



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	FILING or 371(c) DATE	GRP ART UNIT	FIL FEE REC'D	ATTY. DOCKET NO	TOT CLAIMS	IND CLAIMS
61/072,160	03/28/2008		105	105-005900US		

CONFIRMATION NO. 2292

22798

QUINE INTELLECTUAL PROPERTY LAW GROUP, P.C.
P O BOX 458
ALAMEDA, CA 94501

FILING RECEIPT



OC00000029272849

Date Mailed: 04/09/2008

Receipt is acknowledged of this provisional patent application. It will not be examined for patentability and will become abandoned not later than twelve months after its filing date. Any correspondence concerning the application must include the following identification information: the U.S. APPLICATION NUMBER, FILING DATE, NAME OF APPLICANT, and TITLE OF INVENTION. Fees transmitted by check or draft are subject to collection. Please verify the accuracy of the data presented on this receipt. **If an error is noted on this Filing Receipt, please write to the Office of Initial Patent Examination's Filing Receipt Corrections. Please provide a copy of this Filing Receipt with the changes noted thereon. If you received a "Notice to File Missing Parts" for this application, please submit any corrections to this Filing Receipt with your reply to the Notice. When the USPTO processes the reply to the Notice, the USPTO will generate another Filing Receipt incorporating the requested corrections**

Applicant(s)

Pranav Patel, Fremont, CA;
Keith Bjornson, San Mateo, CA;
Kevin Travers, Santa Clara, CA;
Cheryl Heiner, San Mateo, CA;

Power of Attorney: The patent practitioners associated with Customer Number 22798

If Required, Foreign Filing License Granted: 04/07/2008

The country code and number of your priority application, to be used for filing abroad under the Paris Convention, is **US 61/072,160**

Projected Publication Date: None, application is not eligible for pre-grant publication

Non-Publication Request: No

Early Publication Request: No

**** SMALL ENTITY ****

Title

Methods and compositions for nucleic acid sample preparation

PROTECTING YOUR INVENTION OUTSIDE THE UNITED STATES

Since the rights granted by a U.S. patent extend only throughout the territory of the United States and have no effect in a foreign country, an inventor who wishes patent protection in another country must apply for a patent in a specific country or in regional patent offices. Applicants may wish to consider the filing of an international application under the Patent Cooperation Treaty (PCT). An international (PCT) application generally has the same effect as a regular national patent application in each PCT-member country. The PCT process **simplifies** the filing

page 1 of 3

Univ. of Washington EX2049

0001 Guardant Health, Inc v. Univ. of Washington

IPR2022-00817

of patent applications on the same invention in member countries, but **does not result** in a grant of "an international patent" and does not eliminate the need of applicants to file additional documents and fees in countries where patent protection is desired.

Almost every country has its own patent law, and a person desiring a patent in a particular country must make an application for patent in that country in accordance with its particular laws. Since the laws of many countries differ in various respects from the patent law of the United States, applicants are advised to seek guidance from specific foreign countries to ensure that patent rights are not lost prematurely.

Applicants also are advised that in the case of inventions made in the United States, the Director of the USPTO must issue a license before applicants can apply for a patent in a foreign country. The filing of a U.S. patent application serves as a request for a foreign filing license. The application's filing receipt contains further information and guidance as to the status of applicant's license for foreign filing.

Applicants may wish to consult the USPTO booklet, "General Information Concerning Patents" (specifically, the section entitled "Treaties and Foreign Patents") for more information on timeframes and deadlines for filing foreign patent applications. The guide is available either by contacting the USPTO Contact Center at 800-786-9199, or it can be viewed on the USPTO website at <http://www.uspto.gov/web/offices/pac/doc/general/index.html>.

For information on preventing theft of your intellectual property (patents, trademarks and copyrights), you may wish to consult the U.S. Government website, <http://www.stopfakes.gov>. Part of a Department of Commerce initiative, this website includes self-help "toolkits" giving innovators guidance on how to protect intellectual property in specific countries such as China, Korea and Mexico. For questions regarding patent enforcement issues, applicants may call the U.S. Government hotline at 1-866-999-HALT (1-866-999-4158).

LICENSE FOR FOREIGN FILING UNDER

Title 35, United States Code, Section 184

Title 37, Code of Federal Regulations, 5.11 & 5.15

GRANTED

The applicant has been granted a license under 35 U.S.C. 184, if the phrase "IF REQUIRED, FOREIGN FILING LICENSE GRANTED" followed by a date appears on this form. Such licenses are issued in all applications where the conditions for issuance of a license have been met, regardless of whether or not a license may be required as set forth in 37 CFR 5.15. The scope and limitations of this license are set forth in 37 CFR 5.15(a) unless an earlier license has been issued under 37 CFR 5.15(b). The license is subject to revocation upon written notification. The date indicated is the effective date of the license, unless an earlier license of similar scope has been granted under 37 CFR 5.13 or 5.14.

This license is to be retained by the licensee and may be used at any time on or after the effective date thereof unless it is revoked. This license is automatically transferred to any related applications(s) filed under 37 CFR 1.53(d). This license is not retroactive.

The grant of a license does not in any way lessen the responsibility of a licensee for the security of the subject matter as imposed by any Government contract or the provisions of existing laws relating to espionage and the national security or the export of technical data. Licensees should apprise themselves of current regulations especially with respect to certain countries, of other agencies, particularly the Office of Defense Trade Controls, Department of

State (with respect to Arms, Munitions and Implements of War (22 CFR 121-128)); the Bureau of Industry and Security, Department of Commerce (15 CFR parts 730-774); the Office of Foreign Assets Control, Department of Treasury (31 CFR Parts 500+) and the Department of Energy.

NOT GRANTED

No license under 35 U.S.C. 184 has been granted at this time, if the phrase "IF REQUIRED, FOREIGN FILING LICENSE GRANTED" DOES NOT appear on this form. Applicant may still petition for a license under 37 CFR 5.12, if a license is desired before the expiration of 6 months from the filing date of the application. If 6 months has lapsed from the filing date of this application and the licensee has not received any indication of a secrecy order under 35 U.S.C. 181, the licensee may foreign file the application pursuant to 37 CFR 5.15(b).

PROVISIONAL APPLICATION TRANSMITTAL SHEET

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

U.S. PTO
61/072160
03/28/2008

Transmitted herewith for filing is a provisional patent application under CFR 1.53(c) of Inventors:

First Name and Middle Name [if any] followed by Last Name	RESIDENCE (City/State/Country)
Pranav Patel	Fremont/California/United States
Keith Bjornson	San Mateo/California/United States
Kevin Travers	Santa Clara/California//United States
Cheryl Heiner	San Mateo/California//United States

Title: METHODS AND COMPOSITIONS FOR NUCLEIC ACID SAMPLE PREPARATION

Enclosed are:

- ☒ 40 pages of the application (including cover, description, claims, and abstract).
☒ 07 sheet(s) of [] formal [x] informal drawing(s).
☒ Abstract.
☒ 92 claims.
☒ Small Entity Status is Claimed
☒ Cover Sheet.
☒ Patent Application Filing Acknowledgement postcard.
☐ the invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.
☒ Application Data Entry Sheet.

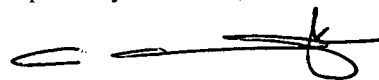
Please charge Deposit Account No. 50-0893 as follows:

- ☒ Provisional Filing fee. \$ 105.00
☒ Provisional Application Size Fee
(Total pages: 48 - 100 = 0/50 = 0 x \$130.00) \$ 0.00
☒ Any additional fees associated with this paper or during the pendency of this application.

Total \$ 105.00

1 extra copy of this sheet is enclosed.

Respectfully submitted,



Christina Onufryk, Ph.D.

Reg No.: 59,842

Applicant's Representative

CORRESPONDENCE ADDRESS:
QUINE INTELLECTUAL PROPERTY LAW GROUP, P.C.
P.O. Box 458
Alameda, CA 94501
Telephone: (510) 337-7871
Fax: (510) 337-7877

Customer No. 22798



22798

PATENT TRADEMARK OFFICE

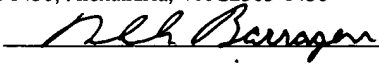
Attorney Docket No. 105-005900US

Client Reference No. 01-005900US

Express Mail Label No. EV 957134874 US

Date of Deposit: March 28, 2008

I hereby certify that this is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR 1.10 on the date indicated above, addressed to: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450

By: 
Deborah Barragan

Application Information

Application Type:: Provisional
Subject Matter::
Suggested Classification::
Suggested Group Art Unit ::
CD-ROM or CD-R?::
Number of CD disks::
Number of copies of CDs::
Sequence submission::
Computer Readable Form
(CRF)?::
Number of copies of CRF::
Title Line One:: METHODS AND COMPOSITIONS FOR
Title Line Two:: NUCLEIC ACID SAMPLE PREPARATION
Attorney Docket Number:: 105-005900US
Request for Early Publication?:: No
Request for Non—Publication?:: No
Suggested Drawing Figure::
Total Drawing Sheets::
Small Entity:: Yes
Petition included?::
Petition Type::
Licensed US GovtAgency::
Contract or Grant Numbers::
Secrecy Order in Parent Appl.?::

Applicant Information

Applicant Authority Type:: Inventor
Primary Citizenship Country:: India
Status::
Given Name:: Pranav
Middle Name::
Family Name:: Patel
Name Suffix::
City of Residence:: Fremont
State or Province of Residence:: California
Country of Residence:: United States
Street of mailing address:: 34203 O'Neil Terrace
City of mailing address:: Fremont
State of mailing address:: California
Country of mailing address:: United States
Postal or Zip Code of mailing
address:: 94555

Applicant Authority Type:: Inventor
Primary Citizenship Country:: United States
Status::
Given Name:: Keith
Middle Name::
Family Name:: Bjornson
Name Suffix::
City of Residence:: San Mateo
State or Province of Residence:: California
Country of Residence:: United States
Street of mailing address:: 833 Newbridge Avenue
City of mailing address:: San Mateo
State of mailing address:: California
Country of mailing address:: United States
Postal or Zip Code of mailing address:: 94401

Applicant Authority Type:: Inventor
Primary Citizenship Country:: United States
Status::
Given Name:: Kevin
Middle Name::
Family Name:: Travers
Name Suffix::
City of Residence:: Santa Clara
State or Province of Residence:: California
Country of Residence:: United States
Street of mailing address:: 3385 Gonzaga Place
City of mailing address:: Santa Clara
State of mailing address:: California
Country of mailing address:: United States
Postal or Zip Code of mailing address:: 95051

Applicant Authority Type:: Inventor
Primary Citizenship Country:: United States
Status::
Given Name:: Cheryl
Middle Name::
Family Name:: Heiner
Name Suffix::
City of Residence:: San Mateo
State or Province of Residence:: California
Country of Residence:: United States
Street of mailing address:: 1952 Ticonderoga Drive

City of mailing address:: San Mateo
State of mailing address:: California
Country of mailing address:: United States
Postal or Zip Code of mailing address:: 94402

Correspondence Information

Correspondence Customer Number:: 22798
Name:: Quine Intellectual Property Law Group
Street of mailing address:: P.O. Box 458
City of mailing address:: Alameda
State or Providence of mailing address:: California
Country of mailing address:: U.S.A.
Postal or Zip Code of mailing address:: 94501
Phone number:: (510) 337-7871
Fax number:: (510) 337-7877
E-Mail address:: jaquine@quinelaw.com

Representative Information

Representative Customer Number::	22798
----------------------------------	-------

Assignee Information

Assignee Name:: Pacific Biosciences of California, Inc.

Express Mail® Label No. EV 957134874 US
Date of Deposit: March 28, 2008
I hereby certify that this is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR 1.10 on the date indicated above, addressed to: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450
By: *Deborah Barragan*
Deborah Barragan

Attorney Docket: 105-005900US

Client Ref. No. 01-005900US

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE
PROVISIONAL Patent Application For

METHODS AND COMPOSITIONS FOR NUCLEIC ACID
SAMPLE PREPARATION

Inventor(s): **Pranav Patel**, a citizen of India, residing at 34203 O'Neil Terrace, Fremont, California 94555;
Keith Bjornson, a citizen of the United States of America, residing at 833 Newbridge Avenue, San Mateo, California 94401;
Kevin Travers, a citizen of the United States of America, residing at 3385 Gonzaga Place, Santa Clara California 95051.;
Cheryl Heiner, a citizen of the United States of America, residing at 1952 Ticonderoga Drive, San Mateo, California 94402.

Assignee: **Pacific Biosciences of California, Inc.**
1505 Adams Drive
Menlo Park, California 94025

Entity: **Small**

QUINE INTELLECTUAL PROPERTY LAW GROUP, P.C.

P.O. Box 458
Alameda, CA 94501
Internet address: www.quinelaw.com

Phone: (510) 337-7871
Fax: (510) 337-7877
E-mail: jaquine@quinelaw.com

**METHODS, COMPOSITIONS AND SYSTEMS FOR NUCLEIC ACID
SAMPLE PREPARATION**

**STATEMENT AS TO RIGHTS TO INVENTIONS MADE UNDER FEDERALLY
SPONSORED RESEARCH AND DEVELOPMENT**

Not applicable.

FIELD OF THE INVENTION

[0001] This invention is in the field of nucleic acid sequencing, particularly the preparation of templates for nucleic acid sequencing.

BACKGROUND OF THE INVENTION

[0002] Nucleic acid sequence data is valuable in myriad applications in biological research and molecular medicine, including determining the hereditary factors in disease, in developing new methods to detect disease and guide therapy (van de Vijver et al. (2002) "A gene-expression signature as a predictor of survival in breast cancer," *New England Journal of Medicine* **347**: 1999-2009), and in providing a rational basis for personalized medicine. Obtaining and verifying sequence data for use in such analyses has made it necessary for sequencing technologies to undergo advancements to expand throughput, lower reagent and labor costs and improve accuracy (*See, e.g.,* Chan, et al. (2005) "Advances in Sequencing Technology" (Review) *Mutation Research* **573**: 13-40, Levene et al. (2003) "Zero Mode Waveguides for Single Molecule Analysis at High Concentrations," *Science* **299**: 682-686).

[0003] Current methods for preparing nucleic acid templates are not optimal for use in high throughput DNA sequencing systems. Conventional cloning and cell culture methods are time consuming and expensive. Lengthy nucleic acid purification protocols currently in use do not reliably produce nucleic acid samples that are sufficiently free of sequencing reaction inhibitors such as salts, carbohydrates and/or proteins. Furthermore, these problems are magnified when such conventional techniques are scaled to the quantities that would be useful for high throughput sequencing technologies. Consequently, there is an increasing demand for efficient, low-cost methods for the preparation of high-quality nucleic acid templates. The present invention provides methods and compositions that would be useful for supplying high throughput DNA sequencing systems with such templates.

