B25, wherein the microorganism has been genetically modified to increase expression of an endogenous *ERG20* gene.

B27. The microorganism of embodiment B24, wherein expression of *ERG10*, *ERG13*, *HMG1*, *ERG12*, *ERG8*, *MVD1*, and/or *IDI1* is regulated by a promoter region of the gene encoding oleate-induced peroxisomal protein (POX18).

B28. The microorganism of embodiment B24 or B27, wherein expression of *ERG10*, *ERG13*, *HMG1*, *ERG12*, *ERG8*, *MVD1*, and/or *IDI1* is regulated by a terminator region of the gene encoding oleate-induced peroxisomal protein (POX18).

B29. The microorganism of embodiment B26, wherein expression of *ERG20* is regulated by a promoter region of the gene encoding oleate-induced peroxisomal protein (POX18).

B30. The microorganism of embodiment B26 or B29, wherein expression of *ERG20* is regulated by a terminator region of the gene encoding oleate-induced peroxisomal protein (POX18).

C1. A genetically modified *Candida viswanathii* yeast, comprising one or more heterologous nucleic acids encoding one or more terpene biosynthesis polypeptides.

C1.1 The genetically modified *Candida viswanathii* yeast of embodiment C1, wherein expression of at least one of the heterologous nucleic acids is regulated by a nucleic acid that provides for fatty acid or alkane induction of expression of the terpene biosynthesis polypeptide.

C1.2 The genetically modified *Candida viswanathii* yeast of embodiment C1, wherein expression of at least one of the heterologous nucleic acids is regulated by a nucleic acid that provides for glucose induction of expression of the terpene biosynthesis polypeptide.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1835 of 3077

C2. The genetically modified *Candida viswanathii* yeast of embodiment C1, C1.1 or C1.2, wherein the one or more heterologous nucleic acids encode one or more terpene biosynthesis polypeptides chosen from terpene synthase, phytoene synthase, geranylgeranyl diphosphate synthase, phytoene desaturase, lycopene cyclase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase, zeaxanthin glucosyltransferase, valencene synthase, and cytochrome p450 reductase.

C3. The genetically modified *Candida viswanathii* yeast of any one of embodiments C1 to C2, wherein the one or more heterologous nucleic acids encode phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase

C4. The genetically modified *Candida viswanathii* yeast of embodiment C3, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a promoter region of the gene encoding a *Candida* hydratase-dehydrogenase-epimerase (HDE).

C4.1 The genetically modified *Candida viswanathii* yeast of embodiment C4, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a terminator region of the gene encoding a *Candida* acyl co-A oxidase 4 (POX4).

C5. The genetically modified *Candida viswanathii* yeast of embodiment C3, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a promoter region of the gene encoding a *Candida* glyceraldehyde-3-phosphate dehydrogenase (GPD).

C6. The genetically modified *Candida viswanathii* yeast of any one of embodiments C1 to C2, wherein the one or more heterologous nucleic acids encode valencene synthase.

C7. The genetically modified *Candida viswanathii* yeast of embodiment C6, wherein expression of the heterologous nucleic acids encoding valencene synthase is regulated by a promoter region of the gene encoding a *Candida* hydratase-dehydrogenase-epimerase (HDE).

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1836 of 3077

C7.1 The genetically modified *Candida viswanathii* yeast of embodiment C7, wherein expression of the heterologous nucleic acids valencene synthase is regulated by a terminator region of the gene encoding a *Candida* acyl co-A oxidase 4 (POX4).

C8. The genetically modified *Candida viswanathii* yeast of embodiment C6, wherein expression of the heterologous nucleic acids encoding valencene synthase is regulated by a promoter region of the gene encoding a *Candida* glyceraldehyde-3-phosphate dehydrogenase (GPD).

C9. The genetically modified *Candida viswanathii* yeast of any one of embodiments C1 to C2, wherein the one or more heterologous nucleic acids encode geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase.

C10. The genetically modified *Candida viswanathii* yeast of embodiment C9, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase is regulated by a promoter region of the gene encoding a *Candida* hydratase-dehydrogenase-epimerase (HDE).

C10.1 The genetically modified *Candida viswanathii* yeast of embodiment C10, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase is regulated by a terminator region of the gene encoding a *Candida* acyl co-A oxidase 4 (POX4).

C11. The genetically modified *Candida viswanathii* yeast of embodiment C9, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase is regulated by a promoter region of the gene encoding a *Candida* glyceraldehyde-3-phosphate dehydrogenase (GPD).

C12. The genetically modified *Candida viswanathii* yeast of any one of embodiments C1 to C2, wherein the one or more heterologous nucleic acids encode geranylgeranyl diphosphate synthase, phytoene desaturase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase, and cytochrome p450 reductase.

