



Diversity and Analysis of Bacterial Terpene Synthases

Yuuki Yamada*, David E. Cane[†], Haruo Ikeda*,¹

*Laboratory of Microbial Engineering, Kitasato Institute for Life Sciences, Kitasato University, Kanagawa, Japan

[†]Department of Chemistry, Box H, Brown University, Providence, Rhode Island, USA

¹Corresponding author: e-mail address: ikeda@ls.kitasato-u.ac.jp

Contents

1. Terpenoid Metabolites from Bacterial Cultures	124
2. Bacterial Terpene Synthases	124
2.1 Monoterpene synthases	125
2.2 Sesquiterpene synthases	126
2.3 Diterpene synthases	128
3. Methods for the Study of Bacterial Terpene Synthases	129
3.1 Bioinformatic analysis of bacterial terpene synthases	129
3.2 Expression of genes encoding bacterial terpene synthases in heterologous hosts	153
References	159

Abstract

Terpenoid compounds are generally considered to be plant or fungal metabolites, although a small number of odorous terpenoid metabolites of bacterial origin have been known for many years. Recently, extensive bacterial genome sequencing and bioinformatic analysis of deduced bacterial proteins using a profile hidden Markov model have revealed more than a hundred distinct predicted terpene synthase genes. Although some of these synthase genes might be silent in the parent microorganisms under normal laboratory culture conditions, the controlled overexpression of these genes in a versatile heterologous host has made it possible to identify the biochemical function of cryptic genes and isolate new terpenoid metabolites.



1. TERPENOID METABOLITES FROM BACTERIAL CULTURES

Tens of thousands of terpenoid metabolites, including monoterpenes, sesquiterpenes, and diterpenes, are present in both terrestrial and marine higher plants, in liverworts, and in fungi, with only a relatively minor fraction having been isolated from prokaryotes. These compounds act as antibiotics, hormones, flavor or odor constituents, and pigments. Some of them also possess other physiologically or commercially important properties, such as vitamins (A, D, E, K, and coenzyme Q) and antitumor agent (paclitaxel) (Glasby, 1982; Sacchettini & Poulter, 1997). The study of bacterial terpenoid metabolites has its roots in the detection of odorous and volatile metabolites (Berthelot & André, 1891). The pioneering study of bacterial terpenoid metabolites as volatile compounds was accomplished by Gerber, who speculated that the typical odors of actinomycetes might be due to the production of terpenes (Gerber, 1967, 1969, 1971, 1973; Gerber & Lechevalier, 1965). In the years that followed, a variety of terpene hydrocarbons and alcohols have been isolated from bacteria (Jáchymová, Votruba, Viden, & Rezanka, 2002; Pollack & Berger, 1996; Schöller, Gürtler, Redersen, Molin & Wilkins, 2002; Wilkins & Schöller, 2009). In this chapter, we review these occurrences and describe *in silico* and bench-level procedures for their analysis.



2. BACTERIAL TERPENE SYNTHASES

Several cyclic monoterpene, sesquiterpene, and diterpene hydrocarbons and alcohols are formed by variations of the common terpene cyclization mechanism that is initiated by enzyme-catalyzed ionization to form allylic cations from the universal acyclic precursors, geranyl diphosphate (GPP), farnesyl diphosphate (FPP), and geranylgeranyl diphosphate (GGPP) that are themselves synthesized by mechanistically related condensations of the 5-carbon building blocks dimethylallyl diphosphate (DMAPP) and isopentenyl diphosphate (IPP). Intramolecular electrophilic attack of the intermediate allylic cations on the central or distal double bonds of the substrate, followed by well-precedented cationic transformations, including hydride shifts and carbon-carbon backbone rearrangements (Wagner-Meerwein reactions), and ultimate quenching of the positive charge either by deprotonation or by capture of a nucleophilic water molecule, can account for the formation of the enormous variety of several hundred cyclic terpene compounds. The initiation of formation of cyclic sesquiterpenes from FPP is illustrated in Fig. 7.1.

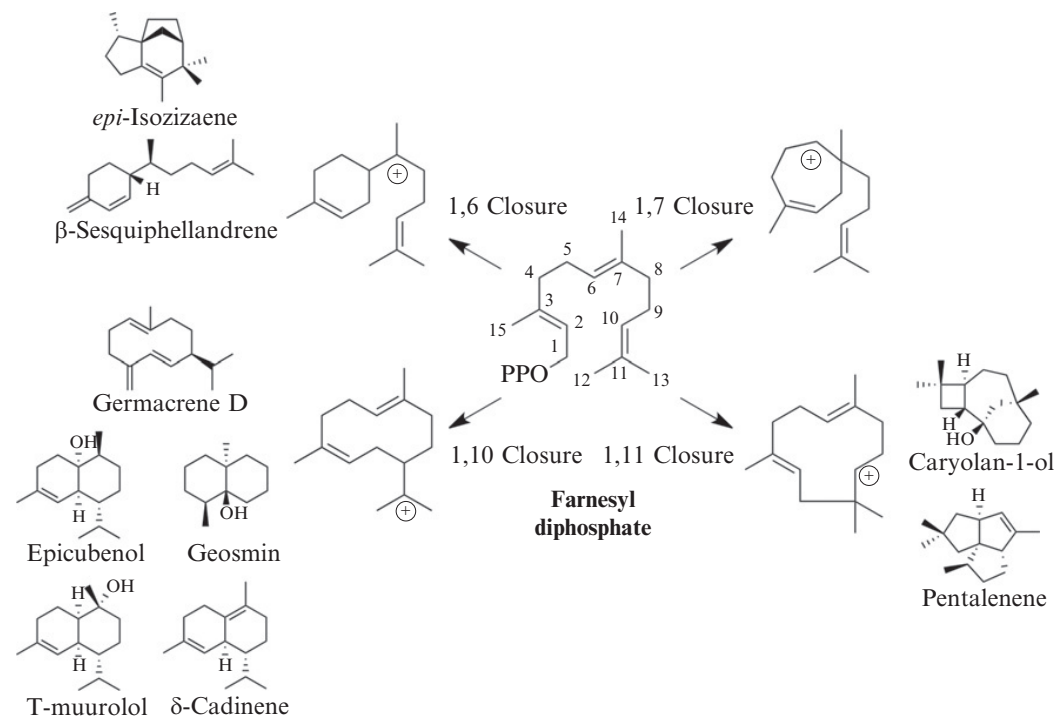


Figure 7.1 Cyclization of farnesyl diphosphate by sesquiterpene synthases.

2.1. Monoterpene synthases

The homomonoterpene synthase, 2-methylisoborneol synthase, was independently isolated from *Streptomyces lasaliensis* NRRL 3382 (Komatsu, Tsuda, Omura, Oikawa, & Ikeda, 2008) and *Streptomyces coelicolor* A3(2) (Wang & Cane, 2008). The 1446-bp *tpc* gene of *S. lasaliensis* NRRL 3382 and the 1323-bp *sco7700* gene of *S. coelicolor* A3(2) encode 481-aa and 440-aa proteins, respectively, with less than 20% identity to pentalenene synthase of *Streptomyces exfoliatus* UC5319, the first bacterial terpene synthase for which the full sequence was determined and thus commonly used as a standard of comparison for newly characterized bacterial terpene synthases. The two 2-methylisoborneol synthases each incorporate the same variants of the two characteristic terpene synthase divalent metal-binding motifs, the acidic amino acid-rich -**DDCYCED**- and the more conventional downstream NSE triad -**NDLYSYTKE**-. All genes encoding 2-methylisoborneol synthase are separated by 16–45 bp from the proximal downstream gene, which was originally annotated as a generic C-methyltransferase. It appears that a two-gene biosynthetic operon is therefore involved in the formation of 2-methylisoborneol. The corresponding downstream gene products of *S. lasaliensis* NRRL 2338 and *S. coelicolor* A3(2) were confirmed to catalyze methylation of GPP at C-2 using *S*-adenosyl-L-methionine, to give exclusively (*E*)-2-methylgeranyl diphosphate (2-MeGPP). The 2-methylisoborneol synthase

then catalyzes cyclization of the methylated acyclic substrate, 2-MeGPP, to generate 2-methylisoborneol. The Gram-negative gamma-proteobacterium *Pseudomonas fluorescens* Pf0-1 harbors an orthologous gene product, Pfl01_1841, which has significant sequence similarity to actinomycete 2-methylisoborneol synthases. Recombinant Pfl01_1841 has been shown to catalyze generation of the homomonoterpene hydrocarbon 2-methylenebornane from 2-MeGPP without formation of any detectable 2-methylisoborneol. Surprisingly, a gene corresponding to the requisite GPP 2-methyltransferase could not be found in the *P. fluorescens* Pf0-1 genome and 2-methylenebornane was not detected in culture extracts (Chou, Ikeda, & Cane, 2011). By contrast, production of 2-methylenebornane has been observed in *Micromonospora olivasterospora* NRRL 8178 and a truncated recombinant protein from *M. olivasterospora* expressed in *E. coli* catalyzes formation of 2-methylenebornane from 2-MeGPP (Ikeda, unpublished data).

2.2. Sesquiterpene synthases

The first studies of a bacterial terpene synthase came from investigations of the biosynthesis of the sesquiterpenoid antibiotic pentalenolactone. Pentalenene synthase, first isolated from *S. exfoliatus* UC5319, was shown to catalyze the cyclization of FPP, pentalenene, the parent sesquiterpene hydrocarbon of the pentalenolactone family of antibiotics (Cane et al., 1994). Extensive experiments with stereospecifically labeled FPP have established the detailed cyclization mechanism (Cane et al., 1990). Eventually purification allowed cloning of the *S. exfoliatus* gene encoding pentalenene synthase and expression in *E. coli*. The crystal structure of pentalenene synthase revealed an all α -helical fold very similar to that of tobacco *epi*-aristolochene synthase, with which pentalenene synthase shares insignificant primary amino acid sequence similarity (Lesburg, Zhai, Cane, & Christianson, 1997). Indeed, this same fold has been found in many terpene synthases of diverse biological origin and is believed to be universal for this class of cyclases. Sequencing of the *Streptomyces avermitilis* genome (Ikeda et al., 2003) revealed a $\sim$ 13.6-kb cluster harboring 13 unidirectionally transcribed protein-coding sequences, one of which, *sav2998*, encodes a 336-aa protein (defined as PtlA) with 76% identity to the previously characterized pentalenene synthase of *S. exfoliatus* UC5319. Consistent with these observations, recombinant PtlA purified from *E. coli* catalyzed the Mg^{2+} -dependent cyclization of FPP to pentalenene (Tetzlaff et al., 2006).