SUMMARY

[0004] The present invention provides methods and compositions that can be useful for supplying high throughput DNA sequencing systems with templates. The methods circumvent the need for costly, labor-intensive cloning and cell culture methods and can be scaled to accommodate template production for a variety of sequencing applications, e.g., sequencing individuals' genomes and/or gene expression profiling (Spinella, et al. (1999) "Tandem arrayed ligation of expressed sequence tags (TALEST): a new method for generating global gene expression profiles." *Nucleic Acids Res* **27**: e22, Velculescu, et al. (1995) "Serial analysis of gene expression." *Science* **270**: 484–487). The methods and compositions provided by the invention can be used to produce either linear or circular single-stranded nucleic acid templates.

[0005] The invention provides a first set of methods of producing a population of single-stranded nucleic acids. In a first set of methods, a genomic DNA, a cDNA, or a DNA concatamer provided. Double-stranded fragments that comprise first exonuclease-sensitive strands and second exonuclease-resistant strands are generated from the genomic DNA, the cDNA, or the concatamer. The methods include eliminating the exonuclease-sensitive strands of the double-stranded fragment to produce the population of single-stranded nucleic acids.

[0006] The genomic DNA, the cDNA, or the concatamer from which the double-stranded fragments are generated can optionally be derived from a eukaryote, a prokaryote, an archaea, a virus, or a phage. Generating the double-stranded fragments that comprise first exonuclease-sensitive strands and second exonuclease-resistant strands can optionally comprise cleaving the genomic DNA, cDNA, or concatamer, e.g., via enzymatic digestion, sonication, mechanical shearing, electrochemical cleavage, and/or nebulization, to produce linear double-stranded nucleic acids. Optionally, exonuclease-sensitive tags can be ligated to the 5' ends of the first strands of the double-stranded nucleic acids and exonuclease-resistant tags can be ligated to the 5' ends of the second strands of the double-stranded nucleic acids to produce the double stranded fragments with first exonuclease-sensitive strands and second exonuclease-resistant strands.

[0007] As used herein, a "tag" refers to a moiety linked to a nucleic acid of interest that can be used as a molecular recognition site to identify or distinguish the nucleic acid in a population, as a means to permit a protein, e.g. a DNA-binding protein, or an enzyme,

e.g., an exonuclease, a restriction enzyme, a nicking enzyme, or the like, to recognize the nucleic acid and perform an activity, and/or as a means by which to separate the nucleic acid from the population. A tag can comprise one or more of a number of moieties, including labeled or modified nucleotides, e.g., fluorescently labeled nucleotides, nucleotide analogs, or the like. Tags can also comprise specific nucleotide sequences, e.g., restriction sites, cis regulatory elements, recognition sites for nucleic acid-binding proteins, sequences capable of forming secondary structures, or the like. The tags herein can optionally comprise one or more ligand, fluorescent label, blocking group, phosphorylated nucleotide, phosphorothioated nucleotide, biotinylated nucleotide, methylated nucleotide, uracil, sequence capable of forming a hairpin structure, oligonucleotide hybridization site, restriction site, DNA promoter, RNA promoter, sample or library identification sequence, and/or cis regulatory sequence.

[0008] Optionally, producing the double stranded fragments with first exonuclease-sensitive strands and second exonuclease-resistant strands can comprise annealing a population of primer pairs, which comprise one primer that comprises an exonuclease-sensitive 5' end and a second primer that comprises an exonuclease-resistant 5' end, to subsequences of the genomic DNA, the cDNA, or the concatamer. The primers can be extended with a polymerase to produce the double stranded fragments with first exonuclease-sensitive strands and second exonuclease-resistant strands. The primers can optionally comprise one or more ligand, fluorescent label, blocking group, phosphorylated nucleotide, phosphorothioated nucleotide, biotinylated nucleotide, methylated nucleotide, uracil, sequence capable of forming a hairpin structure, oligonucleotide hybridization site, restriction site, DNA promoter, RNA promoter, sample or library identification sequence, and/or cis regulatory sequence.

[0009] To produce a population of linear single-stranded nucleic acids, the exonuclease sensitive strand from each double-stranded fragment can be eliminated, e.g., via exonuclease digestion.

[0010] A composition provided by the invention, which is related to the methods described above, comprises an exonuclease and a population of double-stranded nucleic acid fragments that comprise first exonuclease-sensitive strands and second exonuclease-resistant strands. The nucleic acid fragments of the composition comprise overlapping subsequences of a genomic DNA, a cDNA, or a DNA concatamer.

[0011] The invention also provides methods of preparing closed single-stranded nucleic acid loops. One set of methods for preparing closed single-stranded loops includes providing a genomic DNA, a cDNA or a DNA concatamer, and generating double-stranded fragments that comprise first exonuclease-sensitive strands and second exonuclease-resistant strands from the genomic DNA, the cDNA, or the concatamer. The methods include separating first strands of the double-stranded fragments from the second strands to produce single-stranded fragments, and circularizing the resulting single-stranded fragments to produce the closed single-stranded nucleic acid loops.

[0012] The genomic DNA, the cDNA, or the concatamer from which the closed single-stranded nucleic acid loops are produced can optionally be derived from a eukaryote, a prokaryote, an archaea, a virus, or a phage. The population of double-stranded fragments that comprise first exonuclease-sensitive strands and second exonuclease-resistant strands can be optionally produced as described above. Separating the first strands from the second strands of the double-stranded fragments can optionally comprise digesting the first strands of the fragments with an exonuclease to produce the single-stranded fragments.

[0013] As used herein, “circularizing” a nucleic acid refers to forming a loop, e.g., a ring-like structure that does not intersect itself, from a linear nucleic acid fragment, e.g., a single-stranded fragment or double-stranded fragment. A linear nucleic acid fragment can be circularized by joining, e.g., ligating, the ends of the fragment to one another, e.g., to form a “closed loop”, wherein each nucleotide is covalently joined to the nucleotides preceding and following it, or by bringing the ends of the fragment into proximity with one another, e.g., to form a “non-contiguous loop”, wherein at least two nucleotides of the fragment are not joined by a covalent bond.

[0014] For example, in this set of methods, circularizing the single-stranded fragments to produce the closed single-stranded nucleic acid loops can optionally comprise annealing single-stranded nucleic acid splints to the single-stranded fragments. As used herein, a “splint” refers to a short single-stranded nucleic acid oligomer that comprises 3' sequences and 5' sequences that form a perfectly matched homoduplex when hybridized to the sequences at the ends of a longer single-stranded nucleic acid molecule. Following hybridization to the longer single-stranded nucleic acid, the splint functions as a molecular bridge that circularizes the longer molecule by holding its ends in apposition.

[0015] The first ends of the splints can optionally comprise nucleotide sequences complementary to nucleotide sequences at the first ends of the single-stranded fragments, and the second ends of the splints can optionally comprise nucleotide sequences complementary to nucleotide sequences at the second ends of the single-stranded fragments. Thus, annealing the splints to the single-stranded nucleic acids can bring the first and the second ends of each single-stranded fragment into proximity with one another. The first and the second ends of the single-stranded fragments can then be ligated to one another, producing the closed single-stranded nucleic acid loops. The splint can optionally be removed, e.g., via exonuclease digestion.

[0016] In a related aspect, the invention provides a composition that comprises a plurality of single-stranded nucleic acid fragments, which are derived from overlapping subsequences of a genomic DNA, a cDNA, or a DNA concatamer, and a population of single-stranded nucleic acid splints. In the composition, first ends of the splints are annealed to first ends of the fragments and second ends of the splints are annealed to second ends of the fragments, circularizing each fragment by bringing its ends within proximity of one another.

[0017] The invention provides other methods for producing closed single-stranded nucleic acid loops. The methods include providing a genomic DNA, a cDNA, or a DNA concatamer, and generating double-stranded nucleic acid fragments that comprise a gap site from the genomic DNA, cDNA, or concatamer. As described herein, a “gap site” is a nucleotide sequence or nucleotide-associated moiety that permits the introduction of a gap into contiguous nucleic acid strand. For example, a gap site can comprise an unphosphorylated 5' end of a nucleic acid fragment or an enzymatically removable base.

[0018] The fragments that comprise a gap site can be circularized to form double-stranded loops, and non-contiguous strands can be removed from the double-stranded loops to produce the population of closed single-stranded nucleic acid loops. Optionally, the methods include sequencing the circular single-stranded nucleic acids.

[0019] The genomic DNA or the cDNA from which fragments are generated can optionally be derived any of the sources described above. The double-stranded fragments that comprise a gap site can optionally be generated by the strategies described previously.

The gap site can optionally comprise an unphosphorylated 5' end or a sequence recognized by a nicking enzyme, e.g., uracil DNA deglycosylase.

[0020] Circularizing the double-stranded nucleic acid fragments can optionally comprise ligating first ends of the fragments to second ends of the fragments to form a population of double-stranded loops. Removing non-contiguous strands from the double-stranded loops can comprise digesting the non-contiguous strand with, e.g., an exonuclease that specifically cleaves the gap site, to produce the population of single-stranded nucleic acid loops. Optionally, removing non-contiguous strands from the double stranded loops can comprise cleaving sacrificial strands of the loops with, e.g., a nicking enzyme that specifically recognizes the gap site, e.g., uracil DNA deglycosylase, to produce non-contiguous strands, which are then specifically digested with an exonuclease.

[0021] As used herein, "sacrificial strands" refer to the strands in double-stranded nucleic acid molecules, e.g., linear or circular molecules, that are cleaved and removed, e.g., via displacement or enzymatic digestion, to produce single-stranded nucleic acid molecules.

[0022] A related composition provided by the invention comprises a plurality of double-stranded circularized fragments that comprise non-contiguous strands and contiguous strands. The circularized fragments of the composition comprise overlapping sequences of a genomic DNA, a cDNA, or DNA concatamer, optionally derived from a eukaryote.

[0023] Single-stranded circular nucleic acids can also be produced by providing a genomic DNA, a cDNA, or a DNA concatamer, e.g., derived from any of the sources described above, generating a population of double-stranded fragments from the genomic DNA, the cDNA, or the concatamer. The double-stranded fragments generated in these methods comprise ends with a rolling-circle replication (RCR) protein recognition sequence. These methods include copying target strands of the double-stranded fragments to produce a population of single-stranded copies, and circularizing the single-stranded copies to produce the population of closed single-stranded nucleic acid loops.

[0024] As used herein, "copying" refers to the process of replicating a nucleic acid molecule to generate a new nucleic acid that comprises a sequence complementary to that of the original. As used herein, "target strands" refer to the strands in double strand nucleic acid molecules, e.g., linear or circular, that serve as the templates from which new

molecules are synthesized. The nucleic acids that are synthesized are the “copied strands.” The rate of nucleotide misincorporation during the synthesis of the copied strands is assumed to be approximately 0%.

[0025] The double-stranded fragments with ends that comprise an RCR protein recognition sequence can optionally be generated as described in previous methods. The RCR protein recognition sequence can optionally comprise a sequence that is bound by an RCR protein that comprises a histidine-U-histidine-U-U-U amino acid motif, wherein U is a bulky hydrophobic amino acid residue, e.g., cisA.

[0026] Copying the target strands can comprise nicking secondary strands of the double-stranded fragments, displacing the secondary strands, and replicating sequences of the target strands to produce the population of single-stranded copies. The secondary strands of the target strands can be nicked with an RCR protein, e.g., cisA, and displaced a replisome, e.g., a replisome that comprises a single-stranded DNA binding protein (SSB), a helicase, a polymerase, and an RCR protein. Displacing the secondary strands can optionally comprise unwinding the secondary strands from the target strands of the double-stranded fragments with a replisome.

[0027] Replicating the target strand can comprise synthesizing nucleic acid strands that are complementary to the target strands to produce the population of single-stranded copies. Circularizing the single-stranded copies, e.g., the copied strands and the secondary strands, to produce the closed single-stranded nucleic acid loops can comprise ligating the ends of the single-stranded copies with an RCR initiation protein, e.g., cisA.

[0028] In a related aspect, compositions provided by the invention include a plurality of double-stranded nucleic acids, each of which comprise an overlapping subsequence of a genomic DNA, a cDNA, or a DNA concatamer. The nucleic acids of the composition comprise a rolling-circle replication (RCR) protein recognition sequence, e.g., a sequence that is bound by an RCR protein that comprises a histidine-U-histidine-U-U-U amino acid motif, wherein U is a bulky hydrophobic amino acid residue, e.g., cisA. The nucleic acids of the composition can be closed loops or linear. The nucleic acids can optionally comprise genomic DNA or a cDNA that is derived from any of the sources described above.

[0029] Methods of producing a population of single-stranded linear nucleic acids can include providing a genomic DNA, a cDNA, or a DNA concatamer, generating a population of double-stranded fragments from the genomic DNA, the cDNA, or the concatamer. The fragments generated in these methods have ends that comprise a sequence recognized by a nicking enzyme. The methods include circularizing the double-stranded fragments to produce a set of closed double-stranded nucleic acid loops, cleaving sacrificial strands of the closed loops with the nicking enzyme to produce nicked loops, and copying template strands of the nicked loops to produce copied strands that comprise sequences that are recognized by the nicking enzyme. The copied strands are then nicked with a nicking enzyme to produce the single-stranded linear nucleic acids.

[0030] The genomic DNA, cDNA, or concatamer from which the double-stranded fragments are generated can optionally be derived from a eukaryote, a prokaryote, an archaea, a virus or a phage. The double-stranded fragments with ends that comprise a sequence recognized by a nicking enzyme can be produced using previously described strategies.

[0031] Circularizing the double-stranded fragments can optionally comprise ligating the ends the fragments to produce the closed loops. The sacrificial strands of the closed double-stranded loops can be optionally cleaved by a nicking enzyme that recognizes the nicking sequence. Copying the target strands can optionally comprise displacing the sacrificial strands and replicating the template strands, e.g., synthesizing nucleic acid strands that are complementary to the template strands, with, e.g., a strand displacing polymerase, to produce the copied strands.

[0032] In a related aspect, the invention provides compositions that comprise a plurality of closed double-stranded nucleic acid loops, which comprise overlapping fragments of a genomic DNA, a cDNA, or a DNA concatamer. The double-stranded loops of the composition also comprise a sequence that is specifically recognized by a nicking enzyme. The fragments of genomic DNA, the cDNA, or concatamer of the closed double-stranded nucleic acid loops can optionally be derived from a eukaryote.

[0033] Methods for generating closed single-stranded nucleic acid loops can also include providing a genomic DNA, a cDNA, or a DNA concatamer, producing a population of double-stranded nucleic acid fragments from the genomic DNA, the cDNA, or the

concatamer, providing first hairpin oligonucleotides to first ends of the nucleic acid fragments and providing second hairpin oligonucleotides to second ends of the nucleic acid fragments to generate the set of single-stranded circular nucleic acids. The genomic DNA, the cDNA, or the concatamer from which the closed single-stranded nucleic acid loops are produced can be derived from any of the previously described sources. The population of double-stranded nucleic acid fragments can be optionally generated according to strategies elaborated above.

[0034] Providing the double-stranded nucleic acid fragments with hairpin oligonucleotide segments can comprise annealing the first hairpin oligonucleotides to the first ends of the fragments, annealing the second hairpin oligonucleotides to the second ends of the fragments, and ligating the hairpin oligonucleotides to the ends of the fragments to generate the set of closed single-stranded nucleic acid loops that comprise regions of internal complementarity. Providing the hairpins to the fragments can optionally comprise linking 5' strands of the first hairpins to 3' strands of the first ends of the fragments, linking 3' strands the first hairpins to 5' strands of the first ends of the fragments, linking 5' strands of the second hairpins to 3' strands of the second ends of the fragments, and linking 3' strands of the second hairpins to 5' strands of the second ends of the fragments to form a population of closed single-stranded nucleic acid loops.