C13. The genetically modified *Candida viswanathii* yeast of embodiment C12, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1837 of 3077

desaturase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase, and cytochrome p450 reductase is regulated by a promoter region of the gene encoding a *Candida* hydratase-dehydrogenase-epimerase (HDE).

C13.1 The genetically modified *Candida viswanathii* yeast of embodiment C13, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase, and cytochrome p450 reductase is regulated by a terminator region of the gene encoding a *Candida* acyl co-A oxidase 4 (POX4).

C14. The genetically modified *Candida viswanathii* yeast of embodiment C12, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase, and cytochrome p450 reductase is regulated by a promoter region of the gene encoding a *Candida* glyceraldehyde-3-phosphate dehydrogenase (GPD).

C15. The genetically modified *Candida viswanathii* yeast of any one of embodiments C1 to C14, wherein the one or more heterologous nucleic acids encoding the one or more terpene biosynthesis polypeptides are endogenously expressed in a microorganism chosen from *Cronobacter* spp, *Callitropsis* spp, *Xanthophyllomyces* spp, *Agrobacterium* spp, and *Pantoea* spp.

C16. The genetically modified *Candida viswanathii* yeast of any one of embodiments C1 to C15, further comprising a genetic modification that alters the expression of one or more nucleic acids encoding one or more endogenous polypeptides.

C17. The genetically modified *Candida viswanathii* yeast of embodiment C16, wherein the one or more endogenous polypeptides are chosen from polypeptides having one or more of the following activities: acyl-CoA synthetase activity, acyl-CoA oxidase activity, ATP-binding cassette transporter activity, carnitine acetyltransferase activity, transport of acetyl-carnitine, acyl-CoA thioesterase activity, acyl-CoA hydrolase activity, aldehyde dehydrogenase activity, monooxygenase activity, or monooxgenase reductase activity.

C18. The genetically modified *Candida viswanathii* yeast of any one of embodiments C1 to C17, wherein the amount and/or activity of a Ras2 protein has been decreased.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1838 of 3077

C19. The genetically modified *Candida viswanathii* yeast of any one of embodiments C1 to C18, wherein the yeast has been genetically modified to reduce or eliminate expression of an endogenous *RAS2* gene.

C20. The genetically modified *Candida viswanathii* yeast of any one of embodiments C1 to C19, wherein the amount and/or activity of an Faa1 protein has been decreased.

C21. The genetically modified *Candida viswanathii* yeast of any one of embodiments C1 to C20, wherein the yeast has been genetically modified to reduce or eliminate expression of an endogenous *FAA1* gene.

C22. The genetically modified *Candida viswanathii* yeast of any one of embodiments C1 to C21, wherein the amounts and/or activities of one or more proteins chosen from acetyl-CoA C-acetyltransferase, HMG-CoA synthase, HMG-CoA reductase, mevalonate kinase, phosphomevalonate kinase, diphosphomevalonate decarboxylase, and isopentyl diphosphate delta isomerase, have been increased.

C23. The genetically modified *Candida viswanathii* yeast of any one of embodiments C1 to C22, wherein the yeast has been genetically modified to increase expression of one or more endogenous genes chosen from *ERG10*, *ERG13*, *HMG1*, *ERG12*, *ERG8*, *MVD1*, and *IDI1*.

C24. The genetically modified *Candida viswanathii* yeast of any one of embodiments C1 to C23, wherein the amount and/or activity of dimethylallyltranstransferase or farnesyl diphosphate synthetase has been increased.

C25. The genetically modified *Candida viswanathii* yeast of any one of embodiments C1 to C24, wherein the yeast has been genetically modified to increase expression of an endogenous *ERG20* gene.

C26. The genetically modified *Candida viswanathii* yeast of embodiment C23, wherein expression of *ERG10*, *ERG13*, *HMG1*, *ERG12*, *ERG8*, *MVD1*, and/or *IDI1* is regulated by a promoter region of the gene encoding oleate-induced peroxisomal protein (POX18).

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1839 of 3077

C27. The genetically modified *Candida viswanathii* yeast of embodiment C23 or C26, wherein expression of *ERG10*, *ERG13*, *HMG1*, *ERG12*, *ERG8*, *MVD1*, and/or *IDI1* is regulated by a terminator region of the gene encoding oleate-induced peroxisomal protein (POX18).

C28. The genetically modified *Candida viswanathii* yeast of embodiment C25, wherein expression of *ERG20* is regulated by a promoter region of the gene encoding oleate-induced peroxisomal protein (POX18).