The gene encoding geosmin synthase in *S. coelicolor* A3(2) was independently characterized by two groups (Cane & Watt, 2003, Gust, Challis, Fowler, Kieser, & Chater, 2003). The 2181-bp *sco6073* encodes an unusually large 726-aa protein in which both the N-terminal (366 aa) and C-terminal (339 aa) halves showed about 30% identity to *S. exfoliatus* UC5319 pentalenene synthase. Similarly, the 2178-bp *sav2163* of *S. avermitilis* encodes a 725-aa protein with 77% identity to the *S. coelicolor* A3(2) geosmin synthase (SCO6073) and 27% identity in both its N-terminal (363 aa) and C-terminal (329 aa) halves to pentalenene synthase (Cane, He, Kobayashi, Omura, & Ikeda, 2006). The N-terminal domains of both synthases possess the characteristic conserved metal-binding domains, the acidic amino acid-rich motif **-DDHFLE-** and the NSE triad **-NDLFSYQRE-**, while the C-terminal halves contain an unusual acidic amino acid-rich motif, **-DDYYP-**, as well as a canonical **-NDVFSYQKE-** sequence in both synthases. The N-terminal domain has been shown to catalyze cyclization of FPP to generate germacradienol and germacrene D, while the C-terminal domain is responsible for the proton-initiated sequential retro-Prins fragmentation of germacradienol with loss of the 2-propanol side-chain as acetone to give an octalin derivative, which is then converted to geosmin by proton-initiated hydration and hydride migration (Jiang & Cane, 2008, Nawrath et al., 2008). The sequence of geosmin synthase is in fact highly conserved across a wide variety of bacterial genera, with more than 75 predicted proteins from actinomycetes, myxobacteria, and cyanobacteria exhibiting >55% sequence identity to the *S. coelicolor* A3(2) SCO6073.

Two genes, *S. coelicolor* A3(2) *sco5222* and *S. avermitilis* *sav3032*, encode orthologous 361-aa and 363-aa proteins with 24% and 25% identity, respectively, to *S. exfoliatus* pentalenene synthase. Both deduced proteins, SCO5222 and SAV_3032, possess two common conserved metal-binding motifs, acidic amino acid-rich **-DDRHD-** and the downstream NSE triad **-NDLCSLPKE-**. The two corresponding recombinant proteins each catalyze cyclization of FPP to generate a tricyclic sesquiterpene hydrocarbon identified as (+)-*epi*-isozizaene (Lin, Hopson, & Cane, 2006, Takamatsu et al., 2011). The *sco5222* and *sav3032* genes are each translationally coupled with the downstream genes, *sco5223* and *sav3031*, encoding the cytochrome P450s CYP170A1 and CYP170A2, respectively. The two-gene operon for albaflavenone biosynthesis is highly conserved and

widely distributed, with orthologous pairs of translationally coupled sesquiterpene synthase/cytochrome P450 genes with 56–100% identity evident in 10 species of streptomycetes; production of *epi*-isozizaene and/or albaflavenone has been shown in several of these streptomycetes (Moody et al., 2012; Takamatsu et al., 2011).

Epicubenol synthase was originally isolated by biochemical assay-guided methods from *Streptomyces* sp. LL-B7 that was known to produce cadin-4-en-1-ol (Gerber, 1971). The partially purified sesquiterpene synthase catalyzes the Mg^{2+} -dependent cyclization of FPP to (+)-epicubenol (Cane, Tandon, & Prabhakaran, 1993).

2.3. Diterpene synthases

Diterpene synthases, which have been cloned predominantly from higher plants and fungi, have been reviewed (MacMillan & Beale, 1999). They can be classified into at least four groups based on the combination of two fundamental types of cyclization reaction: (i) type-A, (ii) type-B, (iii) type-A-type-B, and (iv) type-B-type-A. The cyclization reaction catalyzed by a type-A synthase is initiated by ionization of the diphosphate of GGPP to give an allylic cation, followed by cyclization and deprotonation to generate an olefinic diterpene hydrocarbon or capture of a water molecule to generate a diterpene alcohol. The type-B reaction is initiated by protonation at the C14,15-double bond of GGPP distal to the allylic diphosphate.

Bacterial genes encoding diterpene synthases were first recognized in connection with the observation that, whereas most actinomycetes lack the mevalonate (MVA) pathway and utilize the methylerythritol phosphate pathway to provide DMAPP and IPP, the key precursors for terpenoid metabolism are present in certain strains of *Streptomyces*. Genes encoding well-characterized MVA pathway enzymes are found in such organisms. One such strain, *Kitasatospora griseola* MF730-N6, was found to produce the diterpenoid metabolite, terpentecin (Dairi et al., 2001). The gene cluster for the MVA pathway lies downstream of a set of five genes required for terpentecin biosynthesis. Since two genes, *orf11* and *orf12*, encode a 499-aa protein with 29% identity to *ent*-kaurene synthase of *Phaeosphaeria* sp. and a 311-aa protein with 25% identity to *S. exofolius* pentalenene synthase, ORF11 and ORF12 correspond to type-B and type-A synthases, respectively. The deduced type-A ORF12 protein displays the universally conserved pair of divalent metal-binding motifs, an acidic amino acid-rich -DDRWD-, and the downstream triad -NDYYSWGRE-. Incubation of

ORF11, ORF12, and GGPP in the presence of magnesium ions produced the diterpene hydrocarbon, terpentetriene (Hamano et al., 2002).

A second bacterial diterpene synthase is involved in the biosynthesis of viguiepinol in *Streptomyces* sp. KO-3988. The five genes implicated in its biosynthesis lie upstream of the gene cluster for the MVA pathway (Kawasaki et al., 2004) and transformants of *Streptomyces lividans* carrying the five genes produced the diterpenoid metabolite, viguiepinol (Kawasaki et al., 2006). Two of these genes, *orf2* and *orf3*, encode a 511-aa protein with 32% identity to *ent*-copalyl diphosphate synthase of the fungus *Gibberella fujikuroi* and a 295-aa protein with no significant homology to any known terpene synthase, respectively. Recombinant ORF2 catalyzes the Mg^{2+} -dependent cyclization of GGPP to generate copalyl diphosphate, corresponding to a type-B synthase, while the type-A synthase ORF3 converts copalyl diphosphate to pimara-9(11),15-diene. Two conserved metal-binding motifs, an acidic amino acid-rich -DDHVE- and the downstream triad -NDLATFERE-, are present in the deduced ORF3 protein.

The cyclooctatin biosynthetic gene cluster of *Streptomyces melanosporofaciens* MI614-43F2 consists of four genes, *cotB1*, *cotB2*, *cotB3*, and *cotB4*, encoding GGPP synthase, a type-A diterpene synthase, and two cytochrome P450s, respectively, but no type-B terpene synthase. The *cotB2* gene encodes a 307-aa protein, which shows no significant sequence similarity to any of the known terpene synthases, displays only one of the conserved metal-binding motifs, the triad sequence -NDFYSYDRE-, found in the C-terminal region, but no identifiable upstream acidic amino acid-rich motif. Recombinant CotB2 catalyzes the Mg^{2+} -dependent cyclization of GGPP to the diterpene alcohol, cyclooctat-9-en-7-ol (Kim et al., 2011).



3. METHODS FOR THE STUDY OF BACTERIAL TERPENE SYNTHASES

3.1. Bioinformatic analysis of bacterial terpene synthases

3.1.1 Properties of terpene synthases and database search procedures

Monoterpene, sesquiterpene, and diterpene synthases from plants and fungi have a strongly conserved amino acid sequence homology. Conversely, bacterial terpene synthases not only exhibit no significant overall sequence amino acid similarity to those from plants and fungi but also usually exhibit relatively low mutual sequence similarity to each other. Despite the very substantial differences in overall primary amino acid sequence, terpene synthases from plants, fungi, and bacteria typically display two highly

conserved metal-binding motifs: an acidic amino acid-rich motif, **-DDxx[DE]-** or **-DDxxx[DE]-**, usually located either 80–120 or 320–360 amino acids downstream of the N-termini of bacterial terpene synthases or the larger eukaryotic terpene synthases, respectively, and a triad of residues, **-[ND]Dxx[ST]xx[KR][DE]-**, located 140 ± 5 amino acids downstream of the acidic amino acid-rich motif (Felicetti & Cane, 2004). The first conserved metal-binding motif in a typical bacterial sesquiterpene synthase is the acidic amino acid-rich domain with a high proportion of aromatic amino acids, **-WFF[VL][FW]DD[LR][FH]D-** (pentalenene and *epi*-isozizaene synthases) or **-WVF[FY]FDDHFLE-** (germacradienol/geosmin synthase). Although the upstream conserved motifs in diterpene synthases (**-LIVNDDRWD-**) and in the monoterpene synthases (**-AVDDxxx[DE]-**) also display an acidic amino acid-rich domain, the content of aromatic amino acids is lower than that in sesquiterpene synthases. The downstream motif, **-ND[IL]xSxx[KR]E-**, is conserved in all three classes of bacterial terpene synthases (Komatsu et al., 2008). Crystallographic analysis of monoterpene and sesquiterpene synthases from bacteria, fungi, and plants have established that these two metal-binding motifs lie at opposite sides of the rim of the deep active-site cavity and are responsible for cooperative binding of three divalent cations and the diphosphate moiety of the substrate (GPP or FPP), precisely positioning the acyclic allylic diphosphate substrate and activating it for the ionization that triggers the cyclization cascade (Christianson, 2006, 2008).

In general, new proteins with novel functions can be recognized by application of the BLAST algorithm for local sequence alignment to deduced protein sequences in the public databases. In this manner, such BLAST searches of bacterial databases using known bacterial terpene synthases as the query have in fact revealed a significant number of presumptive synthases of unknown function. On the other hand, this search strategy may frequently miss cases of low overall sequence similarity. In spite of low overall primary sequence similarity, significant conserved metal-binding motifs such as the acidic amino acid-rich domain and the triad domain have been found in essentially all known bacterial, fungal, and plant terpene synthases. As an alternative to the widely used method of local sequence alignment, therefore, we have adopted a search method based on hidden Markov models (HMMs) and Pfam search (Finn et al., 2010) for the primary recognition of desired terpene synthases. The profile HMM can distinguish members of the relevant protein functional families from nonmembers with a high degree of accuracy (Sonnhammer, Eddy, Birney, Bateman, & Durbin, 1998).

The first successful trial using a profile HMM was demonstrated in the bioinformatics-based discovery of 2-methylisoborneol synthase from bacterial databases (Komatsu et al., 2008). Presumptive bacterial terpene synthases were first harvested from the 2008 NCBI databases of bacterial proteins by searching on the basis of the profile HMM using as a model PF03936 (terpene synthase family, metal-binding domain). From 1,922,990 predicted proteins, 41 proteins were initially selected based on strong alignment matches and very low *E*-values. These proteins apparently segregated into three major groups on the basis of phylogenetic analysis. Group I containing 12 proteins was provisionally assigned as monoterpene synthases, subsequently shown to be 2-methylisoborneol synthases. Group II, by far the largest group consisting of 27 proteins, included the known sesquiterpene synthases, pentalenene, germacradienol/geosmin, and *epi*-isozizaene synthases as well as several presumptive sesquiterpene synthases of unknown biochemical function. Group III contained two diterpene synthases, including terpentetriene synthases. Since the initial HMM analysis, among the presumptive sesquiterpene synthases SAV_76 of *S. avermitilis* has been characterized as a new functional protein, avermitilol synthase (Chou et al., 2010) and SGR_2079 of *Streptomyces griseus* IFO 13350 has been shown to be (+)-caryolan-1-ol synthase (Nakano, Horinouchi, & Ohnishi, 2011). Searching for an orthologous 2-methylisoborneol synthase in the genome sequence of the 2-methylisoborneol-producing cyanobacterium *Pseudanabaena limnetica* str. Castaic Lake turned out to be initially unsuccessful. Since the original PF03936 profile had been based on an alignment of terpene synthase sequences that included those from both plants and fungi, a new set of HMM parameters was generated using as training set of the 41 bacterial terpene synthase sequences that had been identified by the first trial HMM experiment (Komatsu et al., 2008), resulting in the effective recognition of the cyanobacterial 2-methylisoborneol synthase using the new profile of HMM (Giglio, Chou, Ikeda, Cane, & Monis, 2011).