[0035] The first hairpins can optionally comprise first single-stranded terminal sequences that are complementary to second single-stranded terminal sequences at the first ends of the fragments, and the second hairpins can optionally comprise third single-stranded terminal sequences that are complementary to fourth single stranded terminal sequences at the second ends of the fragments. The first or second hairpins can optionally comprise or encode more or more ligand, fluorescent label, blocking group, phosphorylated nucleotide, phosphorothioated nucleotide, biotinylated nucleotide, methylated nucleotide, uracil, sequence capable of forming hairpin secondary structure, oligonucleotide hybridization site, restriction site, DNA promoter, RNA promoter, sample or library identification sequence, and/or cis regulatory sequence.

[0036] Compositions comprising first hairpin oligonucleotides, second hairpin oligonucleotides, and a plurality of double-stranded nucleic acid fragments that comprise overlapping subsequences of a genomic DNA, a cDNA, or a DNA concatamer, are provided by the invention. The first or second hairpins of the compositions can optionally comprise

or encode any of the moieties described previously. The first hairpins of the compositions can optionally comprise first single-stranded terminal sequences complementary to second single-stranded terminal sequences at the first ends of the fragments, and the second hairpins of the composition can optionally comprise third single-stranded terminal sequences complementary to fourth single stranded terminal sequences at the second ends of the fragments. The fragments of the composition can optionally comprise a genomic DNA, a cDNA, or a DNA concatamer that is derived from, e.g., a eukaryote.

[0037] Linear single-stranded nucleic acids can be produced by methods that include providing a genomic DNA, a cDNA, or a DNA concatamer, generating a set of double-stranded nucleic acid fragments from the genomic DNA, the cDNA, or the concatamer, and nicking sacrificial strands of the double-stranded nucleic acid fragments to produce nicked fragments. In these methods the sacrificial strands are displaced from the fragments, and sequences between nick sites and ends on un-nicked strands that are exposed by the displacement of the sacrificial strands are copied with a strand displacing polymerase to produce the linear single-stranded nucleic acids.

[0038] The genomic DNA, the cDNA, or the concatamer from which the fragments are generated can optionally be derived from a eukaryote. The double-stranded fragments can optionally be generated using any of the procedures described above. Nicking the sacrificial strands of the fragments can optionally comprise cleaving the sacrificial strands with an enzyme to produce the nicked fragments. Displacing the sacrificial strands can optionally comprise removing the sacrificial strands from the un-nicked strands with a strand-displacing polymerase, e.g., PolI, BstI, Phi29, or Phi29-like polymerases, e.g., described in United States Patent Application No. 11/645,223, entitled POLYMERASES FOR NUCLEOTIDE ANALOGUE INCORPORATION. The exposed sequences can be copied by replicating sequences between the nick sites and the ends on the un-nicked strands with the strand-displacing polymerase. Replicating can optionally comprise synthesizing nucleic acid strands that are complementary to the un-nicked strands, to produce the linear single-stranded nucleic acids.

BRIEF DESCRIPTION OF THE DRAWINGS

[0039] **Figure 1** illustrates methods and compositions for generating single-stranded nucleic acid fragments from double-stranded fragments.

- [0040] **Figure 2** illustrates methods and compositions for generating closed single-stranded nucleic acid loops from double-stranded fragments.
- [0041] **Figure 3** illustrates methods and compositions for producing double-stranded templates from which closed single-stranded nucleic acid loops are copied.
- [0042] **Figure 4** illustrates methods and compositions for generating closed single-stranded loops from double-stranded fragments.
- [0043] **Figure 5** illustrates methods and compositions for producing closed double-stranded nucleic acid loops from which linear single-stranded nucleic acids are copied.
- [0044] **Figure 6** illustrates methods and compositions for generating single-stranded nucleic acid loops from double-stranded fragments and hairpin oligonucleotides.
- [0045] **Figure 7** illustrates methods and compositions for generating linear single-stranded nucleic acids from double-stranded fragments using strand-displacing polymerases.

DETAILED DESCRIPTION

- [0046] Collecting reliable sequence data using high-throughput sequencing technologies depends in part on the availability of methods for the rapid and efficient production of high-quality nucleic acid templates. The present invention provides methods and compositions that can be useful in supplying templates to such high throughput DNA sequencing systems. The methods circumvent the need for costly, labor-intensive cloning and cell culture methods, which can limit sample production from matching the capacities of modern sequencing systems (such systems are reviewed in, e.g., Chan, et al. (2005) "Advances in Sequencing Technology" *Mutation Research* **573**: 13-40, and described in Levene et al. (2003) "Zero Mode Waveguides for Single Molecule Analysis at High Concentrations," *Science* **299**: 682-686). Accordingly, the reduction of sequencing costs from current levels is a benefit of the improved methods provided herein. The methods can be scaled to accommodate template production for a variety of sequencing applications, e.g., sequencing individuals' genomes, gene expression profiling (Spinella, et al. (1999) "Tandem arrayed ligation of expressed sequence tags (TALEST): a new method for generating global gene expression profiles." *Nucleic Acids Res* **27**: e22, Velculescu, et al. (1995) "Serial analysis of gene expression." *Science* **270**: 484-487) and others.
- [0047] The methods and compositions provided by the invention can be used to produce either linear single-stranded nucleic acids or closed single-stranded nucleic acid

loops. Single-stranded templates can be preferable to double-stranded nucleic acids for sequencing because the thermodynamic stability of, e.g., homoduplex DNA, can promote the rapid reannealing of, e.g., double-stranded DNA that has been denatured in preparation for a sequencing reaction. Such reannealing can reduce the efficiency with which a primer can hybridize to a template. Furthermore, the reannealing of double-stranded DNA can also impede polymerase-catalyzed extension of a sequencing reaction, decreasing the speed of the reaction and the accuracy of the results .

[0048] The detailed description is organized to first elaborate the various methods and compositions that are provided by the invention for the preparation of templates for high-speed, high-capacity sequencing platforms. Broadly applicable molecular biological techniques that can be used to perform any of the sets of methods will be described thereafter.

METHODS AND COMPOSITIONS FOR GENERATING SINGLE-STRANDED NUCLEIC ACID FRAGMENTS FROM DOUBLE-STRANDED FRAGMENTS

[0049] Because of the thermodynamic stability of homoduplex DNA, double-stranded DNA that has been denatured, e.g., in preparation for a sequencing reaction, can rapidly reanneal. Such reannealing can reduce the efficiency with which a sequencing primer can hybridize to a template and impede the polymerase-catalyzed extension of a sequencing reaction. Using single-stranded nucleic acid templates in sequencing reactions obviates these complications. The invention provides a set of methods wherein single-stranded linear nucleic acids are produced from double-stranded nucleic acid fragments.

[0050] An outline of these methods is schematically illustrated in **Figure 1**. In a first aspect, genomic DNA, cDNA or a DNA concatamer **100**, e.g., derived from a prokaryote, archaea, or, more typically a eukaryote, e.g., a human or other mammalian patient, is provided to the methods.

[0051] Population of double-stranded nucleic acid fragments **105** is generated from the genomic DNA, cDNA, or concatamer **100**. The double-stranded fragments comprise first exonuclease-sensitive strands **112** and second exonuclease-resistant strands **110**.

[0052] The fragments can be generated in any of a number of ways. For example, the genomic DNA, cDNA, or concatamer can be cleaved, e.g., via enzymatic digestion, sonication, mechanical shearing, electrochemical cleavage, and/or nebulization, to produce the fragments, and exonuclease-sensitive tags can be attached to the 5' ends of the first strands of the fragments and exonuclease-resistant tags can be attached to the 5' ends of the

second strands of the fragments, e.g., with a ligase, via primer extension, via chemical linkage, and the like. Optionally, the double-stranded fragments can be produced by annealing a population of nested primer pairs, e.g., which pairs comprise one primer that comprises an exonuclease-sensitive 5' end and a second primer that comprises an exonuclease-resistant 5' end, to the genomic DNA, the cDNA or the concatamer, and extending the primers with a polymerase.

[0053] The tags and/or primers used in generating the double-stranded fragments that comprise first exonuclease-sensitive strands and second exonuclease-resistant strands can comprise one or more ligand, fluorescent label, blocking group, phosphorylated nucleotide, phosphorothioated nucleotide, biotinylated nucleotide, methylated nucleotide, uracil, a sequence capable of forming hairpin secondary structure, oligonucleotide hybridization site, restriction site, DNA promoter, RNA promoter, sample or library identification sequence, cis regulatory sequence, and/or the like.

[0054] Exonuclease-sensitive strands **112** are eliminated from each of the double-stranded fragments, e.g., via digestion with an exonuclease **115**, thereby producing single-stranded linear nucleic acid fragments **120** that can subsequently be sequenced in high-throughput sequencing systems. The invention provides related composition **117** that includes an exonuclease and overlapping double-stranded fragments of a genomic DNA, cDNA, or concatamer that comprise first exonuclease-sensitive strands and second exonuclease-resistant strands.

METHODS AND COMPOSITIONS FOR GENERATING SINGLE-STRANDED NUCLEIC ACID LOOPS FROM DOUBLE-STRANDED FRAGMENTS

[0055] **Figure 2** depicts a schematic for a set of methods provided by the invention in which closed single-stranded nucleic acid loops are generated from double-stranded fragments. As described previously, single-stranded nucleic acid templates are preferable to double-stranded nucleic acids for sequencing because the rapid reannealing of, e.g., double-stranded DNA that has been denatured in preparation for a sequencing reaction, can reduce the efficiency with which a primer can hybridize to a template and impede polymerase-catalyzed extension of a sequencing reaction.

[0056] In these methods, a population of double-stranded nucleic acid fragments **210** comprising first exonuclease-sensitive strands **202** and second exonuclease-resistant strands **201** are generated from a genomic DNA, cDNA, or a DNA concatamer, e.g., of tandem ligated expressed sequence tags (ESTs), **200**, using any of the strategies described

above. The genomic DNA, the cDNA or the concatamer can optionally be derived from a prokaryote, a eukaryote, an archaea, or a virus.

[0057] The first and second strands of the double-stranded fragments are separated from one another, e.g., via digestion with exonuclease 235 to produce a set of single-stranded nucleic acid fragments 240. The single-stranded fragments are then circularized to produce a population of single-stranded nucleic acid loops, e.g., for sequencing.

[0058] The single-stranded fragments can be circularized using any number of strategies. However, in a preferred embodiment, the fragments are annealed to nucleic acid splints 245. The splint can comprise first ends that comprise sequences 220 that are complementary to sequences 215 at the first ends of the single-stranded nucleic acid fragments and second ends that comprise sequences 230 that are complementary to sequences 225 at the second ends of the single-stranded fragments. Annealing the fragments to the splints brings the two ends of each fragment within close proximity of one another (composition 250), which increases the efficiency of ligation, wherein the two ends are covalently linked. The splints can be removed, e.g., via digestion with exonuclease 235 to produce single-stranded nucleic acid loops 255.

[0059] Composition 250, which is related to these methods, is also provided by the invention. The composition comprises a population of single-stranded nucleic acid fragments 240, which comprise overlapping subsequences of a genomic DNA, a cDNA, or a DNA concatamer, and a population of single-stranded nucleic acid splints 245. In this composition, the first ends of the splints are annealed to first ends of the fragments and second ends of the splints are annealed to second ends of the fragments, such that the first and second ends of each fragment are brought within proximity of one another. The genomic DNA, the cDNA, or the concatamer from which the single-stranded fragments are derived can optionally be, e.g., eukaryotic, prokaryotic, or archaeal.

[0060] A second set of methods for generating single-stranded nucleic acid loops from double-stranded fragments is illustrated in **Figure 3**. In these methods, a population of double-stranded nucleic acid fragments comprising a gap site is optionally generated from, e.g., a prokaryotic, eukaryotic, archaeal, or viral genomic DNA, cDNA, or DNA concatamer. As described herein, a “gap site” is a nucleotide sequence or nucleotide-associated moiety that permits the introduction of a gap into contiguous nucleic acid strand. In one embodiment, double stranded fragments 350, comprising first phosphorylated strands 335 and second unphosphorylated strands, are produced from genomic DNA, cDNA, or

concatamer **300**. In this embodiment, the unphosphorylated strand comprises the gap site. Optionally, set of fragments **305**, which comprise ends that encode sequence **306**, which is recognized by nicking enzyme **315**, can be generated from genomic DNA, cDNA, or concatamer **300**. In this alternative embodiment, sequence **306** comprises the gap site. Either population of fragments, e.g., **305** or **350**, can be produced using any of the strategies described previously.

[0061] Fragments **350**, which comprise first phosphorylated strands and second unphosphorylated strands, can be circularized, e.g., with a ligase, to produce population of double-stranded loops that comprise single non-contiguous strands **325**. Optionally, fragments **305** can be circularized, e.g., with a ligase, to produce closed double-stranded closed loops **311**. The double-stranded loops can then be cleaved with nicking enzyme **315**, e.g., uracil deglycosylase, at sequence **306** to produce the population of double-stranded loops that comprise single non-contiguous strands **325**. The non-contiguous strands can then be removed, optionally via exonuclease **320** digestion, to produce single-stranded nucleic acid loops **326** that can be used, e.g., in high-throughput sequencing systems.

[0062] Related composition **325** is provided by the invention. The composition comprises a set of double-stranded circularized overlapping fragments of a genomic DNA, a cDNA, or DNA concatamer. The circularized fragments of the composition each comprise non-contiguous strands and contiguous strands.

METHODS AND COMPOSITIONS FOR PRODUCING DOUBLE-STRANDED LINEAR TEMPLATES FROM WHICH CLOSED SINGLE-STRANDED NUCLEIC ACID LOOPS ARE COPIED

[0063] The invention provides methods and compositions for generating a population of double-stranded fragments and for copying sequences on the fragments to produce closed single-stranded loops, e.g., for sequencing. As described above, using single-stranded templates in, e.g., high-throughput sequencing systems, can be preferable to double-stranded nucleic acids for sequencing because the thermodynamic stability of, e.g., homoduplex DNA, can promote the rapid reannealing of, e.g., double-stranded DNA that has been denatured in preparation for a sequencing reaction. Such reannealing can reduce the efficiency with which a primer can hybridize to a template and impede polymerase-catalyzed extension of a sequencing reaction.

[0064] In one embodiment, illustrated in **Figure 4**, double-stranded fragments **410** that comprise a rolling-circle replication (RCR) protein recognition sequence **405** are

generated from genomic DNA, cDNA, or a DNA concatamer **400**. The strategies for generating fragments that are elaborated above and apply to this embodiment, as well.