C29. The genetically modified *Candida viswanathii* yeast of embodiment C25 or C28, wherein expression of *ERG20* is regulated by a terminator region of the gene encoding oleate-induced peroxisomal protein (POX18).

D1. A method for producing a terpene comprising:

contacting the genetically modified microorganism of any one of embodiments A1 to A31 with a feedstock comprising a carbon source, and

culturing the microorganism under conditions in which the terpenes are produced from the feedstock.

D2. A method for producing a terpene comprising:

contacting the genetically modified microorganism of any one of embodiments B1 to B30 with a feedstock comprising a carbon source, and

culturing the microorganism under conditions in which the terpenes are produced from the feedstock.

D3. A method for producing a terpene comprising:

contacting the genetically modified *Candida viswanathii* yeast of any one of embodiments C1 to C29 with a feedstock comprising a carbon source, and

culturing the microorganism under conditions in which the terpenes are produced from the feedstock.

D4. The method of any one of embodiments D1 to D3, wherein the feedstock comprises one or more fatty acids.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1840 of 3077

D5. The method of embodiment D4, wherein the feedstock comprises one or more saturated fatty acids or one or more unsaturated fatty acids; or one or more saturated fatty acids and one or more unsaturated fatty acids.

D6. The method of embodiment D4 or D5, wherein the feedstock comprises one or more fatty acids chosen from oleic acid, palmitoleic acid, erucic acid, linoleic acid, palmitic acid, caproic acid, enanthic acid, caprylic acid pelargonic acid, capric acid, undecylic acid, lauric acid, myristic acid, pentadecanoic acid, margaric acid, stearic acid arachidic acid, behenic acid, tridecylic acid, and linolenic acid.

D7. The method of embodiment D6 wherein the feedstock comprises oleic acid.

D8. The method of any one of embodiments D1 to D3, wherein the feedstock comprises one or more sugars.

D9. The method of D8, wherein the feedstock comprises glucose.

D10. The method of any one of embodiments D1 to D3, wherein the feedstock comprises one or more alkane hydrocarbons.

D10.1 The method of embodiment D10, wherein the feedstock comprises one or more alkane hydrocarbons chosen from one or more of hexane, heptane, nonane, decane, undecane, dodecane, tridecane, tetradecane, pentadecane, hexadecane, heptadecane, and octadecane.

D11. The method of embodiment D10, wherein the feedstock comprises one or more alkane hydrocarbons chosen from C10 to C14 alkane hydrocarbons, nonane and octadecane.

D12. The method of any one of embodiments D1 to D3, wherein the feedstock comprises one or more vegetable oils, derivatives thereof, or byproducts thereof.

D13. The method of embodiment D12, wherein the feedstock comprises crude palm oil (CPO).

D14. The method of embodiment D12, wherein the feedstock comprises palm oil fatty acid distillate (PFAD).

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1841 of 3077

D15. The method of any one of embodiments D1 to D14, wherein the terpene is chosen from lycopene, beta carotene and astaxanthin.

D16. The method of any one of embodiments D1 to D15, wherein the yield or titer of the terpene is between about 0.001 g/L to about 2.0 g/L.

D16.1 The method of any one of embodiments D1 to D15, wherein the yield or titer of the terpene is between about 0.05 g/L to about 0.5 g/L.

D17. The method of any one of embodiments D1 to D15, wherein the yield or titer of the terpene is between about 0.001 g/L to about 0.2 g/L.

D18. The method of embodiment D16, wherein the yield or titer of lycopene is between about 0.03 g/L to about 2.0 g/L.

D18.1 The method of embodiment D16.1, wherein the yield or titer of lycopene is between about 0.05 g/L to about 0.5 g/L.

D19. The method of embodiment D17, wherein the yield or titer of lycopene is between about 0.001 g/L to about 0.2 g/L.

D20. The method of embodiment D16, wherein the yield or titer of beta carotene is between about 0.1 g/L to about 0.2 g/L.

D21. The method of embodiment D17, wherein the yield or titer of beta carotene is between about 0.01 g/L to about 0.1 g/L.

D22. The method of embodiment D17, wherein the yield or titer of beta carotene is between about 0.001 g/L to about 0.01 g/L.

D23. The method of embodiment D16, wherein the yield or titer of astaxanthin is between about 0.005 g/L to about 0.01 g/L.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1842 of 3077

D24. The method of embodiment D17, wherein the yield or titer of astaxanthin is between about 0.001 g/L to about 0.005 g/L.

D25. The method of any one of embodiments D1 to D24, further comprising isolating the terpene.

E1. A genetically modified yeast comprising one or more heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase.

E2. A genetically modified yeast comprising one or more heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase.