3.1.2 Genome mining of bacterial terpene synthases

To date, more than 2700 bacterial genome sequences have been completed (<http://www.genomesonline.org/>). The newest bacterial protein sequence databases have been processed by HMM Pfam search using the above-described second-generation model specific for bacterial terpene synthases. A phylogenetic tree of the harvested terpene synthase sequences illustrating the grouping of over 140 candidate bacterial sequences is shown in Fig. 7.2. The majority of the presumptive terpene synthases are sesquiterpene

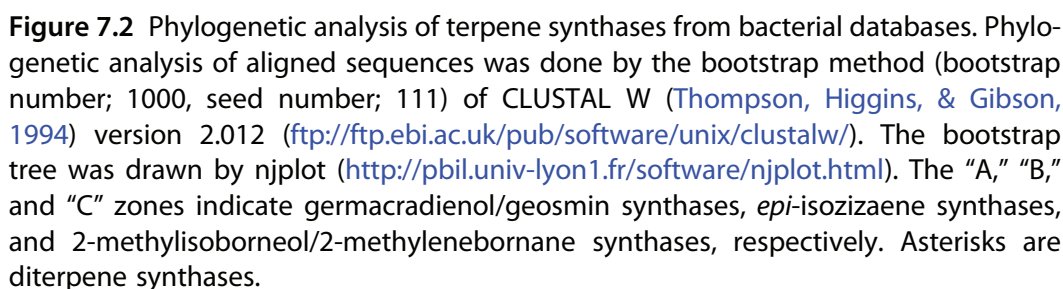


Table 7.1 (type-A synthase) and Table 7.2 (type-B synthase) summarize presumptive terpene synthases from the newest bacterial genome databases.

Table 7.1 Predicted type-A terpene synthases from genome-sequenced bacteria

Microorganism	Class or phylum	Order	Type of synthase ^a	Locus tag or protein ID	Accession number	E-value ^b
<i>Actinosynnema mirum</i> DSM 43827	<i>Actinobacteria</i>	<i>Actinomycetales</i>	Germacradienol/ geosmin synthase	Amir_0977	YP_003098781	9.80×10^{-231}
				Amir_3801	YP_003101527	2.40×10^{-9}
<i>Amycolatopsis mediterranei</i> U32	<i>Actinobacteria</i>	<i>Actinomycetales</i>	Germacradienol/ geosmin synthase	AMED_1325	YP_003763541	3.50×10^{-234}
				AMED_3240	YP_003765432	4.70×10^{-63}
<i>Catenulispora acidiphila</i> DSM 44928	<i>Actinobacteria</i>	<i>Actinomycetales</i>	Germacradienol/ geosmin synthase	Caci_6200	YP_003116895	9.90×10^{-228}
				Caci_4612	YP_003115314	4.10×10^{-66}
				Caci_3530	YP_003114277	1.40×10^{-7}
<i>Frankia alni</i> ACN14a	<i>Actinobacteria</i>	<i>Actinomycetales</i>	Germacradienol/ geosmin synthase	FRAAL6507	YP_716636	3.20×10^{-251}
				FRAAL1336	YP_711586	3.50×10^{-73}
<i>Frankia</i> sp. CcI3	<i>Actinobacteria</i>	<i>Actinomycetales</i>	Germacradienol/ geosmin synthase	Franci3_4231	YP_483306	9.70×10^{-251}
				Franci3_4335	YP_483410	3.40×10^{-21}
<i>Frankia</i> sp. EAN1pec	<i>Actinobacteria</i>	<i>Actinomycetales</i>	Germacradienol/ geosmin synthase	Franean1_5559	YP_001509819	2.60×10^{-251}

Continued

Table 7.1 Predicted type-A terpene synthases from genome-sequenced bacteria—cont'd

Microorganism	Class or phylum	Order	Type of synthase	Locus tag or protein ID	Accession number	E-value
<i>Kitasatospora setae</i> KM-6054	<i>Actinobacteria</i>	<i>Actinomycetales</i>	Germacradienol/ geosmin synthase	KSE_46080	YP_004906345	1.40×10^{-245}
			2-MIB/2-MB synthase ^c	KSE_70210	YP_004908735	2.00×10^{-69}
				KSE_12950	YP_004903082	6.90×10^{-58}
				KSE_00200t	YP_004901829	1.30×10^{-57}
				KSE_62070	YP_004907928	2.50×10^{-24}
<i>Kribbella flavida</i> DSM 17836	<i>Actinobacteria</i>	<i>Actinomycetales</i>	Germacradienol/ geosmin synthase	Kfla_4247	YP_003382082	2.60×10^{-251}
<i>Kutzneria</i> sp. 744	<i>Actinobacteria</i>	<i>Actinomycetales</i>	Germacradienol/ geosmin synthase	KUTG_04531	–	3.10×10^{-216}
			2-MIB/2-MB synthase ^c	KUTG_08607	–	1.40×10^{-48}
				KUTG_08053	–	7.80×10^{-32}
<i>Micromonospora carbonacea</i> ATCC 39149	<i>Actinobacteria</i>	<i>Actinomycetales</i>		MCAG_03139	ZP_04606882	1.40×10^{-68}
			2-MIB/2-MB synthase ^c	MCAG_05469	ZP_04609212	2.30×10^{-64}
				MCAG_01060	ZP_04604803	4.60×10^{-13}
<i>Micromonospora</i> sp. M42	<i>Actinobacteria</i>	<i>Actinomycetales</i>		MCBG_03612	–	4.30×10^{-63}
				MCBG_05692	–	3.10×10^{-56}
<i>Nocardopsis dassonvillei</i> subsp. <i>dassonvillei</i> DSM 43111	<i>Actinobacteria</i>	<i>Actinomycetales</i>	2-MIB/2-MB synthase ^c	Ndas_2620	YP_003680543	7.20×10^{-64}

<i>Rubrobacter xylanophilus</i> DSM 9941	<i>Actinobacteria</i>	<i>Rubrobacterales</i>	RxyL_0493	YP_643279	1.90×10^{-80}
<i>Saccharopolyspora erythraea</i> NRRL 2338	<i>Actinobacteria</i>	<i>Actinomycetales</i>	SACE_3187	YP_001105388	1.10×10^{-244}
		Germacradienol/ geosmin synthase			
		Germacradienol/ geosmin synthase	SACE_4907	YP_001107098	1.80×10^{-240}
		Germacradienol/ geosmin synthase	SACE_3977	YP_001106173	1.80×10^{-202}
		2-MIB synthase ^{c,d1}	SACE_3722	YP_001105919	1.20×10^{-52}
<i>Salinispora arenicola</i> CNS-205	<i>Actinobacteria</i>	<i>Actinomycetales</i>	Sare_1287	YP_001536181	4.10×10^{-16}
<i>Stactebrandtia nassauensis</i> DSM 44728	<i>Actinobacteria</i>	<i>Actinomycetales</i>	Snas_1127	YP_003509930	4.70×10^{-243}
		Germacradienol/ geosmin synthase			
		2-MIB/2-MB synthase	Snas_1991	YP_003510780	2.20×10^{-65}
<i>Streptomyces albus</i> J1074	<i>Actinobacteria</i>	<i>Actinomycetales</i>	SSHG_04630	ZP_06593727	2.10×10^{-245}
		Germacradienol/ geosmin synthase			
		<i>epi</i> -Isoizaene synthase	SSHG_04343	ZP_06593440	9.70×10^{-78}

Continued

Table 7.1 Predicted type-A terpene synthases from genome-sequenced bacteria—cont'd

Microorganism	Class or phylum	Order	Type of synthase	Locus tag or protein ID	Accession number	E-value
<i>Streptomyces avermitilis</i> MA-4680	<i>Actinobacteria</i>	<i>Actinomycetales</i>	Germacradienol/ geosmin synthase ^{d2}	SAV_2163	NP_823339	4.30×10^{-255}
			<i>epi</i> -Isoizaene synthase ^{d3}	SAV_3032	NP_824208	3.60×10^{-81}
			Pentalenene synthase ^{d4}	SAV_2998	NP_824174	2.90×10^{-77}
			Avermitilol synthase ^{d5}	SAV_76	NP_821250	4.70×10^{-75}
<i>Streptomyces bingchengensis</i> BCW-1	<i>Actinobacteria</i>	<i>Actinomycetales</i>	Germacradienol/ geosmin synthase	SBI_02068	ADI05189	9.20×10^{-234}
			2-MIB/2-MB synthase ^c	SBI_08957	ADI12075	5.50×10^{-75}
			Pentalenene synthase	SBI_09679	ADI12797	1.90×10^{-71}
				SBI_01080	ADI04201	7.50×10^{-66}
<i>Streptomyces cattleya</i> NRRL 8005	<i>Actinobacteria</i>	<i>Actinomycetales</i>	Germacradienol/ geosmin synthase	SCAT_2953	YP_004912463	1.00×10^{-250}
				SCAT_p0091	YP_004919385	4.30×10^{-36}

<i>Streptomyces clavuligerus</i> ATCC 27064	<i>Actinobacteria</i>	<i>Actinomycetales</i>	Germacradienol/ geosmin synthase	SCLAV_0159	ZP_06913794	1.60×10^{-237}
				SCLAV_p0635	ZP_05005402	4.00×10^{-82}
			1,8-Cineole synthase ^{d6}	SCLAV_p0982	ZP_05003209	7.00×10^{-78}
				SCLAV_p1407	ZP_06776581	6.70×10^{-77}
			(-)- δ -Cadinene synthase ^{d7}	SCLAV_p0328	ZP_05004823	1.40×10^{-75}
			(+)-T-Muurolol synthase ^{d7}	SCLAV_p0068	ZP_05006242	6.10×10^{-70}
			Linalool synthase ^{d8}	SCLAV_p1185	ZP_06776363	1.70×10^{-69}
				SCLAV_p0985	ZP_06776164	2.70×10^{-62}
				SCLAV_p0574	ZP_06775755	1.40×10^{-55}
				SCLAV_p0571	ZP_06775752	4.90×10^{-48}
				SCLAV_p1173	ZP_06776351	9.10×10^{-42}
				SCLAV_p0491	ZP_05004575	4.40×10^{-41}
				SCLAV_p0765	ZP_06775944	7.30×10^{-29}
				SCLAV_p1429	ZP_05002948	1.30×10^{-20}
				SCLAV_p1169	ZP_05007982	2.30×10^{-10}
				SCLAV_5601	ZP_06775069	6.80×10^{-8}
				SCLAV_5671	ZP_06775139	1.50×10^{-2}

Continued

Table 7.1 Predicted type-A terpene synthases from genome-sequenced bacteria—cont'd