[0065] Composition **410** is provided by the invention. The nucleic acids of the composition can optionally be closed circular loops or linear. Optionally, the RCR protein recognition sequence can comprise a sequence recognized by an RCR protein that comprises a histidine-U-histidine-U-U-U motif, wherein U is a bulky hydrophobic amino acid. Optionally, the RCR protein can be a cisA protein.

[0066] Target strands **420** of each of the double-stranded fragments are repeatedly copied to produce single-stranded loops **435**. Copying target strands can comprise nicking secondary strands **416** of each double-stranded fragment, e.g., with RCR protein **415**, displacing the secondary strands, and replicating sequences of the target strands to produce a population of single-stranded copies. The RCR protein that nicks the secondary strands can optionally be an RCR protein that comprises a histidine-U-histidine-U-U-U amino acid motif, wherein U is a bulky hydrophobic amino acid, e.g., cisA. The secondary strands can optionally be displaced by replisome **425**, e.g., a replisome comprising a single-stranded DNA-binding protein (SSB), a helicase, a polymerase, and an RCR protein. The replisome can optionally replicate the sequences of the target strand, e.g., by synthesizing nucleic acid strands that are complementary to the target strands to produce the population of single-stranded copies **430**. Single-stranded copies **430** and secondary strands **416** can optionally be circularized, e.g., with an RCR protein, e.g., a cisA, which can optionally ligate the strands of each single-stranded copy together to produce closed single-stranded nucleic acid loops **435**.

METHODS AND COMPOSITIONS FOR PRODUCING CLOSED DOUBLE-STRANDED LOOPS FROM WHICH LINEAR SINGLE-STRANDED NUCLEIC ACIDS ARE COPIED

[0067] **Figure 5** illustrates a set of methods for producing double-stranded circular templates from which linear single-stranded nucleic acid copies, e.g., single-stranded nucleic acid templates for high-throughput sequencing systems, are synthesized. Single-stranded nucleic acid templates can be preferable to double-stranded nucleic acids for sequencing for the reasons that have been described above.

[0068] In this set of methods, double-stranded nucleic acid fragments **510** are produced from genomic DNA, cDNA, or a DNA concatamer **500**. The double-stranded

fragments have ends that comprise sequence **520** that is recognized by nicking enzyme **543**. The fragments can be generated using any of the previously described strategies.

[0069] The double-stranded fragments are circularized to produce a set of closed double-stranded loops **530**. Sacrificial strands **544** of each loop can be cleaved, optionally by nicking enzyme **543**, and displaced, e.g., by strand-displacing polymerase **540**, e.g., any of the strand polymerases described elsewhere herein. The strand displacing polymerase can also replicate the template strands, e.g., by synthesizing nucleic acid strands that are complementary to the template strands, thereby producing copied strands **550**. The copied strands each comprise the sequences that are recognized by nicking enzyme **543**, and, accordingly, are cleaved by nicking enzyme **543** to produce copied strands **545**.

[0070] Composition **530**, which is related to the above methods, is provided by the invention. Composition **530** comprises closed double-stranded nucleic acid loops that comprise overlapping subsequences of a genomic DNA, a cDNA, or a DNA concatamer, e.g., derived from a eukaryote. The loops include sequence **520** that is recognized by nicking enzyme **543**.

METHODS AND COMPOSITIONS FOR GENERATING SINGLE-STRANDED NUCLEIC ACID LOOPS FROM DOUBLE-STRANDED FRAGMENTS AND HAIRPIN OLIGONUCLEOTIDES

[0071] Methods and related compositions for generating single-stranded nucleic acid loops from double-stranded nucleic acid fragments and hairpin oligonucleotides are depicted in **Figure 6**. For the reasons described previously, single-stranded templates can be preferable to double-stranded nucleic acids in sequencing reactions. In this set of methods, double-stranded fragments **610** are generated from a genomic DNA, cDNA, or a DNA concatamer **600**. Procedures for generating double-stranded nucleic acid fragments **610** from genomic DNA, cDNA, or concatamer **600** are described above and can be used in this embodiment, as well. To produce the single-stranded nucleic acid loops, a population of first hairpin oligonucleotides **620** can be provided to one end of each double-stranded fragment, and a population of second hairpin oligonucleotides **625** can be provided to the other end of each fragment, as shown in composition **635**. The closed single-stranded nucleic acid loops produced by this method will comprise regions of internal complementarity.

[0072] Composition **630** is provided by the invention. Composition **635** comprises a population of double stranded fragments **610**, which comprise overlapping sequences of

genomic DNA, cDNA, or DNA concatamer **600**, which can optionally be derived from a eukaryote. Composition **635** includes first hairpin oligonucleotides **620**, and second hairpin oligonucleotides **625**. Hairpin oligonucleotides **620** and/or **625** of composition **630** can comprise or encode any of the moieties or sequences that have been described previously. Hairpins **620** of the composition **630** can optionally comprise single-stranded terminal sequences that are complementary to second single-stranded terminal sequences at first ends of fragments **610**, and hairpins **625** of composition **630** can optionally comprise third single-stranded terminal sequences that are complementary to fourth single-stranded terminal sequences at the second ends of fragments **610**.

[0073] Providing hairpin oligonucleotides to the ends of the fragments can optionally comprise annealing the first hairpins to first ends of the fragments, annealing second hairpins to second ends of the fragments, and ligating the hairpin oligonucleotides to the ends of the fragments to produce population of single-stranded nucleic acid loops **635**. Optionally, providing the hairpins to the ends of the fragments can comprise linking 5' strands of the first hairpins to 3' strands at the first ends of the fragments, linking 3' strands of the first hairpins to 5' strands at the first ends of the fragments, linking 5' strands of the second hairpins to 3' strands at the second ends of the fragments, and linking 3' strands of the second hairpins to 5' strands at the second ends of the fragments.

METHODS AND COMPOSITIONS FOR GENERATING SINGLE-STRANDED NUCLEIC ACIDS FROM DOUBLE-STRANDED FRAGMENTS USING STRAND-DISPLACING POLYMERASES

[0074] **Figure 7** depicts a schematic for a set of methods of generating single-stranded nucleic acids from double-stranded fragments using strand-displacing polymerases. Sequencing reactions comprising single-stranded templates can proceed more efficiently for the reasons described elsewhere herein. In this set of methods, genomic DNA, cDNA, or a DNA concatamer **700** is provided, and double-stranded fragments **705** that comprise overlapping subsequences of the genomic DNA, cDNA, or concatamer can be produced using any of the strategies that have been previously described. The methods include nicking sacrificial strands **703** of double-stranded fragments **705** to produce population of nicked fragments **720**, e.g., fragments that comprise one non-contiguous strand. The sacrificial strands of nicked fragments **720** can be displaced, e.g., by strand displacing polymerase **725**, e.g., any of the strand displacing polymerases described elsewhere herein. The sequences on contiguous strands **704** of nicked fragments **720** that are exposed by the

displacement of sacrificial strands **703** can be copied, e.g., by strand displacing polymerase **725**. Strand-displacing polymerase **725** can replicate the sequences between the nick sites and ends of contiguous strands **704** that are exposed by the displacement of sacrificial strands **703**, e.g., by synthesizing complementary copies of the contiguous strands, to produce population of single-stranded nucleic acids **730**. Composition **730** can be provided, e.g., to high-throughput sequencing systems.

PREPARING GENOMIC DNA

[0075] As described above, the single-stranded nucleic acids, e.g., linear or circular nucleic acids, that are provided by the methods described herein, e.g., for use in single molecule sequencing reactions, can be derived from a genomic DNA. Genomic DNA can be prepared from any source by three steps: cell lysis, deproteinization and recovery of DNA. These steps are adapted to the demands of the application, the requested yield, purity and molecular weight of the DNA, and the amount and history of the source. Further details regarding the isolation of genomic DNA can be found in Berger and Kimmel, Guide to Molecular Cloning Techniques, Methods in Enzymology volume 152 Academic Press, Inc., San Diego, CA (Berger); Sambrook et al., Molecular Cloning - A Laboratory Manual (3rd Ed.), Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 2008 ("Sambrook"); Current Protocols in Molecular Biology, F.M. Ausubel et al., eds., Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc ("Ausubel"); Kaufman et al. (2003) Handbook of Molecular and Cellular Methods in Biology and Medicine Second Edition Ceske (ed) CRC Press (Kaufman); and The Nucleic Acid Protocols Handbook Ralph Rapley (ed) (2000) Cold Spring Harbor, Humana Press Inc (Rapley). In addition, many kits are commercially available for the purification of genomic DNA from cells, including WizardTM Genomic DNA Purification Kit, available from Promega; Aqua PureTM Genomic DNA Isolation Kit, available from BioRad; Easy-DNATM Kit, available from Invitrogen; and DnEasyTM Tissue Kit, which is available from Qiagen.

PREPARING cDNA

[0076] The single-stranded linear or circular nucleic acids that can be prepared by the methods described herein, e.g., for use with high-throughput sequencing systems can also be derived from a cDNA, e.g. cDNAs prepared from mRNA obtained from, e.g., a eukaryotic subject or a specific tissue derived from a eukaryotic subject. Data obtained from sequencing the nucleic acid templates derived from a cDNA library, e.g., using a high-

throughput sequencing system, can be useful in identifying, e.g., novel splice variants of a gene of interest or in comparing the differential expression of, e.g., splice isoforms of a gene of interest, e.g., between different tissue types, between different treatments to the same tissue type or between different developmental stages of the same tissue type.

[0077] mRNA can typically be isolated from almost any source using protocols and methods described in, e.g., Sambrook and Ausubel. The yield and quality of the isolated mRNA can depend on, e.g., how a tissue is stored prior to RNA extraction, the means by which the tissue is disrupted during RNA extraction, or on the type of tissue from which the RNA is extracted. RNA isolation protocols can be optimized accordingly. Many mRNA isolation kits are commercially available, e.g., the mRNA-ONLY™ Prokaryotic mRNA Isolation Kit and the mRNA-ONLY™ Eukaryotic mRNA Isolation Kit (Epicentre Biotechnologies), the FastTrack 2.0 mRNA Isolation Kit (Invitrogen), and the Easy-mRNA Kit (BioChain). In addition, mRNA from various sources, e.g., bovine, mouse, and human, and tissues, e.g. brain, blood, and heart, is commercially available from, e.g., BioChain (Hayward, CA), Ambion (Austin, TX), and Clontech (Mountainview, CA).

[0078] Once the purified mRNA is recovered, reverse transcriptase is used to generate cDNAs from the mRNA templates. Methods and protocols for the production of cDNA from mRNAs, e.g., harvested from prokaryotes as well as eukaryotes, are elaborated in *cDNA Library Protocols*, I. G. Cowell, et al., eds., Humana Press, New Jersey, 1997, Sambrook and Ausubel. In addition, many kits are commercially available for the preparation of cDNA, including the Cells-to-cDNA™ II Kit (Ambion), the RETROscript™ Kit (Ambion), the CloneMiner™ cDNA Library Construction Kit (Invitrogen), and the Universal RiboClone® cDNA Synthesis System (Promega). Many companies, e.g., Agencourt Bioscience and Clontech, offer cDNA synthesis services.

PREPARING CONCATAMERS OF TANDEM SHORT EXPRESSED SEQUENCE TAGS (ESTs)

[0079] The nucleic acids that are distributed to low copy number reaction volumes in preparation for single molecule sequencing can also be derived from concatenated short expressed sequence tags (ESTs) prepared from, e.g., a cDNA library. A short sequence tag, e.g., 10-14bp, can contain sufficient information to uniquely identify a transcript, provided that the tag is obtained from a unique sequence within the transcript. Short sequence tags can be linked together to form long serial molecules that can be prepared, e.g., using

the methods described herein, for sequencing, e.g., using a high-throughput sequencing system, e.g., a ZMW. Quantitation of the number of times a particular tag is observed provides the expression level of the corresponding transcript. Thus, sequencing the nucleic acid templates derived from concatenated short ESTs, e.g., using a high-throughput system such as a ZMW, can be useful in analyzing global gene expression patterns of, e.g., a tissue at different developmental stages, tissues in different organs from a common genotype, common tissues of different genotypes, common tissues that have been exposed to different treatments, and the like. In addition, sequencing templates derived from concatamers of short ESTs can eliminate the need for a practitioner to carry out laborious and time-consuming *in vivo* cloning and cell culturing techniques that are common for other EST-based systems for the analysis of global gene expression, e.g. SAGE (Velculescu, et al. (1995) "Serial analysis of gene expression." *Science* **270**: 484-487) and TALEST (Spinella, et al (1999) "Tandem arrays ligation of expressed sequence tags (TALEST): a new method for generating global gene expression profiles." *Nucl Acid Res* **27**: e22).

[0080] Preparing concatenated ESTs can comprise preparing a cDNA library, e.g., as described above in the section entitled "PREPARING cDNA". Typically, the prepared cDNA can then be digested with a restriction enzyme that would be expected to cleave most transcripts at least once, e.g., a restriction enzyme with a 4-base pair recognition site. The 3'-most cDNA fragments are then captured and ligated to adapter molecules that each contain a type-II restriction site, e.g., BsgI, and a second restriction site. Digestion of the adapter-ligated cDNAs, e.g., with BsgI, produces DNA fragments that consist of the adapter itself and an additional 10-12 nucleotides of unknown cDNA sequence separated from the adapter by the restriction site originally used to digest the cDNA. The fragments can then be ligated to a second adapter containing a second restriction site at one end and degenerate overhangs, e.g., which render the second adapter compatible with all possible cDNA sequences, e.g., produced by the BsgI digestion, at the other. The resulting double-tagged DNA molecules can be digested with enzymes that recognize the restriction sites on the adapters and ligated together to form concatamers that can then be prepared, e.g., using the methods described herein, for sequencing, e.g., using a high-throughput system. Additional information and methods describing the preparation of concatamers comprising short ESTs can be found in, e.g., Velculescu, et al. (1995) "Serial analysis of gene expression." *Science* **270**: 484-487; Spinella, et al (1999) "Tandem arrays ligation of expressed sequence tags (TALEST): a new method for generating global gene expression profiles." *Nucl Acid Res*

27: e22; WIPO Patent Application Number WO/2004/024953; and Unneberg, et al. (2003) "Transcript identification by analysis of short sequence tags – influence of tag length, restriction site, and transcript database." *Nucl Acids Res* **31**: 2217-2226.

CLEAVING NUCLEIC ACID FRAGMENTS TO PRODUCE FRAGMENTS

[0081] In some embodiments of the invention described herein, nucleic acid fragments are generated from a genomic DNA or a cDNA. There exist a plethora of ways of generating nucleic acid fragments from a genomic DNA, a cDNA, or a DNA concatamer. These include, but are not limited to, mechanical methods, such as sonication, mechanical shearing, nebulization, hydroshearing, and the like; enzymatic methods, such as exonuclease digestion, restriction endonuclease digestion, and the like; and electrochemical cleavage. These methods are further explicated in Sambrook and Ausubel.

COPYING NUCLEIC ACID SEQUENCES

[0082] Certain embodiments of the invention utilize optional steps in which populations of double-stranded nucleic acid fragments are produced by copying subsequences, e.g. overlapping subsequences, of a genomic DNA, a cDNA, or a DNA concatamer. A variety of nucleic acid amplification and/or copying methods are known in the art and can be implemented to perform these steps.