E3. A genetically modified yeast comprising one or more heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase, and cytochrome p450 reductase.

E4. The genetically modified yeast of embodiment E1, E2 or E3, wherein the amount and/or activity of a Ras2 protein has been decreased.

E5. The genetically modified yeast of any one of embodiments E1 to E4, wherein the yeast has been genetically modified to reduce or eliminate expression of an endogenous *RAS2* gene.

E6. The genetically modified yeast of any one of embodiments E1 to E5, wherein the amount and/or activity of an Faa1 protein has been decreased.

E7. The genetically modified yeast of any one of embodiments E1 to E6, wherein the yeast has been genetically modified to reduce or eliminate expression of an endogenous *FAA1* gene.

E8. The genetically modified yeast of any one of embodiments E1 to E7, wherein the amounts and/or activities of one or more proteins chosen from acetyl-CoA C-acetyltransferase, HMG-CoA synthase, HMG-CoA reductase, mevalonate kinase, phosphomevalonate kinase, diphosphomevalonate decarboxylase, and isopentyl diphosphate delta isomerase, have been increased.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1843 of 3077

E9. The genetically modified yeast of any one of embodiments E1 to E8, wherein the yeast has been genetically modified to increase expression of one or more endogenous genes chosen from *ERG10*, *ERG13*, *HMG1*, *ERG12*, *ERG8*, *MVD1*, and *IDI1*.

E10. The genetically modified yeast of any one of embodiments E1 to E9, wherein the amount and/or activity of dimethylallyltranstransferase or farnesyl diphosphate synthetase has been increased.

E11. The genetically modified yeast of any one of embodiments E1 to E10, wherein the yeast has been genetically modified to increase expression of an endogenous *ERG20* gene.

E12. The genetically modified yeast of any one of embodiments E1 to E11, wherein the yeast is *Candida viswanathii.*

E13. The genetically modified yeast of any one of embodiments E1 to E11, wherein the yeast is *Blastobotrys adeninivorans*.

E14. Use of the genetically modified yeast of any one of embodiments E1 to E13 for the production of a terpene.

E15. Use of the genetically modified yeast of any one of embodiments E1 and E4 to E14 for the production of lycopene.

E16. Use of the genetically modified yeast of any one of embodiments E2 and E4 to E14 for the production of beta carotene.

E17. Use of the genetically modified yeast of any one of embodiments E3 to E14 for the production of astaxanthin.

* * *

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1844 of 3077

The entirety of each patent, patent application, publication and document referenced herein hereby is incorporated by reference. Citation of the above patents, patent applications, publications and documents is not an admission that any of the foregoing is pertinent prior art, nor does it constitute any admission as to the contents or date of these publications or documents. Their citation is not an indication of a search for relevant disclosures. All statements regarding the date(s) or contents of the documents is based on available information and is not an admission as to their accuracy or correctness.

Modifications may be made to the foregoing without departing from the basic aspects of the technology. Although the technology has been described in substantial detail with reference to one or more specific embodiments, those of ordinary skill in the art will recognize that changes may be made to the embodiments specifically disclosed in this application, yet these modifications and improvements are within the scope and spirit of the technology.

The technology illustratively described herein suitably may be practiced in the absence of any element(s) not specifically disclosed herein. Thus, for example, in each instance herein any of the terms "comprising," "consisting essentially of," and "consisting of" may be replaced with either of the other two terms. The terms and expressions which have been employed are used as terms of description and not of limitation, and use of such terms and expressions do not exclude any equivalents of the features shown and described or portions thereof, and various modifications are possible within the scope of the technology claimed. The term "a" or "an" can refer to one of or a plurality of the elements it modifies (e.g., "a reagent" can mean one or more reagents) unless it is contextually clear either one of the elements or more than one of the elements is described. The term "about" as used herein refers to a value within 10% of the underlying parameter (i.e., plus or minus 10%), and use of the term "about" at the beginning of a string of values modifies each of the values (i.e., "about 1, 2 and 3" refers to about 1, about 2 and about 3). For example, a weight of "about 100 grams" can include weights between 90 grams and 110 grams. Further, when a listing of values is described herein (e.g., about 50%, 60%, 70%, 80%, 85% or 86%) the listing includes all intermediate and fractional values thereof (e.g., 54%, 85.4%). Thus, it should be understood that although the present technology has been specifically disclosed by representative embodiments and optional features, modification and variation of the concepts herein disclosed may be resorted to by those skilled in the art, and such modifications and variations are considered within the scope of this technology.

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1845 of 3077

Certain embodiments of the technology are set forth in the claim(s) that follow(s).

LCY Biotechnology Holding, Inc.
Ex. 1002
Page 1846 of 3077