Microorganism	Class or phylum	Order	Type of synthase	Locus tag or protein ID	Accession number	E-value
<i>Streptomyces coelicolor</i> A3(2)	<i>Actinobacteria</i>	<i>Actinomycetales</i>	Germacradienol/ geosmin synthase ^{d9}	SCO6073	NP_630182	4.80×10^{-249}
			<i>epi</i> -Isozizaene synthase ^{d10}	SCO5222	NP_629369	8.20×10^{-83}
			2-MIB synthase ^{c,d11}	SCO7700	NP_733742	1.60×10^{-67}
			Germacradienol/ geosmin synthase	Sfla_6028	YP_004926931	4.60×10^{-253}
<i>Streptomyces flavogriseus</i> ATCC 33331	<i>Actinobacteria</i>	<i>Actinomycetales</i>	Germacradienol/ geosmin synthase	Sfla_5399	YP_004926313	3.00×10^{-77}
				Sfla_4535	YP_004925455	8.80×10^{-71}
			2-MIB/2-MB synthase ^c	Sfla_5667	YP_004926578	4.20×10^{-70}
				Sfla_1617	YP_004922572	1.10×10^{-62}
<i>Streptomyces ghanaensis</i> ATCC 14672	<i>Actinobacteria</i>	<i>Actinomycetales</i>	Germacradienol/ geosmin synthase	SSFG_01626	ZP_06575915	2.80×10^{-249}
			<i>epi</i> -Isozizaene synthase	SSFG_02456	ZP_06576746	3.00×10^{-84}
			Germacradienol/ geosmin synthase	SSRG_01130	ZP_07309957	1.80×10^{-251}
<i>Streptomyces griseoflavus</i> Tü4000	<i>Actinobacteria</i>	<i>Actinomycetales</i>	<i>epi</i> -Isozizaene synthase	SSRG_02017	ZP_07310844	8.80×10^{-82}

<i>Streptomyces griseus</i> IFO 13350	<i>Actinobacteria</i>	<i>Actinomycetales</i>	Germacradienol/ geosmin synthase	SGR_6839	YP_001828351	3.00×10^{-258}
				SGR_6065	YP_001827577	9.20×10^{-80}
			2-MIB synthase ^c	SGR_1269	YP_001822781	1.30×10^{-74}
			(+)-Caryolan-1-ol synthase ^{d12}	SGR_2079	YP_001823591	4.60×10^{-72}
<i>Streptomyces hygroscopicus</i> ATCC 53653	<i>Actinobacteria</i>	<i>Actinomycetales</i>	Germacradienol/ geosmin synthase	SSOG_01998	ZP_07293917	2.60×10^{-264}
			Germacradienol/ geosmin synthase	SSOG_04571	ZP_07296488	2.50×10^{-257}
			2-MIB/2-MB synthase ^c	SSOG_08875	ZP_07300792	6.20×10^{-76}
				SSOG_01163	ZP_07293082	2.80×10^{-69}
<i>Streptomyces lividans</i> TK24	<i>Actinobacteria</i>	<i>Actinomycetales</i>	Germacradienol/ geosmin synthase	SSPG_01553	ZP_05522837	4.90×10^{-249}
			<i>epi</i> -Isoizaene synthase	SSPG_02461	ZP_06528571	7.80×10^{-83}
				SSPG_07524	ZP_06533634	3.30×10^{-69}
			2-MIB/2-MB synthase ^c	SSPG_00143	ZP_06526253	1.20×10^{-57}
<i>Continued</i>						

Table 7.1 Predicted type-A terpene synthases from genome-sequenced bacteria—cont'd

Microorganism	Class or phylum	Order	Type of synthase	Locus tag or protein ID	Accession number	E-value
<i>Streptomyces pristinaespiralis</i> ATCC 25486	<i>Actinobacteria</i>	<i>Actinomycetales</i>	Germacradienol/ geosmin synthase	SSDG_03228	ZP_06913794	1.70×10^{-252}
				SSDG_01120	ZP_06913376	2.30×10^{-85}
				SSDG_02809	ZP_06911744	1.70×10^{-50}
<i>Streptomyces roseosporus</i> NRRL 15998	<i>Actinobacteria</i>	<i>Actinomycetales</i>	Caryolan-1-ol synthase	SSGG_00557	ZP_06582730	2.60×10^{-82}
				SSGG_05086	ZP_06587258	1.50×10^{-71}
<i>Streptomyces scabiei</i> 87.22	Actinobacteria	Actinomycetales	Germacradienol/ geosmin synthase	SCAB20121	YP_003487693	1.40×10^{-249}
			2-MIB/2-MB synthase ^c	SCAB5041	YP_003486275	2.30×10^{-68}
				SCAB82161	YP_003493696	3.40×10^{-54}
				SCAB73741	YP_003492893	4.70×10^{-54}
<i>Streptomyces svicens</i> ATCC 29083	<i>Actinobacteria</i>	<i>Actinomycetales</i>	Germacradienol/ geosmin synthase	SSEG_08185	ZP_06920565	8.20×10^{-255}
				SSEG_08483	ZP_06919672	2.80×10^{-81}
<i>Streptomyces venezuelae</i> ATCC 10712	<i>Actinobacteria</i>	<i>Actinomycetales</i>	Germacradienol/ geosmin synthase	SVEN_0269	CCA53556	6.10×10^{-236}
				SVEN_7111	CCA60397	1.70×10^{-71}
				SVEN_0552	CCA53839	4.90×10^{-59}

<i>Streptomyces violaceusniger</i> Tü4113	<i>Actinobacteria</i>	<i>Actinomycetales</i>	Germacradienol/ geosmin synthase	Strvi_9069	YP_004818644	2.50×10^{-252}
			2-MIB/2-MB synthase ^c	Strvi_5648	YP_004815441	7.70×10^{-74}
				Strvi_5748	YP_004815539	3.70×10^{-59}
			Cyclooctat-9-en-7-ol synthase	Strvi_5779	YP_004815569	1.50×10^{-5}
<i>Streptomyces viridochromogenes</i> DSM 40736	<i>Actinobacteria</i>	<i>Actinomycetales</i>	Germacradienol/ geosmin synthase	SSQG_06187	ZP_07307300	6.90×10^{-249}
			epi-Isoizaene synthase	SSQG_05270	ZP_07306383	1.70×10^{-84}
				SSQG_07226	ZP_07308339	5.70×10^{-71}
			2-MIB/2-MB synthase ^c	SSQG_00350	ZP_07301463	1.60×10^{-53}
<i>Streptomyces</i> sp. AA4	<i>Actinobacteria</i>	<i>Actinomycetales</i>		SSQG_00965	ZP_07302078	1.10×10^{-27}
			Germacradienol/ geosmin synthase	SSMG_01007	ZP_07276967	4.20×10^{-220}
			2-MIB/2-MB synthase ^c	SSMG_03159	ZP_07279119	1.10×10^{-68}
<i>Streptomyces</i> sp. C	<i>Actinobacteria</i>	<i>Actinomycetales</i>	Germacradienol/ geosmin synthase	SSNG_05608	ZP_07289987	5.50×10^{-230}
			2-MIB/2-MB synthase ^c	SSNG_06212	ZP_07290591	5.40×10^{-69}
				SSNG_07281	ZP_07291660	1.30×10^{-24}

Continued

Table 7.1 Predicted type-A terpene synthases from genome-sequenced bacteria—cont'd

Microorganism	Class or phylum	Order	Type of synthase	Locus tag or protein ID	Accession number	E-value
<i>Streptomyces</i> sp. E14	<i>Actinobacteria</i>	<i>Actinomycetales</i>	<i>epi</i> -Isoizaene synthase	SSTG_03324	ZP_06709883	2.00×10^{-77}
				SSTG_04057	ZP_06710616	2.40×10^{-56}
<i>Streptomyces</i> sp. Mg1	<i>Actinobacteria</i>	<i>Actinomycetales</i>	Germacradienol/ geosmin synthase	SSAG_06177	ZP_05001875	7.50×10^{-234}
				SSAG_00457	ZP_04996155	6.60×10^{-75}
				SSAG_06488	ZP_05002186	5.00×10^{-72}
				SSAG_05118	ZP_05000816	2.90×10^{-69}
				SSAG_06899	ZP_05002597	1.70×10^{-56}
<i>Streptomyces</i> sp. SirexAA	<i>Actinobacteria</i>	<i>Actinomycetales</i>	Germacradienol/ geosmin synthase	SSAG_00644	ZP_04996342	3.70×10^{-51}
				SACTE_0245	ZP_06273105	6.80×10^{-257}
				SACTE_0873	ZP_06275238	4.40×10^{-81}
				SACTE_4655	ZP_06270281	1.10×10^{-66}
				SACTE_5791	ZP_06273747	2.50×10^{-26}
<i>Streptomyces</i> sp. SPB74	<i>Actinobacteria</i>	<i>Actinomycetales</i>	<i>epi</i> -Isoizaene synthase	SSBG_02327	ZP_06823519	4.30×10^{-78}

<i>Streptomyces</i> sp. SPB78	<i>Actinobacteria</i>	<i>Actinomycetales</i>	Germacradienol/ geosmin synthase	SSLG_03862	ZP_07273435	2.90×10^{-247}
			<i>epi</i> -Isoizaene synthase	SSLG_04610	ZP_07274183	1.00×10^{-81}
<i>Streptosporangium</i> <i>roseum</i> DSM 43021	<i>Actinobacteria</i>	<i>Actinomycetales</i>	Germacradienol/ geosmin synthase	Sros_0075	YP_00333587	2.40×10^{-242}
				Sros_6866	YP_003342315	1.20×10^{-79}
				Sros_2608	YP_003338321	4.30×10^{-69}
				Sros_3191	YP_003338879	8.60×10^{-15}
				Sros_2610	YP_003338323	5.90×10^{-5}
<i>Thermomonospora</i> <i>curvata</i> DSM 43183	<i>Actinobacteria</i>	<i>Actinomycetales</i>		Tcur_3107	YP_003300687	9.50×10^{-60}
<i>Anabaena variabilis</i> ATCC 29413	<i>Cyanobacteria</i>	<i>Nostocaceae</i>		Ava_1982	YP_322499	2.70×10^{-70}
<i>Burkholderia</i> <i>pseudomallei</i> 668	<i>Betaproteobacteria</i>	<i>Burkholderiales</i>		BURPS668_A0947	YP_001061946	2.30×10^{-69}
				BURPS668_A1721	YP_001062716	6.70×10^{-40}
<i>Chitinophaga</i> <i>pinensis</i> DSM 2588	<i>Sphingobacteriia</i>	<i>Sphingobacteriales</i>		Cpin_4730	YP_003124367	5.60×10^{-70}
				Cpin_3949	YP_003123612	1.70×10^{-57}
				Cpin_1797	YP_003121494	6.00×10^{-44}
				Cpin_4100	YP_003123761	8.50×10^{-42}
<i>Haliangium</i> <i>odhraceum</i> DSM 14365	<i>Deltaproteobacteria</i>	<i>Myxococcales</i>	Germacradienol/ geosmin synthase	Hoch_1257	YP_003265710	6.80×10^{-219}

Continued

Table 7.1 Predicted type-A terpene synthases from genome-sequenced bacteria—cont'd