[0083] The most widely used *in vitro* technique among these methods is polymerase chain reaction (PCR), which requires the addition of a template of interest, e.g., a DNA comprising the sequence that is to be amplified, nucleotides, oligonucleotide primers, buffer, and an appropriate polymerase to an amplification reaction mix. In PCR, the primers anneal to complementary sequences on denatured template DNA and are extended with a thermostable DNA polymerase to copy the sequence of interest. As a result, a nucleic acid that comprises a sequence complementary to that of the template strand is synthesized. Repeated cycles of PCR can generate a myriad of copies. Primers ideally comprise sequences that are complementary to the template. However, they can also comprise sequences that are not complementary, but which comprise e.g., restriction sites, cis regulatory sites, oligonucleotide hybridization sites, protein binding sites, DNA promoters, RNA promoters, sample or library identification sequences, and the like. Primers can comprise modified nucleotides, such as methylated, biotinylated, or fluorinated nucleotides; and nucleotide analogs, such as dye-labeled nucleotides, non-hydrolysable nucleotides, and nucleotides comprising heavy atoms. Primers can be custom synthesized by commercial suppliers as described below in NUCLEIC ACID TAGS. As described

below in NUCLEIC ACID TAGS, PCR can be a useful means by which to attach tags to fragments. Further details regarding PCR and its uses are described in *PCR Protocols A Guide to Methods and Applications* (Innis et al. eds) Academic Press Inc. San Diego, CA (1990) (Innis); Chen et al. (ed) *PCR Cloning Protocols, Second Edition* (Methods in Molecular Biology, volume 192) Humana Press; and in Viljoen et al. (2005) *Molecular Diagnostic PCR Handbook* Springer, ISBN 1402034032.

[0084] Additional methods that can be used to amplify, or copy, nucleic acids include strand displacement amplification (SDA), multiple-displacement amplification (MDA), rolling circle replication (RCR). Some methods use RCR to copy single-stranded nucleic acids, e.g., which will be used as templates in sequencing reactions, from double-stranded templates. In RCR, DNA replication is initiated by an initiator protein, e.g., cis A, which nicks one strand of the double-stranded, circular DNA molecule at a specific nucleotide sequence called the double-strand origin, or DSO. The initiator protein remains bound to the 5' phosphate end of the nicked strand, and the free 3' hydroxyl end is released to serve as a primer for DNA synthesis by DNA polymerase III. Using the un-nicked strand as a template, replication proceeds around the circular DNA molecule, displacing the nicked strand as single-stranded DNA. Displacement of the nicked strand is carried out by a replisome, e.g., a multiprotein complex that comprises a single-stranded DNA binding protein (SSB), a helicase, a polymerase, and an RCR initiation protein, e.g., cisA.

[0085] Further details regarding Rolling Circle Amplification can be found in Demidov, et al. (2002) "Rolling-circle amplification in DNA diagnostics: the power of simplicity," *Expert Rev Mol Diagn* **2**: 89-94; Demidov and Broude (eds) (2005) *DNA Amplification: Current Technologies and Applications*. Horizon Bioscience, Wymondham, UK; and Bakht et al. (2005) "Ligation-mediated rolling-circle amplification-based approaches to single nucleotide polymorphism detection" *Expert Rev Mol Diagn* **5**: 111-116.

NUCLEIC ACID TAGS

[0086] In some methods provided by this invention, tags, e.g., tags comprising phosphorylated 5' ends, can be attached to ends of nucleic acid fragments in order to protect the fragments from, e.g., degradation by an exonuclease. The invention also provides methods wherein tags comprising specific sequences can be attached to the ends of fragments so that the fragments can be, e.g., recognized by and nicked by, an e.g., an RCR protein. In other methods, tags comprising, e.g., complementary single stranded terminal

sequences, are attached to the ends of fragments to, e.g., promote more efficient circularization of the fragments. Nucleic acid tags can comprise any of a plethora of ligands, such as high-affinity DNA-binding proteins; modified nucleotides, such as methylated, biotinylated, or fluorinated nucleotides; and nucleotide analogs, such as dye-labeled nucleotides, non-hydrolysable nucleotides, or nucleotides comprising heavy atoms. Such reagents are widely available from a variety of vendors, including Perkin Elmer, Jena Bioscience and Sigma-Aldrich. Nucleic acid tags can also include oligonucleotides that comprise specific sequences, such as restriction sites, cis regulatory sites, oligonucleotide hybridization sites, protein binding sites, and the like. Such oligonucleotide tags can be custom synthesized by commercial suppliers such as Operon (Huntsville, AL), IDT (Coralville, IA) and Bioneer (Alameda, CA). The methods that can be used to join tags to nucleic acids of interest include chemical linkage, ligation (described below in LIGATION OF NUCLEIC ACIDS), and extension of a primer by a polymerase (described above in PRODUCING NUCLEIC ACID FRAGMENTS). Further details regarding nucleic acid tags and the methods by which they are attached to nucleic acids of interest are elaborated in Sambrook and Ausubel.

LIGATION OF NUCLEIC ACIDS

[0087] In some embodiments of the invention, ligation steps can be performed to, e.g., attach tags to the ends of nucleic acid fragments, or to, e.g., attach the first and second ends of fragments to each other to generate closed nucleic acid loops that can be used, e.g., in sequencing reactions, e.g., in a high-throughput sequencing system.

[0088] Ligation is a method by which DNAs, RNAs, or DNAs and RNAs are joined with a covalent bond. Ligations are performed by incubating the nucleic acid fragments to be joined in the presence of buffer, rATP, and a ligase enzyme capable of catalyzing the ligation reaction of interest. Further details regarding these techniques can be found in Sambrook and Ausubel. Furthermore, a plethora of enzymes, each capable of catalyzing a unique type of ligation reaction, are commercially available. For example, CircLigase™, from Epicentre Biotechnologies, is capable of catalyzing the intramolecular ligation of single-stranded DNA fragments; T4 RNA ligase 1, available from New England Biosciences, is capable of ligating single-stranded RNAs to other single-stranded RNAs and single-stranded RNAs to single-stranded DNAs; and T4 DNA ligase, available from many

commercial sources, is capable of catalyzing both inter- and intramolecular ligation of double-stranded DNAs.

[0089] While the foregoing invention has been described in some detail for purposes of clarity and understanding, it will be clear to one skilled in the art from a reading of this disclosure that various changes in form and detail can be made without departing from the true scope of the invention. For example, all the techniques and apparatus described above can be used in various combinations. All publications, patents, patent applications, and/or other documents cited in this application are incorporated by reference in their entirety for all purposes to the same extent as if each individual publication, patent, patent application, and/or other document were individually indicated to be incorporated by reference for all purposes.

CLAIMS

WHAT IS CLAIMED IS:

1. A method of producing a population of linear single-stranded nucleic acids, the method comprising:
 - providing a genomic DNA, a cDNA, or a DNA concatamer,
 - generating double-stranded fragments from the genomic DNA, the cDNA, or the concatamer, wherein first strands of the fragments are exonuclease-sensitive and second strands of the fragments are exonuclease-resistant; and,
 - eliminating the first strands from the double-stranded fragments to produce the population of linear single-stranded nucleic acids.
2. The method of claim 1, wherein the genomic DNA, the cDNA, or the concatamer is derived from a eukaryote.
3. The method of claim 1, wherein generating the double-stranded fragments that comprise the first exonuclease-sensitive strands and the second exonuclease-resistant strands comprises:
 - cleaving the genomic DNA, the cDNA, or the concatamer to produce linear double-stranded nucleic acids, and;
 - ligating exonuclease-sensitive tags to the 5' ends of the first strands of the linear double-stranded nucleic acids and ligating exonuclease-resistant tags to the 5' ends of the second strands of the linear double-stranded nucleic acids, thereby producing the double-stranded fragments with the first exonuclease-sensitive strands and the second exonuclease-resistant strands.
4. The method of claim 3, wherein cleaving the genomic DNA, the cDNA, or the concatamer comprises one or more of: enzymatic digestion, sonication, mechanical shearing, electrochemical cleavage, or nebulization of the genomic DNA, the cDNA, or the concatamer to produce the linear double-stranded nucleic acids.
5. The method of claim 3, wherein the tags comprise one or more moieties selected from: a ligand, a fluorescent label, a blocking group, a phosphorylated nucleotide, a phosphorothioated nucleotide, a biotinylated nucleotide, a methylated nucleotide, a uracil, a sequence capable of forming hairpin secondary structure, an oligonucleotide hybridization site, a restriction site, a DNA promoter, an RNA promoter, a sample or library identification sequence, and a cis regulatory sequence.

6. The method of claim 1, wherein generating the double-stranded fragments that comprise the first exonuclease-sensitive strands and the second exonuclease-resistant strands comprises:

annealing a population of primer pairs, which primer pairs comprise first primers with exonuclease-sensitive 5' ends and second primers with exonuclease-resistant 5' ends, to subsequences of the genomic DNA, the cDNA, or the concatamer; and,

extending the primers with a polymerase to produce the double stranded fragments with the first exonuclease-sensitive strands and the second exonuclease-resistant strands.

7. The method of claim 6, wherein the primers comprise one or more moieties selected from: a ligand, a fluorescent label, a blocking group, a phosphorylated nucleotide, a phosphorothioated nucleotide, a biotinylated nucleotide, a methylated nucleotide, a uracil, a sequence capable of forming hairpin secondary structure, an oligonucleotide hybridization site, a restriction site, a DNA promoter, an RNA promoter, a sample or library identification sequence, and a cis regulatory sequence.

8. The method of claim 1, wherein the first strands are eliminated by digestion with an exonuclease, thereby producing the population of linear single-stranded nucleic acids.

9. A composition comprising:

a population of double-stranded nucleic acid fragments, which fragments comprise overlapping subsequences of a genomic DNA, a cDNA, or a DNA concatamer, wherein the fragments comprise first exonuclease-sensitive strands and second exonuclease-resistant strands; and,

an exonuclease.

10. A method of preparing closed single-stranded nucleic acid loops, the method comprising:

providing a genomic DNA, a cDNA, or a DNA concatamer;

generating double-stranded fragments from the genomic DNA, the cDNA, or the concatamer, wherein first strands of the fragments are exonuclease-sensitive and second strands of the fragments are exonuclease-resistant; and,

separating the first strands of the fragments from the second strands to produce single-stranded fragments; and,

circularizing the single-stranded fragments to produce the closed single-stranded nucleic acid loops.

11. The method of claim 10, wherein the genomic DNA, the cDNA, or the concatamer is derived from a eukaryote.

12. The method of claim 10, wherein generating the double-stranded fragments that comprise the first exonuclease-sensitive strands and the second exonuclease-resistant strands comprises:

cleaving the genomic DNA, the cDNA, or the concatamer to produce linear double-stranded nucleic acids, and;

ligating exonuclease-sensitive tags to 5' ends of first strands of the linear double-stranded nucleic acids and ligating exonuclease-resistant tags to 5' ends of second strands of the linear double-stranded nucleic acids, thereby producing the population of double-stranded fragments with the first strands that comprise the exonuclease-sensitive 5' ends and the second strands that comprise the exonuclease-resistant 5' ends.

13. The method of claim 12, wherein cleaving the genomic DNA, the cDNA, or the concatamer comprises one or more of: enzymatic digestion, sonication, mechanical shearing, electrochemical cleavage, or nebulization of the genomic DNA, the cDNA, or the concatamer to produce the linear double-stranded nucleic acids.

14. The method of claim 12, wherein the tags comprise one or more moieties selected from: a ligand, a fluorescent label, a blocking group, a phosphorylated nucleotide, a phosphorothioated nucleotide, a biotinylated nucleotide, a methylated nucleotide, a uracil, a sequence capable of forming hairpin secondary structure, an oligonucleotide hybridization site, a restriction site, a DNA promoter, an RNA promoter, a sample or library identification sequence, and a cis regulatory sequence.

15. The method of claim 10, wherein generating the double-stranded fragments that comprise the first exonuclease-sensitive strands and the second exonuclease-resistant strands comprises:

annealing a population of primer pairs, which primer pairs comprise first primers that comprise exonuclease-sensitive 5' ends and second primers that comprise exonuclease-resistant 5' ends, to subsequences of the genomic DNA, the cDNA or the concatamer; and,

extending the primers with a polymerase to produce the double-stranded fragments with the first strands that comprise the exonuclease-sensitive 5' ends and the second strands that comprise the exonuclease-resistant 5' ends.

16. The method of claim 15, wherein the primers comprise one or more moieties selected from: a ligand, a fluorescent label, a blocking group, a phosphorylated nucleotide, a

phosphorothioated nucleotide, a biotinylated nucleotide, a methylated nucleotide, a uracil, a sequence capable of forming hairpin secondary structure, an oligonucleotide hybridization site, a restriction site, a DNA promoter, an RNA promoter, a sample or library identification sequence, and a cis regulatory sequence.

17. The method of claim 10, wherein separating the first strands from the second strands comprises digesting the first strands of the double-stranded fragments with an exonuclease to produce the single-stranded fragments.

18. The method of claim 10, wherein circularizing the single-stranded fragments to produce the closed single-stranded nucleic acid loops comprises:

annealing single-stranded nucleic acid splints to the single-stranded fragments, wherein first ends of the splints comprise nucleotide sequences complementary to nucleotide sequences at the first ends of the single-stranded fragments, and wherein second ends of the splints comprise nucleotide sequences complementary to nucleotide sequences at the second ends of the single-stranded fragments, such that the first and the second ends of each single-stranded fragment are brought into proximity with one another;

ligating the first and the second ends of the single-stranded fragments to one another; and,

removing the splints, thereby producing the closed single-stranded nucleic acid loops.

19. The method of claim 18, wherein removing the splints comprises digesting the splints with an exonuclease.

20. A composition, comprising:

a plurality of single-stranded nucleic acid fragments, which fragments comprise overlapping subsequences of a genomic DNA, a cDNA, or a DNA concatamer; and,

a population of single-stranded nucleic acid splints, wherein first ends of the splints are annealed to first ends of the fragments and second ends of the splints are annealed to second ends of the fragments, such that the first and second ends of each fragment are brought within proximity of one another.

21. The composition of claim 20, wherein the genomic DNA, the cDNA, or the concatamer is derived from a eukaryote.

22. A method of producing a population of closed single-stranded nucleic acid loops, the method comprising:

providing a genomic DNA, a cDNA, or a DNA concatamer;

generating double-stranded nucleic acid fragments from the genomic DNA, the cDNA, or the concatamer, which fragments comprise a gap site;

circularizing the double-stranded nucleic acid fragments; and,

removing the non-contiguous strands from the double-stranded loops, thereby producing a population of closed single-stranded nucleic acid loops.

23. The method of claim 22, wherein the genomic DNA, the cDNA, or the concatamer is derived from a eukaryote.

24. The method of claim 22, wherein generating the double-stranded fragments that comprise the gap site comprises:

cleaving the genomic DNA, the cDNA, or the concatamer to produce linear double-stranded nucleic acids, and;

ligating tags, which tags comprise the gap site, to the ends of the linear double-stranded nucleic acids to produce the double-stranded fragments that comprise the gap site.