Microorganism	Class or phylum	Order	Type of synthase	Locus tag or protein ID	Accession number	E-value
<i>Herpetosiphon aurantiacus</i> DSM 785	<i>Chloroflexi</i>	<i>Herpetosiphonales</i>		Haur_2988	YP_001545754	2.60×10^{-77}
				Haur_2987	YP_001545753	4.70×10^{-70}
				Haur_4149	YP_001546909	3.70×10^{-19}
<i>Nostoc</i> sp. PCC 7120	<i>Cyanobacteria</i>	<i>Nostocaceae</i>	Germacrene A synthase ^{d13}	aln4685	NP_488725	1.70×10^{-71}
<i>Nostoc punctiforme</i> PCC 73102	<i>Cyanobacteria</i>	<i>Nostocaceae</i>	Germacradienol/ geosmin synthase ^{d14}	Npun_R2756	YP_001866236	4.80×10^{-242}
			8a- <i>epi</i> - α -Selinene synthase ^{d13}	Npun_R3832	YP_001867159	2.20×10^{-70}
<i>Myxococcus fulvus</i> HW-1	<i>Deltaproteobacteria</i>	<i>Myxococcales</i>	Germacradienol/ geosmin synthase	LILAB_15255	YP_004666032	1.80×10^{-243}
<i>Myxococcus xanthus</i> DK1622	<i>Deltaproteobacteria</i>	<i>Myxococcales</i>	Germacradienol/ geosmin synthase	MXAN_6247	YP_634376	3.50×10^{-244}
<i>Oscillatoria</i> sp. PCC 6506	<i>Cyanobacteria</i>	<i>Oscillatoriales</i>	Germacradienol/ geosmin synthase	OSCI_4110022	ZP_07114089	2.10×10^{-243}
<i>Pseudomonas fluorescens</i> Pf0-1	<i>Gammaproteobacteria</i>	<i>Pseudomonadales</i>	2-MB synthase ^{c,d15}	Pf01_1841	YP_347573	7.60×10^{-61}
<i>Roseiflexus castenholzii</i> DSM 13941	<i>Chloroflexi</i>	<i>Chloroflexales</i>		Rcas_0622	YP_001430766	2.20×10^{-89}
				Rcas_2716	YP_001432805	1.80×10^{-33}

<i>Roseiflexus</i> sp. RS-1	<i>Chloroflexi</i>	<i>Chloroflexales</i>	RoseRS_3509	YP_001277817	6.70×10^{-87}
			RoseRS_2648	YP_001276972	3.50×10^{-34}
<i>Sorangium</i> <i>cellulosum</i> 'So ce 56'	<i>Deltaproteobacteria</i>	<i>Myxococcales</i>	Sce1440	YP_001612078	1.10×10^{-238}
		Germacradienol/ geosmin synthase	Sce6369	YP_001617016	2.30×10^{-74}
			Sce8552	YP_001619202	3.40×10^{-74}
<i>Stigmatella</i> <i>aurantiaca</i> DW4/ 3-1	<i>Deltaproteobacteria</i>	<i>Myxococcales</i>	STAUR_1114	YP_003950745	1.1×10^{-240}
<i>Waddlia</i> <i>chondrophila</i> WSU 86-1044	<i>Chlamydia</i>	<i>Chlamydiales</i>	Wcw_0522	YP_003708899	4.00×10^{-32}
<i>Kitasatospora</i> <i>griseola</i> MF730- N6 ^e	<i>Actinobacteria</i>	<i>Actinomycetales</i>	BAB39207	AB048795	1.50×10^{-45}
		Terpentertriene synthase ^{d16}			
<i>Micromonospora</i> <i>olivasterospora</i> NRRL 8178 ^e	<i>Actinobacteria</i>	<i>Actinomycetales</i>	BAK26793	AB634561	1.70×10^{-66}
		2-MB synthase ^{c,d1}			
<i>Streptomyces</i> <i>ambofaciens</i> ISP 5053 ^c	<i>Actinobacteria</i>	<i>Actinomycetales</i>	SAML0357	CAJ89344	3.60×10^{-73}
		2-MB synthase ^{c,d1}			

Continued

Table 7.1 Predicted type-A terpene synthases from genome-sequenced bacteria—cont'd

Microorganism	Class or phylum	Order	Type of synthase	Locus tag or protein ID	Accession number	E-value
<i>Streptomyces citricolor</i> NBRC 13005 ^c	<i>Actinobacteria</i>	<i>Actinomycetales</i>	(-)-Germacradiene-4-ol synthase ^{d17}	BAL14866	AB621338	1.70×10^{-64}
<i>Streptomyces exfoliatus</i> UC5319 ^e	<i>Actinobacteria</i>	<i>Actinomycetales</i>	(-)- <i>epi</i> - α -Bisabolol synthase ^{d17}	BAL14867	AB621339	3.20×10^{-53}
<i>Streptomyces lasaliensis</i> NRRL 3382 ^e	<i>Actinobacteria</i>	<i>Actinomycetales</i>	Pentalenene synthase ^{d18}	PTLS_STRS3	Q55012	2.50×10^{-81}
<i>Streptomyces melanosporofaciens</i> MI614-43F2 ^e	<i>Actinobacteria</i>	<i>Actinomycetales</i>	2-MIB synthase ^{c,d1}	BAI77523	AB547324	3.90×10^{-70}
<i>Streptomyces platensis</i> MA7327 ^c	<i>Actinobacteria</i>	<i>Actinomycetales</i>	Cyclooctat-9-ene-7-ol synthase ^{d19}	BAI44338	AB183750	2.20×10^{-5}
<i>Streptomyces platensis</i> MA7327 ^c	<i>Actinobacteria</i>	<i>Actinomycetales</i>	<i>ent</i> -Kaurene synthase ^{d20}	ACO31279	FJ655920	1.10×10^{-1}
<i>Streptomyces platensis</i> MA7339 ^c	<i>Actinobacteria</i>	<i>Actinomycetales</i>	<i>ent</i> -Atiserene synthase ^{d20}	ACO31274	FJ655920	5.20×10^{-1}
<i>Streptomyces platensis</i> MA7339 ^c	<i>Actinobacteria</i>	<i>Actinomycetales</i>	<i>ent</i> -Atiserene synthase ^{d20}	ADD83014	GQ398492	5.20×10^{-1}
<i>Streptomyces</i> sp. KO-3988 ^e	<i>Actinobacteria</i>	<i>Actinomycetales</i>	Pimaradiene synthase ^{d21}	BAD86798	BAD86798	1.10×10^{-6}

<i>Anabaena ucrainica</i> CHAB1432 ^e	<i>Cyanobacteria</i>	<i>Nostocaceae</i>	Germacradienol/ geosmin synthase	AEA03338	AEA03338	1.20×10^{-238}
<i>Anabaena ucrainica</i> CHAB2155 ^e	<i>Cyanobacteria</i>	<i>Nostocaceae</i>	Germacradienol/ geosmin synthase	AEA03341	AEA03341	1.90×10^{-238}
<i>Oscillatoria limosa</i> LBD 305b ^e	<i>Cyanobacteria</i>	<i>Oscillatoriales</i>	2-MIB/2-MB synthase ^c	ADU79150	ADU79150	3.10×10^{-70}
<i>Phormidium</i> sp. P2r ^e	<i>Cyanobacteria</i>	<i>Oscillatoriales</i>	Germacradienol/ geosmin synthase	ABU93239	ABU93239	1.10×10^{-242}
<i>Planktothricoides</i> <i>raciborskii</i> CHAB3331 ^e	<i>Cyanobacteria</i>	<i>Oscillatoriales</i>	2-MIB/2-MB synthase ^c	AEK21537	AEK21537	8.40×10^{-65}
<i>Pseudanabaena</i> <i>limnetica</i> str. Castaic Lake ^e	<i>Cyanobacteria</i>	<i>Oscillatoriales</i>	2-MIB synthase ^{c,d22}	ADU79148	ADU79148	1.10×10^{-70}
<i>Pseudanabaena</i> sp. dqh15 ^e	<i>Cyanobacteria</i>	<i>Oscillatoriales</i>	2-MIB/2-MB synthase ^c	AEK21533	AEK21533	8.40×10^{-65}
<i>Pseudanabaena</i> sp. NIVA-CYA 111 ^e	<i>Cyanobacteria</i>	<i>Oscillatoriales</i>	2-MIB/2-MB synthase ^c	ADU79149	ADU79149	3.60×10^{-69}

^aPredicted catalytic function was estimated by the phylogenetic analysis.

^bE-value was calculated Hmmpfam search using a second-generation model derived from aligned amino acid sequences of pentalenene, germacradienol/geosmin, *epi*-isoziaene, and 2-methylisoborneol synthases of *Streptomyces* strains.

^c2-MIB, 2-methylisoborneol synthase; 2-MB, 2-methylenebornane synthase.

^d(1–21) Catalytic function was confirmed by *in vitro* enzymatic reaction or heterologous expression. ^{d1}Komatsu et al. (2008), ^{d2}Cane et al. (2006), ^{d3}Takamatsu et al. (2011), ^{d4}Tetzlaff et al. (2006), ^{d5}Chou et al. (2010), ^{d6}Nakano, Kim, and Ohnishi (2011a), ^{d7}Hu, Chou, Hopson, and Cane (2011), ^{d8}Nakano, Kim, and Ohnishi (2011b), ^{d9}Gust et al. (2003), Cane and Watt (2003), ^{d10}Lin et al. (2006), ^{d11}Wang and Cane (2008), ^{d12}Nakano et al. (2011), ^{d13}Agger, Lopez-Gallego, Hoye, and Schmidt-Dannert (2008), ^{d14}Giglio, Jiang, Saint, Cane, and Monis (2008), ^{d15}Chou et al. (2011), ^{d16}Dairi et al. (2001), ^{d17}Nakano, Kudo, Eguchi, and Onishi (2011), ^{d18}Cane et al. (1994), ^{d19}Kim et al. (2011), ^{d20}Smanskia et al. (2011), ^{d21}Kawasaki et al. (2004), and ^{d22}Giglio et al. (2011).

^eGenome sequence has not been examined.

Table 7.2 Predicted type-B terpene synthases from genome-sequenced bacteria

Microorganism	Class or phylum	Order	Type of synthase	Locus tag or protein ID	Accession number	E-value^a
<i>Micromonospora aurantiaca</i> ATCC 27029	<i>Actinobacteria</i>	<i>Actinomycetales</i>	–	Micau_4289	YP_003837378	1.90×10^{-147}
<i>Micromonospora</i> sp. L5	<i>Actinobacteria</i>	<i>Actinomycetales</i>	–	ML5_4014	YP_004083675	2.30×10^{-147}
<i>Mycobacterium africanum</i> GM041182	<i>Actinobacteria</i>	<i>Actinomycetales</i>	–	MAF_33920	YP_004725027	1.00×10^{-194}
<i>Mycobacterium bovis</i> AF2122/97	<i>Actinobacteria</i>	<i>Actinomycetales</i>	–	Mb3411c	NP_857052	2.40×10^{-147}
<i>Mycobacterium bovis</i> BCG str. Pasteur 1173P2	<i>Actinobacteria</i>	<i>Actinomycetales</i>	–	BCG_3448c	YP_979527	2.40×10^{-147}
<i>Mycobacterium bovis</i> BCG str. Tokyo 172	<i>Actinobacteria</i>	<i>Actinomycetales</i>	–	JTY_3448	YP_002646489	2.40×10^{-147}
<i>Mycobacterium canettii</i> CIPT 140010059	<i>Actinobacteria</i>	<i>Actinomycetales</i>	–	MCAN_34031	YP_004746810	8.00×10^{-193}
<i>Mycobacterium tuberculosis</i> CDC1551	<i>Actinobacteria</i>	<i>Actinomycetales</i>	–	MT3487	NP_338009	3.00×10^{-194}

<i>Mycobacterium tuberculosis</i> F11	<i>Actinobacteria</i>	<i>Actinomycetales</i>	–	TBFG_13413	YP_001289340	9.00×10^{-194}
<i>Mycobacterium tuberculosis</i> H37Ra	<i>Actinobacteria</i>	<i>Actinomycetales</i>	–	MRA_3417	YP_001284763	9.00×10^{-194}
<i>Mycobacterium tuberculosis</i> H37Rv	<i>Actinobacteria</i>	<i>Actinomycetales</i>	Tubaculosinol diphosphate synthase ^{b1}	Rv3377c	NP_217894	9.00×10^{-194}
<i>Mycobacterium tuberculosis</i> KZN 1435	<i>Actinobacteria</i>	<i>Actinomycetales</i>	–	TBMG_03428	YP_003033425	9.00×10^{-194}
<i>Saccharomonospora viridis</i>	<i>Actinobacteria</i>	<i>Actinomycetales</i>	–	Svir_08360	YP_003132723	4.10×10^{-149}
<i>Salinispora arenicola</i> CNS-205	<i>Actinobacteria</i>	<i>Actinomycetales</i>	–	Sare_1288	YP_001536182	5.20×10^{-144}
<i>Streptomyces clavuligerus</i> ATCC 27064	<i>Actinobacteria</i>	<i>Actinomycetales</i>	–	SCLAV_5672	ZP_06775140	0
			–	SCLAV_p0490	ZP_06775672	0
<i>Streptosporangium roseum</i> DSM 43021	<i>Actinobacteria</i>	<i>Actinomycetales</i>	–	Sros_3710	YP_003339376	7.00×10^{-163}
			–	Sros_3192	YP_003338880	6.10×10^{-136}

Continued

Table 7.2 Predicted type-B terpene synthases from genome-sequenced bacteria—cont'd

Microorganism	Class or phylum	Order	Type of synthase	Locus tag or protein ID	Accession number	E-value
<i>Thermobispora bispora</i> DSM 43833	<i>Actinobacteria</i>	<i>Actinomycetales</i>	–	Tbis_3256	YP_003653839	4.10×10^{-155}
<i>Thermomonospora curvata</i> DSM 43183	<i>Actinobacteria</i>	<i>Actinomycetales</i>	–	Tcur_3808	YP_003301377	1.40×10^{-148}
<i>Verrucosipora maris</i> AB-18-032	<i>Actinobacteria</i>	<i>Actinomycetales</i>	–	VAB18032_25805	YP_004406859	2.80×10^{-145}
<i>Bradyrhizobium japonicum</i> USDA 110	<i>Alphaproteobacteria</i>	<i>Rhizobiales</i>	–	blr2149	NP_768789	7.70×10^{-70}
<i>Burkholderia multivorans</i> ATCC 17616	<i>Betaproteobacteria</i>	<i>Burkholderiales</i>	–	Bmul_5947	YP_001585899	3.10×10^{-60}
<i>Herpetosiphon aurantiacus</i> DSM 785	<i>Chloroflexi</i>	<i>Herpetosiphonales</i>	–	Haur_2145	YP_001544913	1.20×10^{-55}
<i>Mesorhizobium ciceri</i> biovar <i>biserrulae</i> WSM1271	<i>Alphaproteobacteria</i>	<i>Rhizobiales</i>	–	Mesci_5650	YP_004144784	6.40×10^{-70}

<i>Mesorhizobium loti</i> MAFF303099	<i>Alphaproteobacteria Rhizobiales</i>	–	mlr6369	NP_106893	1.00×10^{-67}
<i>Mesorhizobium opportunistum</i> WSM2075	<i>Alphaproteobacteria Rhizobiales</i>	–	Mesop_6233	YP_004614715	6.40×10^{-70}
<i>Rhizobium etli</i> CFN 42	<i>Alphaproteobacteria Rhizobiales</i>	–	RHE_PD00247	NP_659791	1.20×10^{-71}
<i>Rhizobium etli</i> CIAT 652	<i>Alphaproteobacteria Rhizobiales</i>	–	RHECIAT_PB0000287	YP_001984544	7.30×10^{-72}
<i>Sinorhizobium fredii</i> NGR234	<i>Alphaproteobacteria Rhizobiales</i>	–	NGR_a02760	NP_443949	4.10×10^{-70}
<i>Sorangium cellulosum</i> ‘So ce 56’	<i>Deltaproteobacteria Myxococcales</i>	–	sce4636	YP_001615279	2.50×10^{-66}
<i>Kitasatospora griseola</i> MF730-N6 ^c	<i>Actinobacteria Actinomycetales</i>	<i>ent-Copalyl diphosphate synthase^{b2}</i>	BAB39206	AB048795	0
<i>Streptomyces</i> sp. KO-3988 ^c	<i>Actinobacteria Actinomycetales</i>	<i>ent-Copalyl diphosphate synthase^{b3}</i>	BAD86797	BAD86797	0
<i>Streptomyces platensis</i> MA7327 ^c	<i>Actinobacteria Actinomycetales</i>	<i>ent-Copalyl diphosphate synthase^{b4}</i>	ACO31276	FJ655920	1.60×10^{-188}

Continued

Table 7.2 Predicted type-B terpene synthases from genome-sequenced bacteria—cont'd

Microorganism	Class or phylum	Order	Type of synthase	Locus tag or protein ID	Accession number	E-value
<i>Streptomyces platensis</i> MA7339 ^c	<i>Actinobacteria</i>	<i>Actinomycetales</i>	<i>ent</i> -Copalyl diphosphate synthase ^{b4}	ADD83015	GQ398492	2.50×10^{-193}
<i>Streptomyces lydicus</i> ^c	<i>Actinobacteria</i>	<i>Actinomycetales</i>	— ^d	CBA11579	CBA11579	2.10×10^{-142}
<i>Streptomyces</i> sp. SCSIO1666 ^c	<i>Actinobacteria</i>	<i>Actinomycetales</i>	— ^d	ADY38538	ADY38538	8.80×10^{-127}
<i>Streptomyces</i> sp. 307-9 ^c	<i>Actinobacteria</i>	<i>Actinomycetales</i>	— ^d	ADC79644	ADC79644	8.30×10^{-126}

^aE-value was calculated from Hmmpfam search using a model derived from aligned amino acid sequences of tuberculol diphosphate synthase (Rv337c) and *ent*-copalyl diphosphate synthases (BAB39206 and BAD86797).

^b(1–4) The catalytic function was confirmed by enzymatic reaction or heterologous expression. ^{b1}Nakano, Okamura, Sato, Dai, and Hoshino (2005), ^{b2}Dairi et al. (2001), ^{b3}Kawasaki et al. (2004), and ^{b4}Smanska et al. (2011).

^cGenome sequence has not been examined.

^dEach gene locates in the biosynthetic gene cluster for tirandamycin (ADY38538 and ADC79644) or streptolydigin (CBA11579), but their gene products have no apparent effect on the production of tirandamycin or streptolydigin.

Actinomycetes normally harbor at least two kinds of presumptive terpene synthases, typically germacradienol/geosmin synthase and 2-methylisoborneol/2-methylenebornane synthase. It can be expected, based on the emerging genomic sequence information now available, that these terpene synthases are widely distributed in actinomycetes, as previously indicated by chemical analysis of the volatile terpenoid metabolites from actinomycete cultures. In consideration of the type-A/type-B pair in the diterpene synthase family, Sare_1287/Sare_1288 of *Salinispora arenicola* CNS-205, SCLAV_5671/SCLAV_5672, and SCLAV_p0491/SCLAV_p0490 of *Streptomyces clavuligerus* ATCC 27064, and Sros_3191/Sros_3192 of *Streptosporangium roseum* DSM 43021 might be a combination of type-A and type-B diterpene synthases.

3.2. Expression of genes encoding bacterial terpene synthases in heterologous hosts

As described above, the sequences of presumptive terpene synthases can readily be harvested from bacterial genome data using profile HMM analysis, and the most likely acyclic allylic diphosphate substrate for the catalytic reaction of the presumptive synthase can also be identified with reasonable certainty by phylogenetic analysis based on alignment of the sequences with known terpene synthases. Nonetheless, the actual biochemical function of each presumptive synthase, including the identification of the actual substrate and the characteristic terpene cyclization product, must still be established by direct experiment. Two complementary strategies have been successfully employed toward these ends. *In vitro* incubation with recombinant protein has frequently been used for the evaluation of presumptive synthases. One drawback, however, is that such recombinant proteins expressed in *E. coli* are often obtained predominantly as insoluble inclusion bodies. In some cases, it has nonetheless been possible to successfully refold such inclusion bodies so as to obtain soluble proteins that possess the requisite enzymatic activity (Cane et al., 2006; Cane & Watt, 2003; Lin et al., 2006; Takamatsu et al., 2011). A versatile and powerful alternative strategy for the assignment of biochemical function has been heterologous expression of the synthase gene of interest in specially engineered bacterial host strains. Although normal strains of *E. coli* lack an adequate metabolic pool of isoprenoid precursors, metabolically engineered *E. coli* hosts designed for the efficient expression of terpenoid metabolites have been developed and successfully applied to the production of significant

titers (20–100 mg/l) of terpene metabolites (Harada et al., 2009; Hu et al., 2011; Martin, Pitera, Withers, Newman, & Keasling, 2003).

We have also developed a powerful and versatile alternative heterologous expression system for the production of a wide variety of natural products, including terpenoid metabolites. Since many actinomycetes produce geosmin and/or 2-methylisoborneol as terpenoid metabolites, these organisms seem to possess efficient metabolic pools of the precursors for terpene biosynthesis. The genome-sequenced *S. avermitilis* is used for industrial production of the widely used anthelmintic agent avermectin. Based on the genome data, we have developed engineered strains of *S. avermitilis* from which 1.5 Mb of the left subtelomeric region harboring the majority of natural product biosynthetic gene clusters has been deleted (Komatsu et al., 2008, Komatsu, Uchiyama, Omura, Cane, & Ikeda, 2010). These mega-deletion mutants, designated as “SUKA,” are robust natural product factories that retain the ability of the parent industrial avermectin-producing *S. avermitilis* to synthesize common biological building blocks and export biosynthetic end-products. Indeed, the SUKA mutants are excellent hosts for the introduction of individual biosynthetic genes or even entire biosynthetic gene clusters from *Streptomyces* or even eukaryotic sources, allowing determination of their biochemical function. The resultant transformants can also be used for the production of otherwise inaccessible, complex biosynthetic intermediates (Komatsu et al., 2010).