25. The method of claim 24, wherein cleaving the genomic DNA, the cDNA, or the concatamer comprises: enzymatic digestion, sonication, mechanical shearing, electrochemical cleavage, or nebulization of the genomic DNA, the cDNA, or the concatamer to produce the linear double-stranded nucleic acids.

26. The method of claim 24, wherein the tags comprise one or more moieties selected from: a ligand, a fluorescent label, a blocking group, a phosphorylated nucleotide, a phosphorothioated nucleotide, a biotinylated nucleotide, a methylated nucleotide, a uracil, a sequence capable of forming hairpin secondary structure, an oligonucleotide hybridization site, a restriction site, a DNA promoter, an RNA promoter, a sample or library identification sequence, and a cis regulatory sequence.

27. The method of claim 22, generating the double-stranded fragments that comprise the gap site comprises:

annealing a population of primer pairs, which primers comprise the gap site, to subsequences of the genomic DNA, the cDNA, or the concatamer; and,

extending the primers with a polymerase, thereby producing the generating the double-stranded fragments that comprise the gap site.

28. The method of claim 27, wherein the primer pairs comprise one or more moieties selected from: a ligand, a fluorescent label, a blocking group, a phosphorylated nucleotide, a phosphorothioated nucleotide, a biotinylated nucleotide, a methylated nucleotide, a uracil, a sequence capable of forming hairpin secondary structure, an oligonucleotide hybridization

site, a restriction site, a DNA promoter, an RNA promoter, a sample or library identification sequence, and a cis regulatory sequence.

29. The method of claim 22, wherein the gap site comprises an unphosphorylated 5' end

30. The method of claim 22, wherein the gap site comprises a sequence recognized by a nicking enzyme.

31. The method of claim 30, wherein the nicking enzyme is uracil DNA deglycosylase.

32. The method of claim 22, wherein circularizing the double-stranded nucleic acid fragments comprises ligating first ends of the fragments to second ends of the fragments, thereby forming the population of double-stranded loops.

33. The method of claim 22, wherein removing the non-contiguous strands from the double-stranded loops comprises digesting the non-contiguous strands with an exonuclease that specifically cleaves the gap site, thereby producing the population of closed single-stranded nucleic acid loops.

34. The method of claim 22, wherein removing the non-contiguous strands from the double-stranded loops comprises:

cleaving sacrificial strands of the loops with a nicking enzyme that specifically cleaves the gap site to produce non-contiguous strands; and

digesting the non-contiguous strands with an exonuclease that specifically cleaves the gap site, thereby producing the population of closed single-stranded nucleic acid loops.

35. The method of claim 22, further comprising sequencing the circular single-stranded nucleic acids.

36. A composition, comprising a plurality of double-stranded circularized fragments which fragments comprise overlapping sequences of a genomic DNA or a cDNA, wherein the circularized fragments comprise non-contiguous strands and contiguous strands.

37. A method of producing a population of closed single-stranded nucleic acid loops, the method comprising:

providing a genomic DNA, a cDNA, or a DNA concatamer;

generating double-stranded fragments from the genomic DNA, the cDNA, or the concatamer, which fragments comprise ends that comprise a rolling-circle replication (RCR) protein recognition sequence;

copying target strands of the double-stranded fragments to produce a population of single-stranded copies, and;

circularizing the single-stranded copies to produce the population of closed single-stranded nucleic acid loops.

38. The method of claim 37, wherein the genomic DNA, the cDNA, or the concatamer is derived from a eukaryote.

39. The method of claim 37, wherein generating the population of double-stranded fragments with the ends that comprise the rolling-circle replication (RCR) protein recognition sequence comprises:

cleaving the genomic DNA, the cDNA, or the concatamer to produce linear double-stranded nucleic acids, and;

ligating tags that comprise the rolling-circle replication (RCR) protein recognition sequence to the ends of the linear double-stranded nucleic acids to produce the double-stranded fragments with the ends that comprise the rolling-circle replication (RCR) protein recognition sequence.

40. The method of claim 39, wherein cleaving the genomic DNA, the cDNA, or the concatamer comprises: enzymatic digestion, sonication, mechanical shearing, electrochemical cleavage, or nebulization of the genomic DNA, the cDNA, or the concatamer to produce the linear double-stranded nucleic acids.

41. The method of claim 39, wherein the tags comprise one or more moieties selected from: a ligand, a fluorescent label, a blocking group, a phosphorylated nucleotide, a phosphorothioated nucleotide, a biotinylated nucleotide, a methylated nucleotide, a uracil, a sequence capable of forming hairpin secondary structure, an oligonucleotide hybridization site, a restriction site, a DNA promoter, an RNA promoter, a sample or library identification sequence, and a cis regulatory sequence.

42. The method of claim 37, wherein the double stranded fragments with the ends that comprise the rolling-circle replication (RCR) protein recognition sequence are generated by:

annealing a population of primer pairs that comprise the rolling-circle replication (RCR) protein recognition sequence to subsequences of the genomic DNA, the cDNA, or the concatamer; and,

extending the primers with a polymerase, thereby producing the fragments with the ends that comprise the rolling-circle replication (RCR) protein recognition sequence.

43. The method of claim 42, wherein the primers comprise one or more moieties selected from: a ligand, a fluorescent label, a blocking group, a phosphorylated nucleotide, a phosphorothioated nucleotide, a biotinylated nucleotide, a methylated nucleotide, a uracil, a

sequence capable of forming hairpin secondary structure, an oligonucleotide hybridization site, a restriction site, a DNA promoter, an RNA promoter, a sample or library identification sequence, and a cis regulatory sequence.

44. The method of claim 37, wherein the RCR protein recognition sequence comprises a sequence that is bound by an RCR protein that comprises a histidine-U-histidine-U-U-U amino acid motif, wherein U is a bulky hydrophobic amino acid residue.

45. The method of claim 44 wherein the RCR protein is cisA.

46. The method of claim 37, wherein copying the target strands comprises:
nicking secondary strands of the double-stranded fragments,
displacing the secondary strands, and
replicating sequences of the target strands to produce the population of single-stranded copies.

47. The method of claim 46, wherein nicking the secondary strands of the double-stranded fragments comprises cleaving the strands with an RCR protein.

48. The method of claim 46, wherein displacing the secondary strands comprises unwinding the secondary strands from the target strands of the double-stranded fragments with a replisome.

49. The method of claim 48, wherein the replisome comprises a single-stranded DNA binding protein (SSB), a helicase, a polymerase, and an RCR protein.

50. The method of claim 46, wherein the replicating comprises synthesizing nucleic acid strands that are complementary to the target strands to produce the population of single-stranded copies.

51. The method of claim 37, wherein circularizing the single-stranded copies comprises ligating first ends of the single-stranded copies to second ends of the single-stranded copies with an RCR protein to produce the closed single-stranded nucleic acid loops.

52. A composition comprising a plurality of double-stranded nucleic acids wherein the nucleic acids comprise overlapping subsequences of a genomic DNA, a cDNA, or a DNA concatamer, and wherein the nucleic acids comprise ends that comprise a rolling-circle replication (RCR) protein recognition sequence.

53. The composition of claim 52, wherein the genomic DNA, the cDNA, or the concatamer is derived from a eukaryote.

54. The composition of claim 52, wherein the nucleic acids are closed loops.

55. The composition of claim 52, wherein the nucleic acids are linear.

56. The composition of claim 52, wherein the rolling-circle replication (RCR) protein recognition sequence comprises a sequence that is bound by an RCR protein that comprises a histidine-U-histidine-U-U-U amino acid motif, wherein U is a bulky hydrophobic amino acid residue.

57. The composition of claim 56, wherein the RCR protein is cisA.

58. A method of producing a population of single-stranded linear nucleic acids, the method comprising:

providing a genomic DNA, a cDNA, or a DNA concatamer;

generating a population of double-stranded fragments from the genomic DNA, the cDNA, or the concatamer, which fragments comprise ends that comprise a sequence recognized by a nicking enzyme;

circularizing the fragments to produce a set of closed double-stranded nucleic acid loops,

cleaving sacrificial strands of the closed loops with the nicking enzyme to produce nicked loops;

copying template strands of the nicked loops to produce copied strands that comprise sequences that are recognized by the nicking enzyme; and,

nicking the copied strands with the nicking enzyme to produce the single-stranded linear nucleic acids.

59. The method of claim 58, wherein the genomic DNA, the cDNA, or the concatamer is derived from a eukaryote.

60. The method of claim 58, wherein generating the population of double-stranded fragments with the ends that comprise the sequence recognized by the nicking enzyme comprises:

cleaving the genomic DNA, the cDNA, or the concatamer to produce linear double-stranded nucleic acids, and;

ligating tags, which tags comprise the sequence recognized by the nicking enzyme, to ends of the linear double-stranded nucleic acids to produce the double-stranded fragments with the ends that comprise the sequence recognized by the nicking enzyme.

61. The method of claim 60, wherein cleaving the genomic DNA, the cDNA, or the concatamer comprises: enzymatic digestion, sonication, mechanical shearing, electrochemical cleavage, or nebulization of the genomic DNA, the cDNA, or the concatamer to produce the double-stranded fragments.

62. The method of claim 60, wherein the tags comprise one or more moieties selected from: a ligand, a fluorescent label, a blocking group, a phosphorylated nucleotide, a phosphorothioated nucleotide, a biotinylated nucleotide, a methylated nucleotide, a uracil, a sequence capable of forming hairpin secondary structure, an oligonucleotide hybridization site, a restriction site, a DNA promoter, an RNA promoter, a sample or library identification sequence, and a cis regulatory sequence.

63. The method of claim 58, wherein the double-stranded fragments with the ends that comprise the sequence recognized by the nicking enzyme are generated by:

annealing a population of primer pairs, which primer pairs comprise the sequence recognized by the nicking enzyme, to subsequences of the genomic DNA, the cDNA, or the concatamer; and,

extending the primers with a polymerase, thereby producing the double-stranded fragments with the ends that comprise the sequence recognized by the nicking enzyme.

64. The method of claim 63, wherein the primers comprise one or more moieties selected from: a ligand, a fluorescent label, a blocking group, a phosphorylated nucleotide, a phosphorothioated nucleotide, a biotinylated nucleotide, a methylated nucleotide, a uracil, a sequence capable of forming hairpin secondary structure, an oligonucleotide hybridization site, a restriction site, a DNA promoter, an RNA promoter, a sample or library identification sequence, and a cis regulatory sequence.

65. The method of claim 58, wherein the circularizing comprises ligating first ends of the fragments to second ends of the fragments to produce the closed loops.

66. The method of claim 58, wherein the copying comprises:

displacing the sacrificial strands; and,

replicating the template strands, wherein the replicating comprises synthesizing nucleic acid strands that are complementary to the template strands, to produce the copied strands.

67. A composition comprising a plurality of closed double-stranded nucleic acid loops, wherein the loops comprise overlapping fragments of a genomic DNA, a cDNA, or a DNA concatamer, and wherein the loops comprise a sequence that is specifically recognized by a nicking enzyme.

68. The composition of claim 67, wherein the genomic DNA, the cDNA, or the concatamer is derived from a eukaryote.

69. A method of generating a set of single-stranded circular nucleic acids, the method comprising:

providing a genomic DNA, a cDNA, or a DNA concatamer;

producing a population of double-stranded nucleic acid fragments from the genomic DNA, the cDNA, or the concatamer;

providing first hairpin oligonucleotides to first ends of the double-stranded nucleic acid fragments and providing second hairpin oligonucleotides to second ends of the double-stranded nucleic acid fragments, thereby generating the set of single-stranded circular nucleic acids that comprise regions of internal complementarity.

70. The method of claim 69, wherein the genomic DNA, the cDNA, or the concatamer is derived from a eukaryote.

71. The method of claim 69, wherein producing the population of double-stranded nucleic acid fragments comprises cleaving the genomic DNA, the cDNA, or the concatamer.

72. The method of claim 71, wherein cleaving the genomic DNA, the cDNA, or the concatamer comprises: enzymatic digestion, sonication, mechanical shearing, electrochemical cleavage, or nebulization of the genomic DNA, the cDNA, or the concatamer to produce the double-stranded fragments.

73. The method of claim 69, wherein producing the population of double-stranded nucleic acid fragments comprises:

annealing a population of primer pairs to subsequences of the genomic DNA, the cDNA, or the concatamer; and,

extending the primer pairs with a polymerase to produce the double-stranded fragments.

74. The method of claim 73, wherein the primers comprise one or more moieties selected from: a ligand, a fluorescent label, a blocking group, a phosphorylated nucleotide, a phosphorothioated nucleotide, a biotinylated nucleotide, a methylated nucleotide, a uracil, a sequence capable of forming hairpin secondary structure, an oligonucleotide hybridization site, a restriction site, a DNA promoter, an RNA promoter, a sample or library identification sequence, and a cis regulatory sequence.

75. The method of claim 69, wherein providing the hairpin oligonucleotides to the ends of the double-stranded nucleic acids fragments comprises:

annealing the first hairpin oligonucleotides to the first ends of the fragments;
annealing the second hairpin oligonucleotides to the second ends of the fragments;
and,

ligating the hairpin oligonucleotides to the ends of the fragments to generate the set of single-stranded circular nucleic acids.

76. The method of claim 69, wherein the attaching comprises:

linking 5' strands of the first hairpins to 3' strands of the first ends of the fragments;

linking 3' strands the first hairpins to 5' strands of the first ends of the fragments;

linking 5' strands of the second hairpins to 3' strands of the second ends of the fragments; and,

linking 3' strands of the second hairpins to 5' strands of the second ends of the fragments to form a population of circular single-stranded nucleic acids

77. The method of claim 69, wherein the first hairpins comprise first single-stranded terminal sequences complementary to second single-stranded terminal sequences at the first ends of the fragments and the second hairpins comprise third single-stranded terminal sequences complementary to fourth single stranded terminal sequences at the second ends of the fragments.

78. The method of claim 69, wherein the first or the second hairpins comprise one or more moieties selected from: a ligand, a fluorescent label, a blocking group, a phosphorylated nucleotide, a phosphorothioated nucleotide, a biotinylated nucleotide, a methylated nucleotide, a uracil, a sequence capable of forming hairpin secondary structure, an oligonucleotide hybridization site, a restriction site, a DNA promoter, an RNA promoter, a sample or library identification sequence, and a cis regulatory sequence.

79. A composition comprising first hairpin oligonucleotides, second hairpin oligonucleotides, and a plurality of double-stranded nucleic acid fragments, which fragments comprise overlapping subsequences of a genomic DNA, a cDNA, or a DNA concatamer.

80. The composition of claim 79, wherein the first and/or second hairpin oligonucleotides comprise one or more moieties selected from: a ligand, a fluorescent label, a blocking group, a phosphorylated nucleotide, a phosphorothioated nucleotide, a biotinylated nucleotide, a methylated nucleotide, a uracil, a sequence capable of forming hairpin secondary structure, an oligonucleotide hybridization site, a restriction site, a DNA

promoter, an RNA promoter, a sample or library identification sequence, and a cis regulatory sequence.