Since one of series of SUKA mutants, SUKA17, lacks terpene synthase genes, the mutant is unable to produce its natural spectrum of endogenous terpenoid metabolites. This SUKA17 host strain has been shown to heterologously express genes encoding the 2-methylisoborneol synthase as well as the codon-optimized amorph-4,11-diene synthase of the plant *Artemisia annua*. Terpene productivity was improved by cotranscribing a copy of the native *S. avermitilis* *ptlB* gene which encodes FPP synthase, thereby providing an efficient sources of FPP for the biosynthesis of amorph-4,11-diene (Komatsu et al., 2010) and avermitilol (Chou et al., 2010). For the expression of genes encoding monoterpene, sesquiterpene and diterpene synthases, these genes, including their native or a heterologous ribosome-binding sequence, can be inserted downstream of an appropriate GPP (*gps*), FPP (*ptlB*), or GGPP (*crtE*) synthase gene controlled by a constitutively expressed promoter (*rpsJp*) harbored in an integrating vector pKU1021 (Fig. 7.3). After PCR-amplification of each terpene synthase gene or preparation of the corresponding synthetic DNA

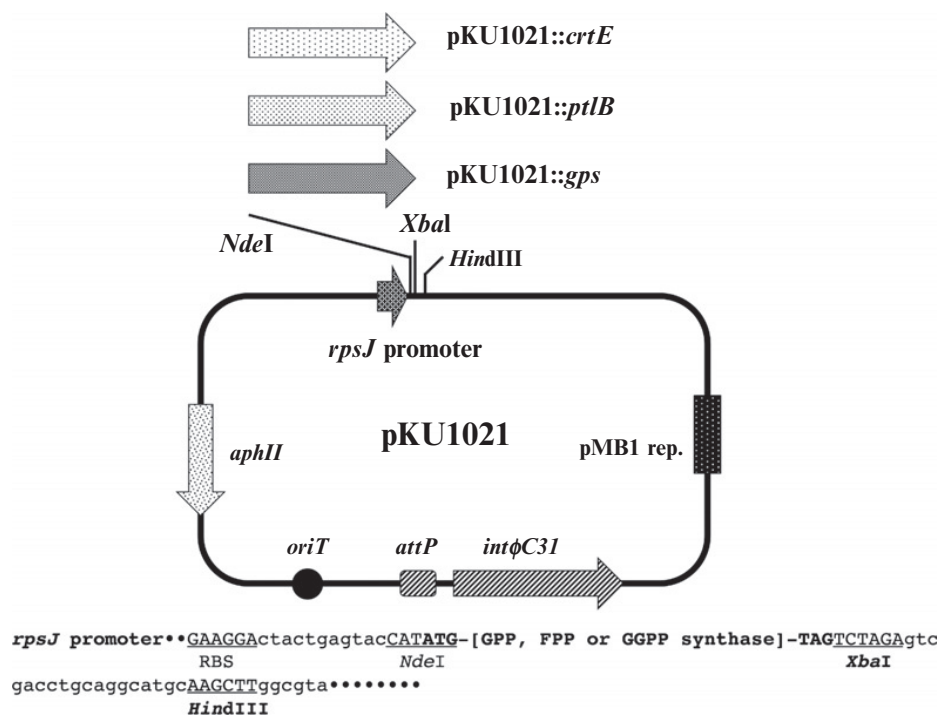


Figure 7.3 An integrating vector for expression of terpene synthase genes. The parent integrating vector pKU1021 was constructed by inserting a constitutively expressed promoter of the *S. avermitilis rpsJ* gene upstream of the multicloning site of pKU460 (Komatsu et al., 2010). Each gene encoding a polyprenyl diphosphate synthase is introduced downstream of the *rpsJ* promoter, in which the ATG of the *NdeI* is the start codon of each polyprenyl diphosphate synthase gene. The terpene synthase gene of interest is prepared by PCR using a pair of primers containing *XbaI* and *HindIII* sequences, respectively, for ligation.

with codons optimized for expression in *Streptomyces*, the amplified DNA can be digested with *XbaI* and *HindIII*, both of which represent rare recognition sites in *Streptomyces* genomes, and the digested DNA is inserted at the *XbaI/HindIII* sites of the expression vector, pKU1021::gps, pKU1021::ptlB, or pKU1021::crtE.

Protocol 1. Construction of expression cassette on integrating vector for terpene synthase gene of interest

1. Putative terpene synthase genes, *sclav_p0982* and *sclav_p0068*, were amplified by PCR with a template DNA from *S. clavuligerus* ATCC 27067 using following the primer pairs: forward primer 5'-CGGCtctagaTGCGAGGT**CGACATG**TCCCTGAAC-3' (lowercase, underlined, and bold characters indicate *XbaI* site, ribosome-binding site, and start codon of *sclav_p0982*, respectively) and reverse primer 5'-CTCGAGaagcttCGGAGGGACGGGCGGGT**CACCAAG**-3'

(lowercase and bold characters are *Hind*III site and stop codon of *sclav_p0982*, respectively) for *sclav_p0982*, and forward primer 5'-C GGCTctagaTGCGAGGTCGAC**ATGT**CCCTGAAC-3' (lowercase, underlined, and bold characters are *Xba*I site, ribosome-binding site, and start codon of *sclav_p0068*, respectively) and reverse primer 5'-GGTCACaagcttCGGGCAGCTCCCGGATGAAGT-3' (lowercase characters are *Hind*III site) for *sclav_p0068* in the presence of 5% of dimethylsulfoxide.

2. After addition of high-fidelity DNA polymerase, the initial denaturation step (95 °C, 2 min) was followed by 25 cycles amplification (95 °C, 0.5 min: 55 °C, 0.5 min, 72 °C, 1 min) and then a final incubation at 72 °C (10 min). The resulting amplicon was purified, digested with *Xba*I and *Hind*III, and ligated into the corresponding site of pKU1021::*gps* (for *sclav_p0982*) or pKU1021::*ptlB* (for *sclav_p0068*) with T4 DNA ligase at 12 °C overnight.
3. Each ligation product was introduced into *E. coli* DH5 α by electroporation and transformants were selected by kanamycin (50 μ g/ml). The DNA sequence of each insert was verified by dideoxy sequencing.

Protocol 2. Introduction of recombinant integrating plasmid into *S. avermitilis* SUKA17

1. To prepare unmethylated DNA preparation, the resulting plasmids, pKU1021::*gps*::*sclav_p0982* and pKU1021::*ptlB*::*sclav_p0068*, were introduced into *E. coli* GM2929 *hsdS*::Tn10 (Komatsu et al., 2010) by electroporation and desired transformants were selected by kanamycin (50 μ g/ml) and chloramphenicol (30 μ g/ml). Unmethylated recombinant plasmid DNA was prepared from the transformants.
2. Preparation of *S. avermitilis* SUKA17 protoplasts was described as previously (Ikeda, Takada, Pang, Tanaka, & Omura, 1993). Protoplasts were transformed by unmethylated recombinant integrating plasmid using 25% polyethyleneglycol 1000. The resulting protoplasts were regenerated by growing on RM14 medium by pouring over 2.5 ml of molten RM14 soft agar. After plates were incubated at 30 °C for 18 h, transformants were selected by overlaying 2.5 ml of RM14 soft agar containing 100 μ g/ml of neomycin and the plates were incubated for at least 5 days.
3. Transformants regenerated to mycelial form were inoculated onto YMS plate (Ikeda, Kotaki, & Omura, 1987) containing 1 μ g/ml of neomycin and incubated at 30 °C until sporulation (about 4 days).

Protocol 3. Cultivation for the production of terpene compounds and analysis of products

1. Spores of transformants of *S. avermitilis* SUKA17 carrying the terpene synthase gene were inoculated into 10 ml of vegetative medium (Komatsu et al., 2008) in 80-ml test tube and grown with reciprocal shaking for 2 days at 28 °C. A 0.1 ml portion of the vegetative culture was inoculated into 10 ml of production medium (Komatsu et al., 2008) in a 125-ml flask and grown on a rotary shaker for 5 days at 28 °C.
2. After cultivation, the culture was separated by centrifugation. The culture broth was extracted with 1 ml of *n*-hexane. The mycelium was extracted with 5 ml of methanol. After removal of mycelium, the methanol extract was reextracted with 1 ml of *n*-hexane.
3. Both organic layers from the culture broth and methanol extract of the mycelium were dehydrated by 0.5 g of anhydrous Na₂SO₄. A 5 µl portion of *n*-hexane extract from the culture broth or the mycelium was analyzed by GC-MS (Shimadzu GC-17A, 70 eV, EI, positive ion mode) using 30 m × 0.25 mm neutral bond-5 capillary column (5% phenylmethylsilicon) and a temperature program of 50–250 °C (temperature gradient of 25 °C/min). Authentic sample of 1,8-cineole was purchased and (+)-T-muuirolol was prepared by enzymatic reaction of recombinant SCLAV_p0068 and FPP in the presence of magnesium ions (Hu et al., 2011).

In this manner, when *sclav_p0982* of *S. clavuligerus* ATCC 27064 was expressed along with a gene encoding FPP synthase, no terpenoid metabolites were found in the culture (middle line in Fig. 7.4), while coexpression with a gene encoding GPP synthase resulted in formation of a large amount of 1,8-cineole in the culture broth (upper line in Fig. 7.4). This approach has also been applied to the expression of *sclav_p0068* of *S. clavuligerus* ATCC 27064, giving rise to the presence of abundant (+)-T-muuirolol in the culture of the transformants (Fig. 7.5). The combination of versatile host, SUKA17, and coexpression of presumptive terpene synthase gene with an appropriate polyprenyl diphosphate synthase gene is therefore exceptionally useful for the discovery and biochemical characterization of new terpene synthases as well as the preparative isolation of terpenoid metabolites. The method has been further enhanced by the coexpression of a wide range of downstream genes mediating late-stage biosynthetic steps, including P450s, nonheme iron, α -ketoglutarate-dependent dioxygenases, flavin-dependent monooxygenases, and dehydrogenases (Jiang et al., 2009; Seo, Zhu, Endo, Ikeda, & Cane, 2011; Tetzlaff et al., 2006).