81. The composition of claim 79, wherein the first hairpins comprise first single-stranded terminal sequences complementary to second single-stranded terminal sequences at the first ends of the fragments and the second hairpins comprise third single-stranded terminal sequences complementary to fourth single stranded terminal sequences at the second ends of the fragments.

82. The composition of claim 79, wherein the genomic DNA, the cDNA, or the concatamer is derived from a eukaryote.

83. A method of producing linear single-stranded nucleic acids, the method comprising:
providing a genomic DNA, a cDNA, or a DNA concatamer;
generating a set of double-stranded nucleic acid fragments from the genomic DNA, the cDNA, or the concatamer;

nicking sacrificial strands of the double-stranded nucleic acid fragments to produce nicked fragments;

displacing the sacrificial strands from the fragments; and,

copying sequences between nick sites and ends on un-nicked strands that are exposed by the displacement of the sacrificial strands with a strand displacing polymerase to produce the linear single-stranded nucleic acids.

84. The method of claim 83, wherein the genomic DNA, the cDNA, or the concatamer is derived from a eukaryote.

85. The method of claim 83, wherein producing the population of double-stranded nucleic acid fragments comprises cleaving the genomic DNA, the cDNA, or the concatamer.

86. The method of claim 85, wherein cleaving the genomic DNA, the cDNA, or the concatamer comprises one or more of: enzymatic digestion, sonication, mechanical shearing, electrochemical cleavage, or nebulization of the genomic DNA, the cDNA, or the concatamer to produce the double-stranded fragments.

87. The method of claim 83, wherein producing the population of double-stranded nucleic acid fragments comprises

annealing a population of primer pairs to subsequences of the genomic DNA, the cDNA, or the concatamer; and,

extending the primer pairs with a polymerase to produce the double-stranded fragments.

88. The method of claim 87, wherein the primers comprise one or more moieties selected from: a ligand, a fluorescent label, a blocking group, a phosphorylated nucleotide, a phosphorothioated nucleotide, a biotinylated nucleotide, a methylated nucleotide, a uracil, a sequence capable of forming hairpin secondary structure, an oligonucleotide hybridization site, a restriction site, a DNA promoter, an RNA promoter, a sample or library identification sequence, and a cis regulatory sequence.

89. The method of claim 83, wherein nicking the sacrificial strands comprises cleaving the sacrificial strands with an enzyme to produce the nicked fragments.

90. The method of claim 83, wherein displacing the sacrificial strands comprises removing the sacrificial strands from the un-nicked strands with the strand-displacing polymerase.

91. The method of claim 83, wherein the exposed sequences are copied by:
replicating sequences between the nick sites and the ends on the un-nicked strands with the strand-displacing polymerase, wherein replicating the sequences comprises synthesizing nucleic acid strands that are complementary to the un-nicked strands, to produce the linear single-stranded nucleic acids.

92. The method of claim 83, wherein the strand polymerase is a Phi29 polymerase, a Phi29-like polymerase, PolI polymerase or a BstI polymerase.

METHODS, COMPOSITIONS AND SYSTEMS FOR SAMPLE PREPARATION

ABSTRACT OF THE DISCLOSURE

Provided are methods and compositions for the production of linear single-stranded nucleic acids, which can be used as templates in high-throughput sequencing systems.. Also provided are methods and compositions for the production of closed single-stranded nucleic acid loops, which can be used as templates in high-throughput sequencing systems.

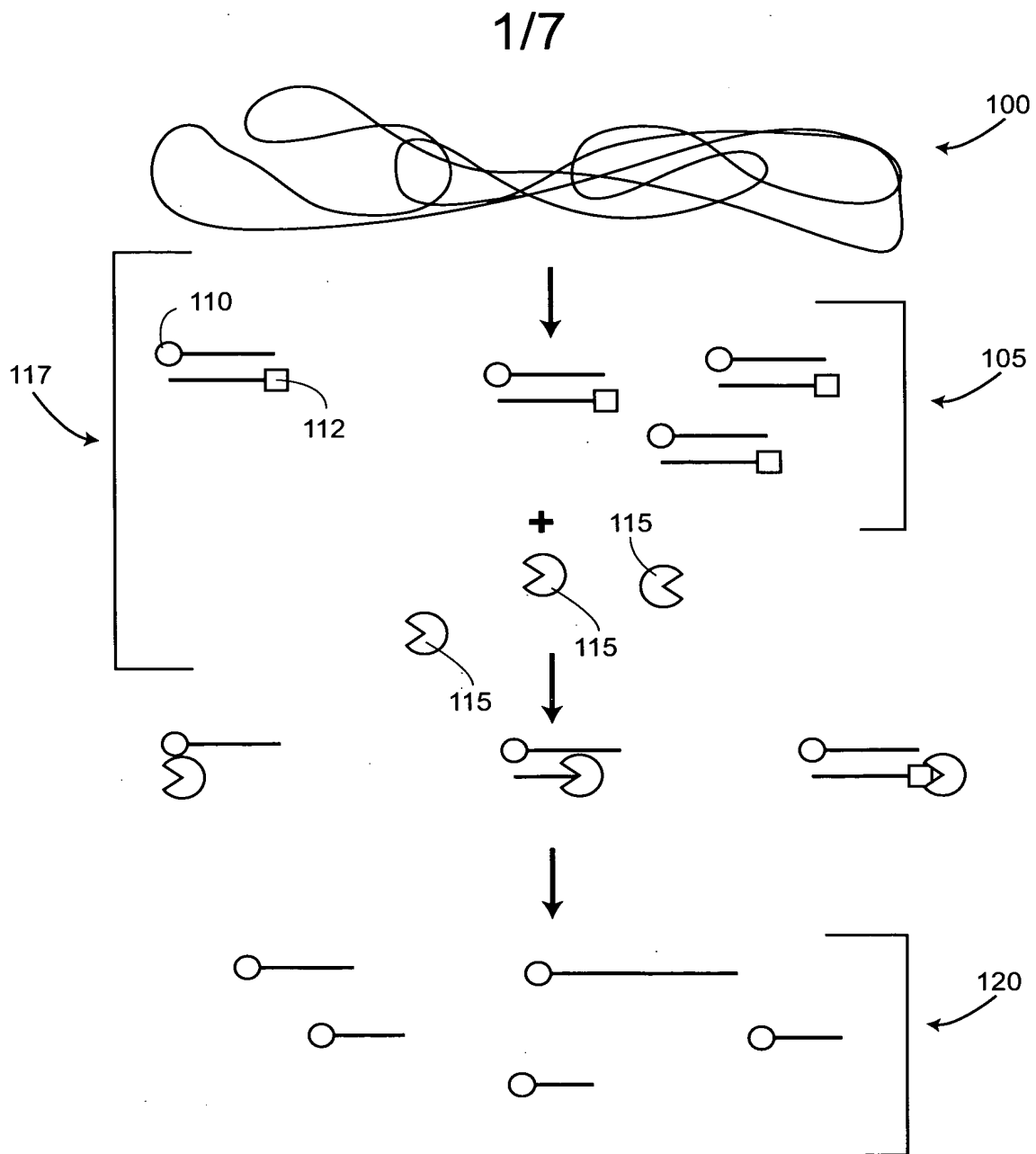
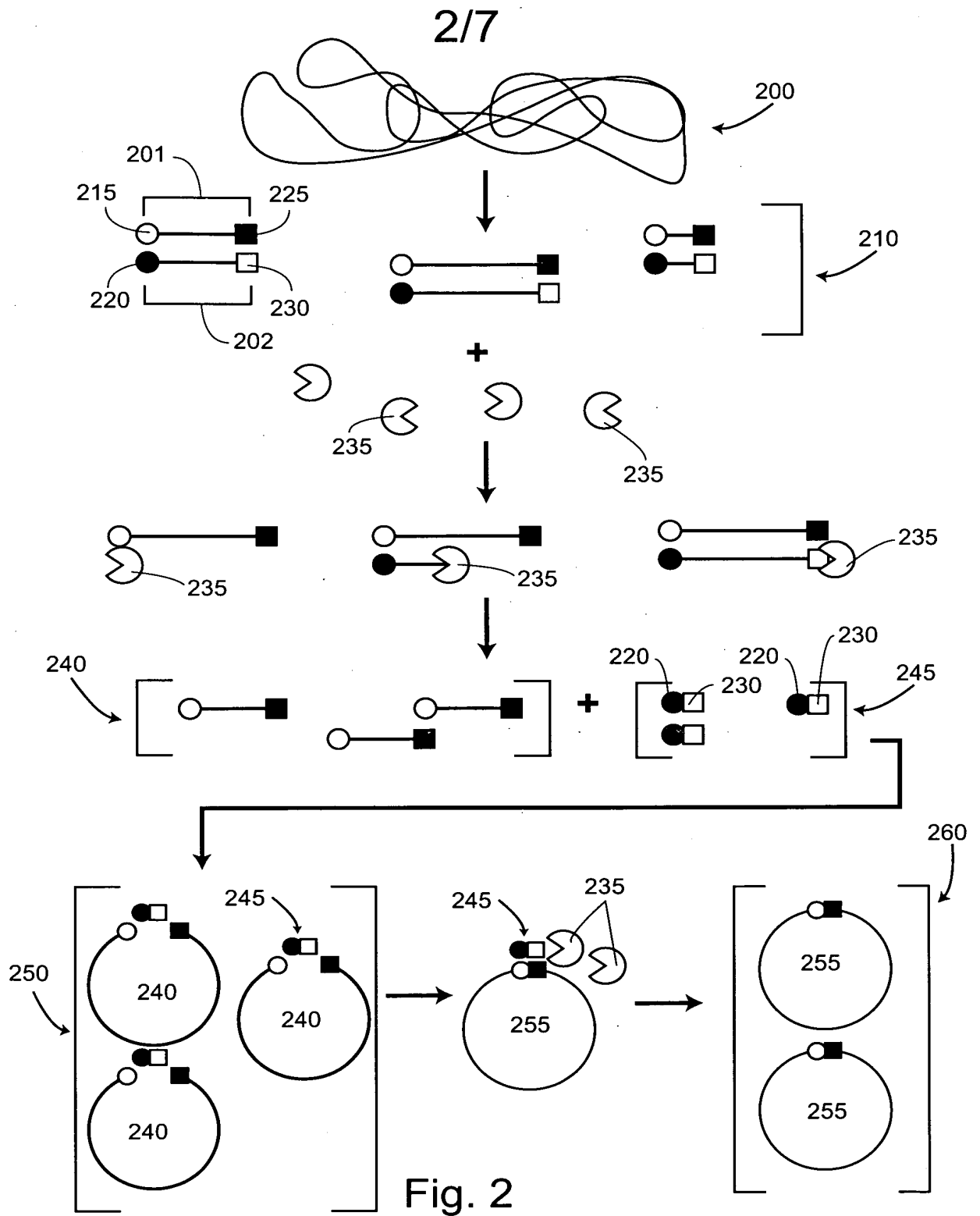


Fig. 1



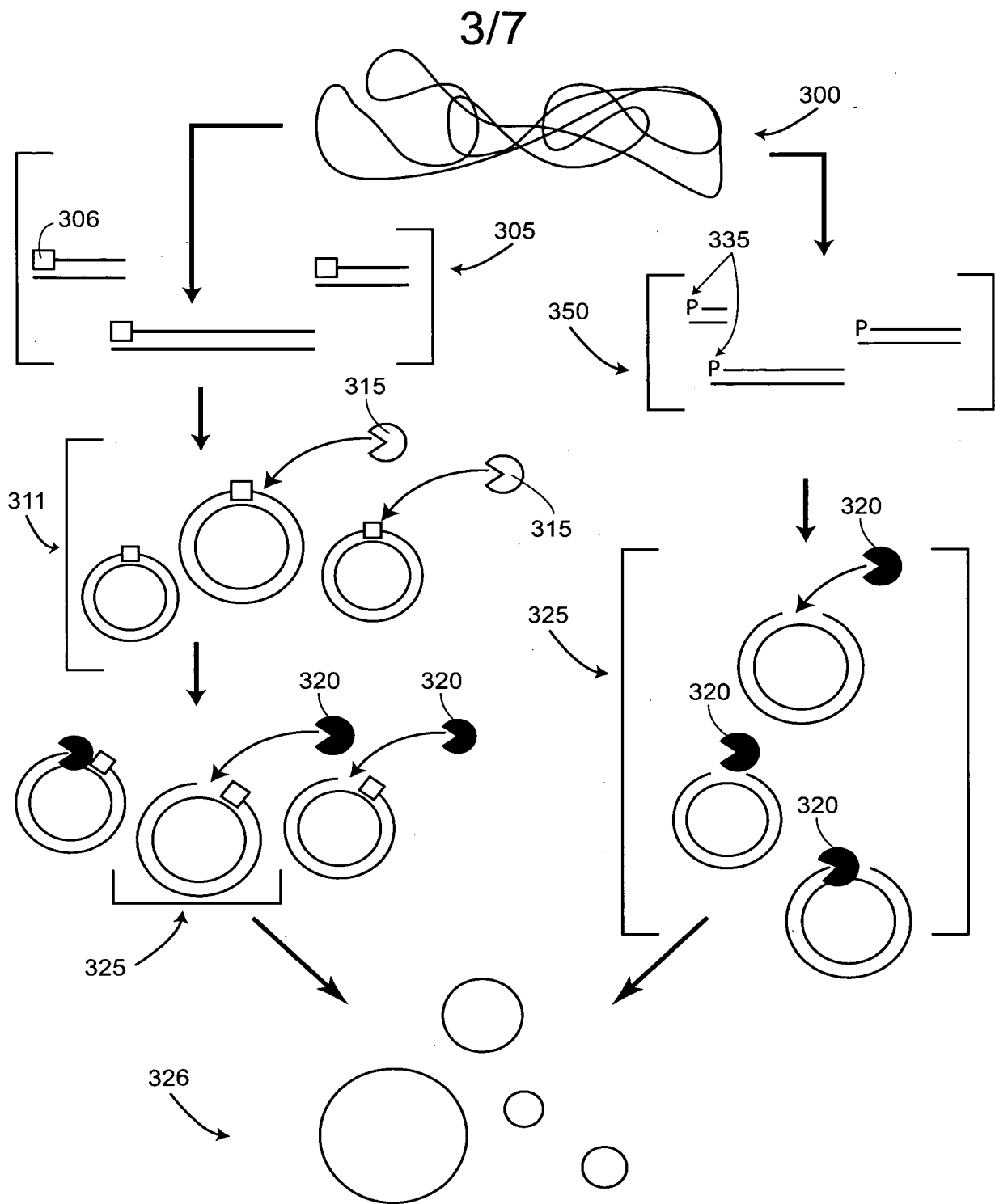


Fig. 3

4/7

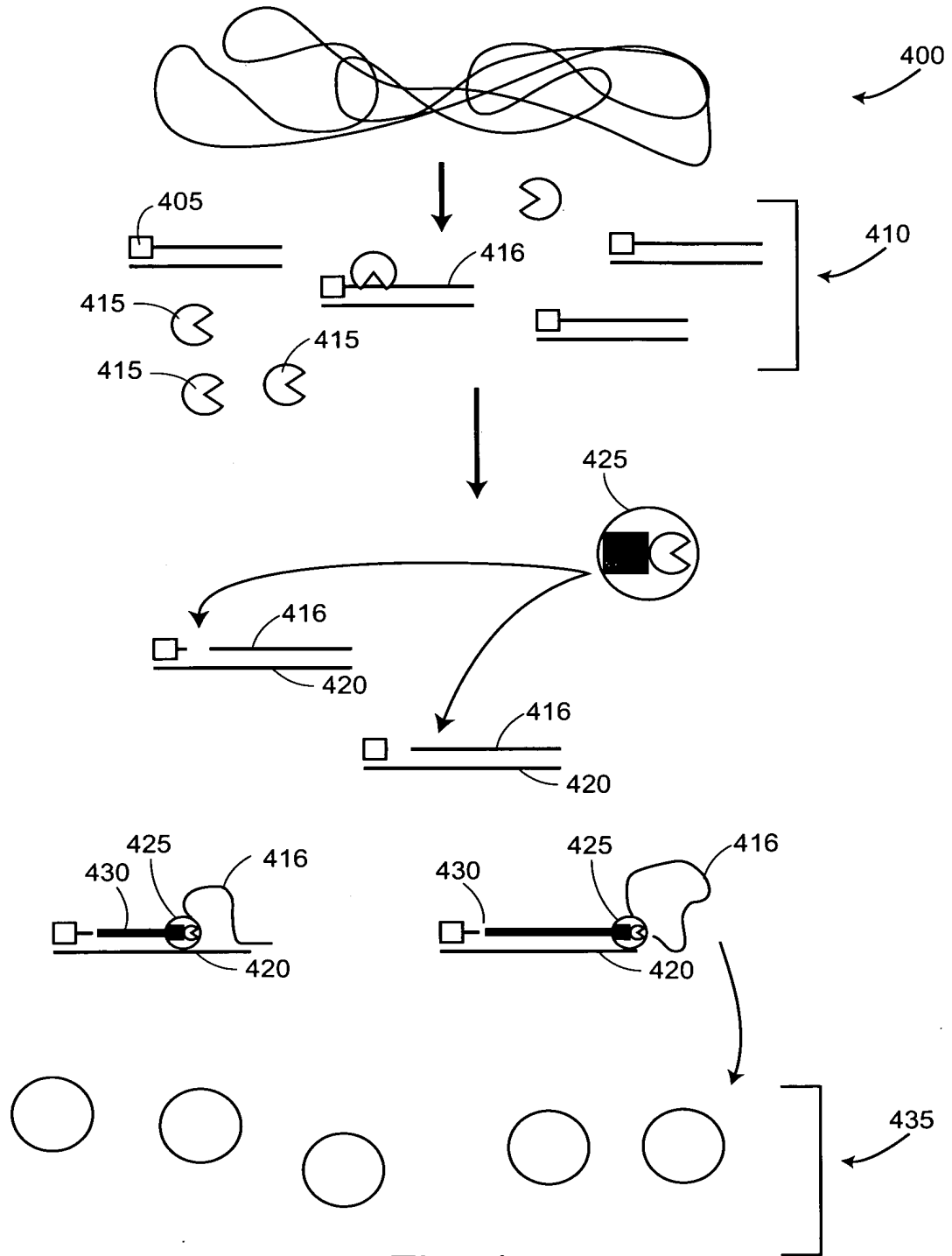


Fig. 4

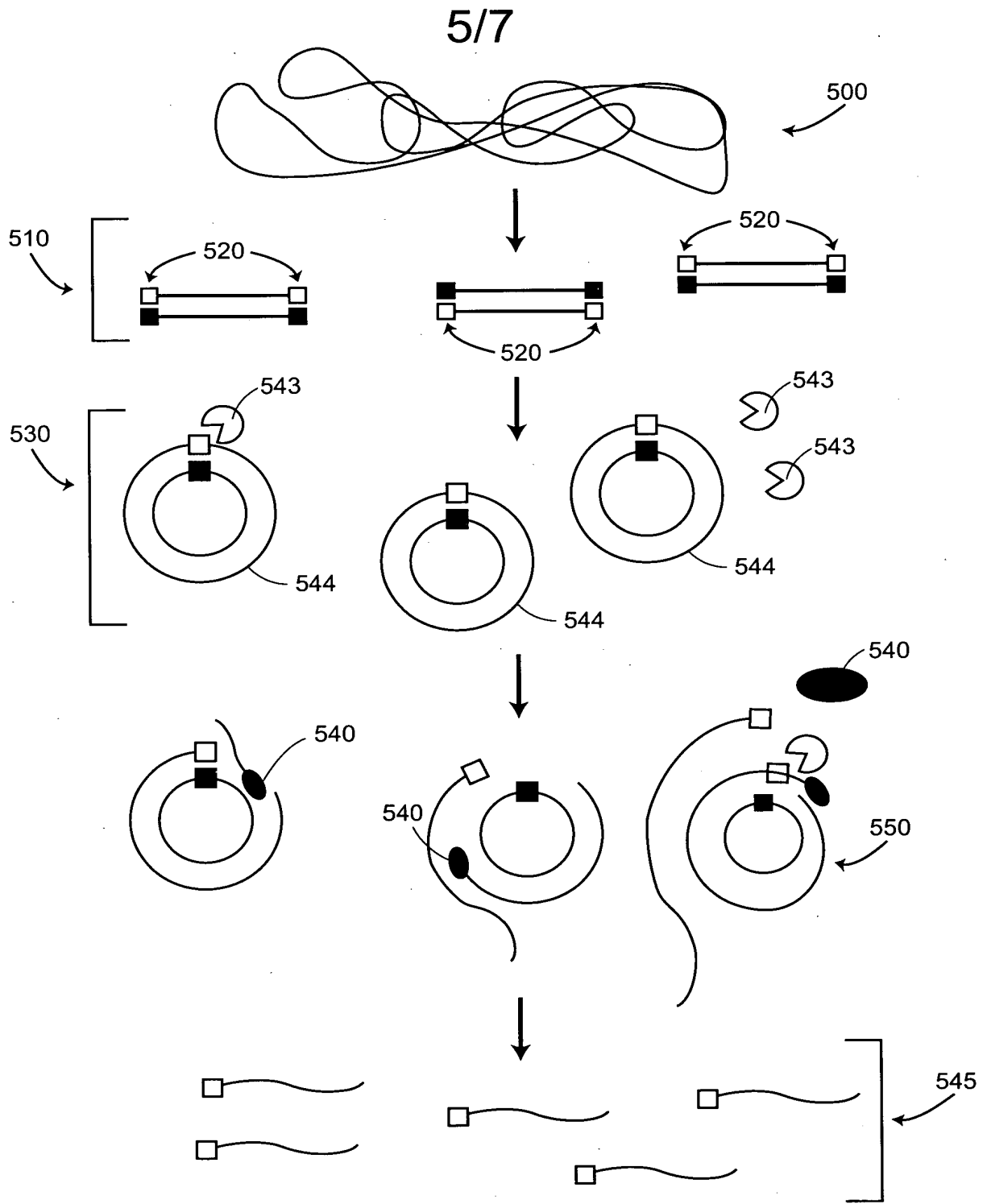
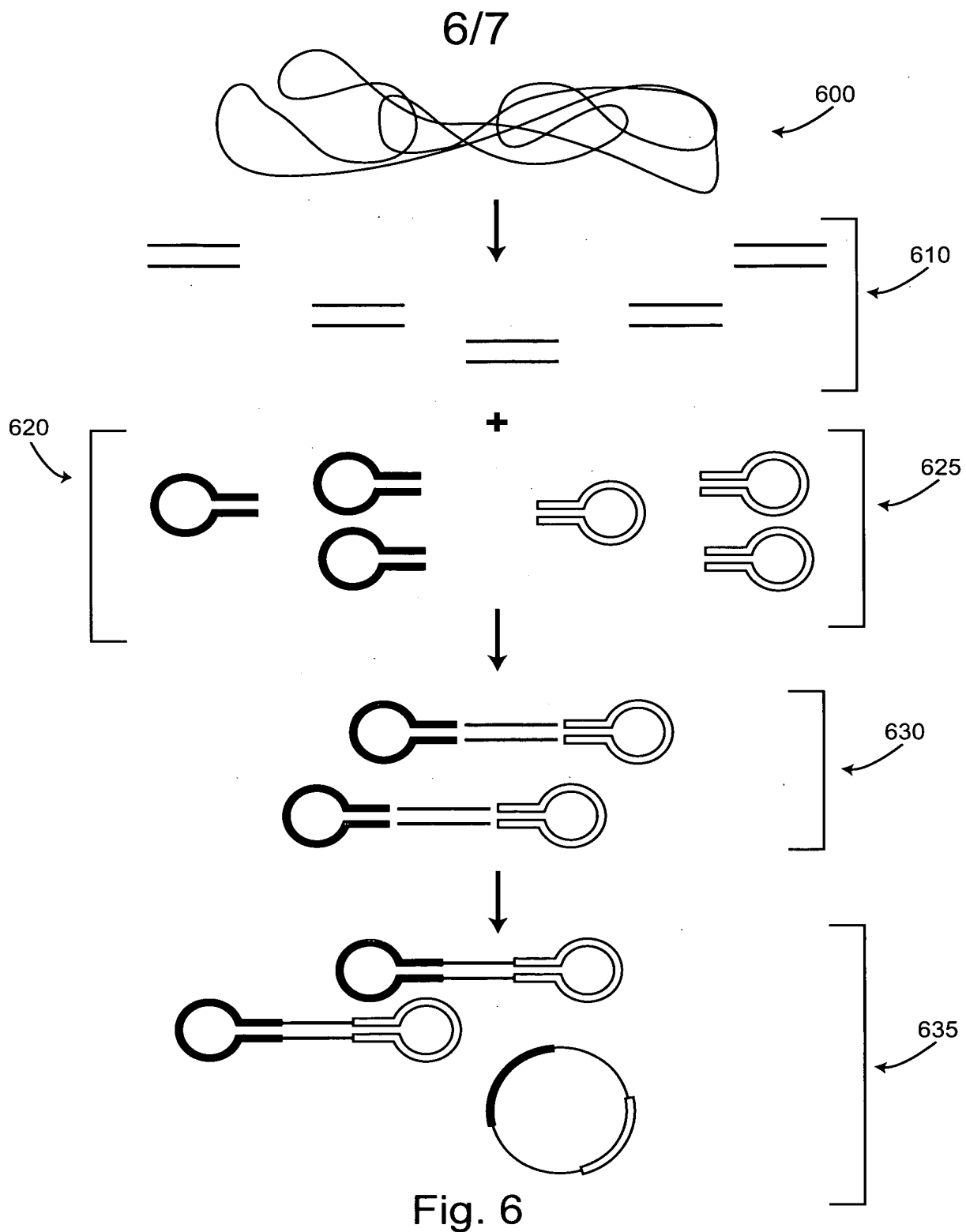


Fig. 5



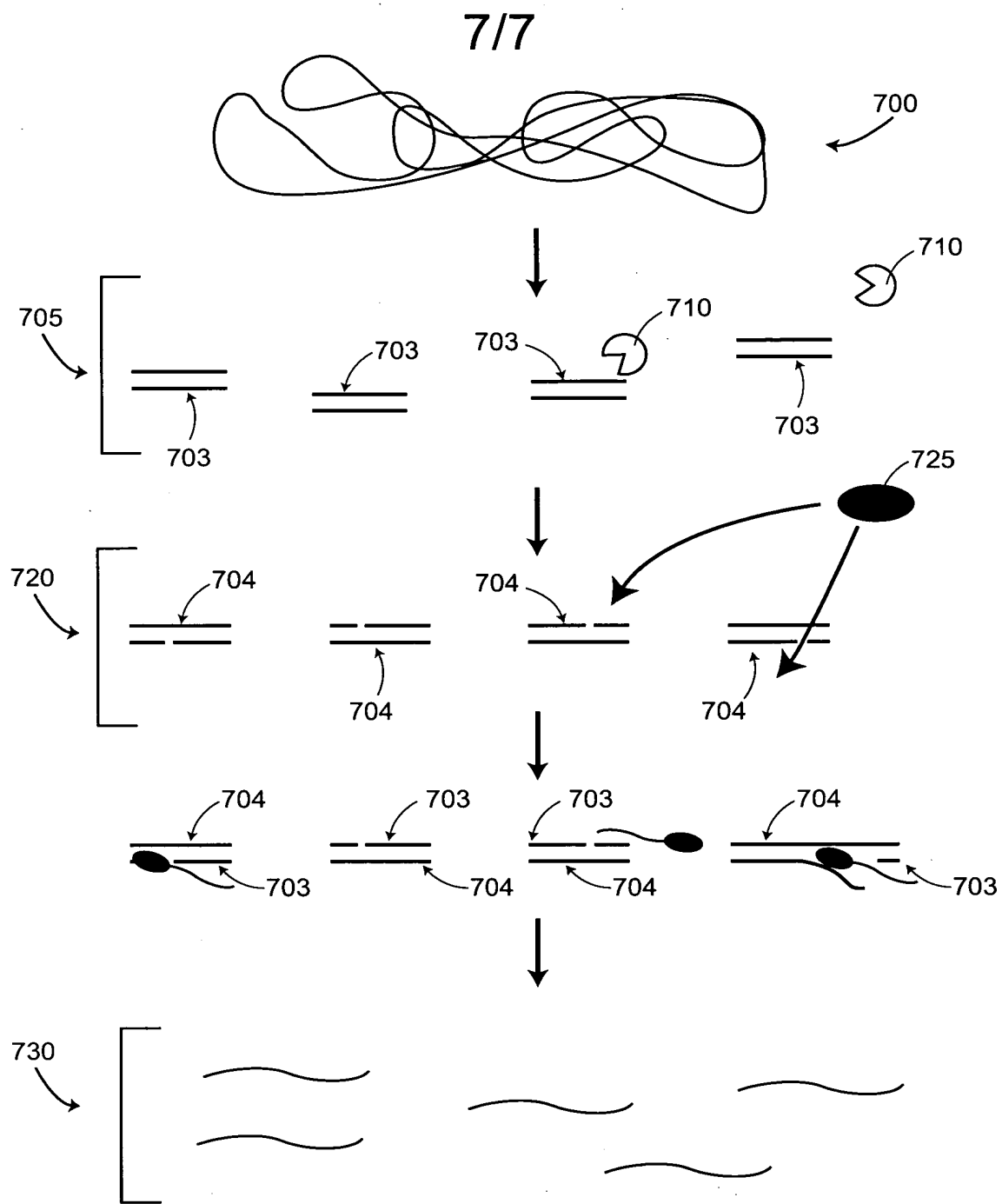


Fig. 7

PATENT APPLICATION SERIAL NO. _____

U.S. DEPARTMENT OF COMMERCE
PATENT AND TRADEMARK OFFICE
FEE RECORD SHEET

04/01/2008 HDESTA1 00000006 500893 61072160
01 FC:2005 105.00 DA

PTO-1556
(5/87)

*U.S. Government Printing Office: 2002- 489-267/69033