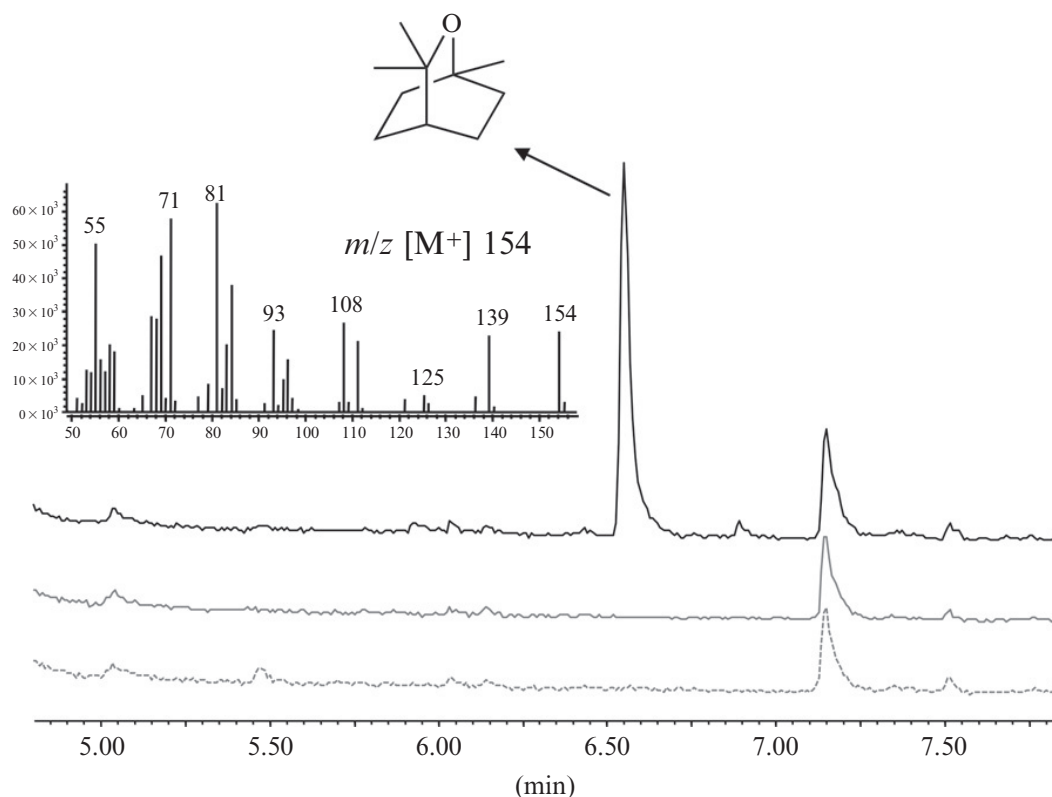


Figure 7.4 GC–MS analysis of an *n*-hexane extract of *S. avermitilis* SUKA17 transformants. The lower, middle, and upper lines indicate chromatograms of transformants carrying pKU1021, pKU1021::ptIB::sclav_p0982, and pKU1021::gps::sclav_p0982 of *S. clavuligerus* ATCC 27064, respectively. A peak eluting at 6.512 min was identical to 1,8-cineole by comparison with an authentic sample of 1,8-cineole.

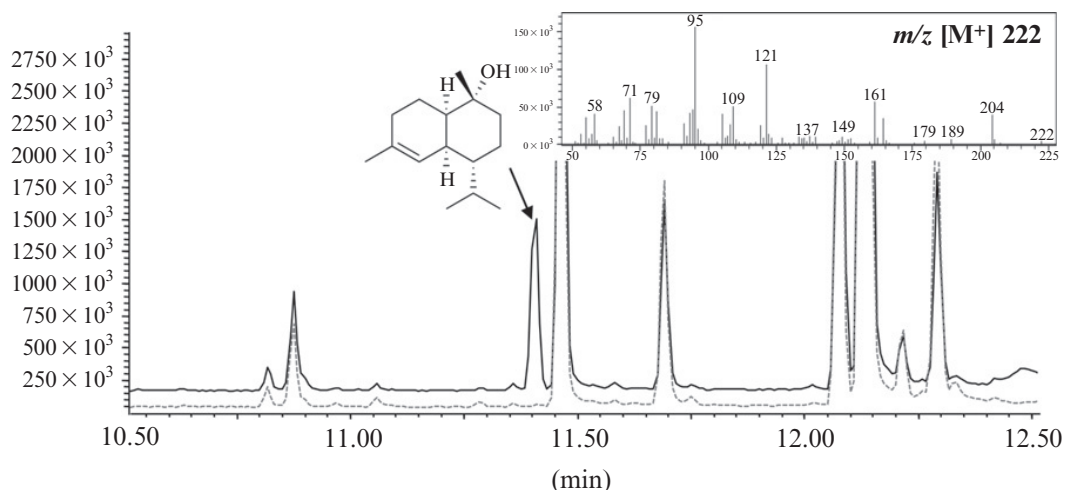


Figure 7.5 GC–MS analysis of an *n*-hexane extract of *S. avermitilis* SUKA17 transformants. The lower and upper lines indicate chromatograms of transformants carrying pKU1021 and pKU1021::ptIB::sclav_p0068 of *S. clavuligerus* ATCC 27064, respectively. A peak eluting at 11.413 min was identical to (+)-T-muurolol by comparison with an authentic sample of (+)-T-muurolol.

REFERENCES

- Agger, S. A., Lopez-Gallego, F., Hoye, T. R., & Schmidt-Dannert, C. (2008). Identification of sesquiterpene synthases from *Nostoc punctiforme* PCC 73102 and *Nostoc* sp. strain PCC 7120. *Journal of Bacteriology*, 190, 6084–6096.
- Berthelot, M., & André, G. (1891). Sur l'odeur proper de la terre. *Comptes Rendus de l'Académie des Sciences*, 112, 598–599.
- Cane, D. E., He, X., Kobayashi, S., Omura, S., & Ikeda, H. (2006). Geosmin biosynthesis in *Streptomyces avermitilis*. Molecular cloning, expression, and mechanistic study of the germacradienol/geosmin synthase. *The Journal of Antibiotics*, 59, 471–479.
- Cane, D. E., Oliver, J. S., Harrison, P. H. M., Abell, C., Hubbard, B. R., Kane, C. T., et al. (1990). The biosynthesis of pentalenene and pentalenolactone. *Journal of the American Chemical Society*, 112, 4513–4524.
- Cane, D. E., Sohng, J.-K., Lamberson, C. R., Rudnicki, S. M., Wu, Z., Lloyd, M. D., et al. (1994). Pentalenene synthase. Purification, molecular cloning, sequencing and high-level expression in *Escherichia coli* of a terpenoid cyclase from *Streptomyces* UC5319. *Biochemistry*, 33, 5846–5857.
- Cane, D. E., Tandon, M., & Prabhakaran, P. C. (1993). Epicubenol synthase and the enzymic cyclization of farnesyl diphosphate. *Journal of the American Chemical Society*, 115, 8103–8106.
- Cane, D. E., & Watt, R. M. (2003). Expression and mechanistic analysis of a germacradienol synthase from *Streptomyces coelicolor* implicated in geosmin biosynthesis. *Proceedings of the National Academy of Sciences of the United States of America*, 100, 1547–1551.
- Chou, W. K., Fanizza, I., Uchiyama, T., Komatsu, M., Ikeda, H., & Cane, D. E. (2010). Genome mining in *Streptomyces avermitilis*: Cloning and characterization of SAV_76, the synthase for a new sesquiterpene, avermitilol. *Journal of the American Chemical Society*, 132, 8850–8851.
- Chou, W. K., Ikeda, H., & Cane, D. E. (2011). Cloning and characterization of Pfl_1841, a 2-methylenebornane synthase in *Pseudomonas fluorescens* PfO-1. *Tetrahedron*, 67, 6627–6632.
- Christianson, D. W. (2006). Structural biology and chemistry of the terpenoid cyclases. *Chemical Reviews*, 106, 3412–3442.
- Christianson, D. W. (2008). Unearthing the roots of the terpenome. *Current Opinion in Chemical Biology*, 12, 141–150.
- Dairi, T., Hamano, Y., Kuzuyama, T., Itoh, N., Furihata, K., & Seto, H. (2001). Eubacterial diterpene cyclase genes essential for production of the isoprenoid antibiotic terpentecin. *Journal of Bacteriology*, 183, 6085–6094.
- Felicetti, B., & Cane, D. E. (2004). Aristolochene synthase: Mechanistic analysis of active site residues by site-directed mutagenesis. *Journal of the American Chemical Society*, 126, 7212–7221.
- Finn, R. D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J. E., et al. (2010). The Pfam protein families database. *Nucleic Acids Research*, 38, D211–D222.
- Gerber, N. N. (1967). Geosmin, an earthy-smelling substance isolated from actinomycetes. *Biotechnology and Bioengineering*, IX, 321–327.
- Gerber, N. N. (1969). A volatile metabolite of actinomycetes, 2-methylisoborneol. *The Journal of Antibiotics*, 22, 508–509.
- Gerber, N. N. (1971). Sesquiterpenoids from Actinomycetes: Cadin-4-en-1-ol. *Phytochemistry*, 10, 185–189.
- Gerber, N. N. (1973). Volatile lactones from *Streptomyces*. *Tetrahedron Letters*, 14, 771–774.
- Gerber, N. N., & Lechevalier, H. A. (1965). Geosmin, an earthy-smelling substance isolated from actinomycetes. *Applied Microbiology*, 13, 935–938.
- Giglio, S., Chou, W. K., Ikeda, H., Cane, D. E., & Monis, P. T. (2011). Biosynthesis of 2-methylisoborneol in cyanobacteria. *Environmental Science and Technology*, 45, 992–998.
- Giglio, S., Jiang, J., Saint, C. P., Cane, D. E., & Monis, P. T. (2008). Isolation and characterization of the gene associated with geosmin production in cyanobacteria. *Environmental Science and Technology*, 42, 8027–8032.

- Glasby, J. S. (1982). *Encyclopedia of the terpenoids*. Chichester: John Wiley & Sons.
- Gust, B., Challis, G. L., Fowler, K., Kieser, T., & Chater, K. F. (2003). PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. *Proceedings of the National Academy of Sciences of the United States of America*, 100, 1541–1546.
- Hamano, Y., Kuzuyama, T., Itoh, N., Furihata, K., Seto, H., & Dairi, T. (2002). Functional analysis of eubacterial diterpene cyclases responsible for biosynthesis of a diterpene antibiotic, terpentecin. *The Journal of Biological Chemistry*, 277, 37098–37104.
- Harada, H., Yu, F., Okamoto, S., Kuzuyama, T., Utsumi, R., & Misawa, N. (2009). Efficient synthesis of functional isoprenoids from acetoacetate through metabolic pathway-engineered *Escherichia coli*. *Applied Microbiology and Biotechnology*, 81, 915–925.
- Hu, Y., Chou, W. K. W., Hopson, R., & Cane, D. E. (2011). Genome mining in *Streptomyces clavuligerus*: Expression and biochemical characterization of two new cryptic sesquiterpene synthases. *Chemistry and Biology*, 18, 32–37.
- Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H., Shiba, T., et al. (2003). Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. *Nature Biotechnology*, 21, 526–531.
- Ikeda, H., Kotaki, H., & Omura, S. (1987). Genetic studies of avermectin biosynthesis in *Streptomyces avermitilis*. *Journal of Bacteriology*, 169, 5615–5621.
- Ikeda, H., Takada, Y., Pang, C.-H., Tanaka, H., & Omura, S. (1993). Transposon mutagenesis by Tn4560 and applications with avermectin-producing *Streptomyces avermitilis*. *Journal of Bacteriology*, 175, 2077–2082.
- Jáchymová, J., Votruba, J., Viden, I., & Rezanka, T. (2002). Identification of *Streptomyces* odor spectrum. *Folia Microbiologica*, 47, 37–41.
- Jiang, J., & Cane, D. E. (2008). Geosmin biosynthesis. Mechanism of the fragmentation-rearrangement in the conversion of germacradienol to geosmin. *Journal of the American Chemical Society*, 130, 428–429.
- Jiang, J., Tetzlaff, C. N., Takamatsu, S., Iwatsuki, M., Komatsu, M., Ikeda, H., et al. (2009). Genome mining in *Streptomyces avermitilis*: A biochemical Baeyer-Villiger reaction and discovery of a new branch of the pentalenolactone family tree. *Biochemistry*, 48, 6431–6440.
- Kawasaki, T., Hayashi, Y., Kuzuyama, T., Furihata, K., Itoh, N., Seto, H., et al. (2006). Biosynthesis of a natural polyketide-isoprenoid hybrid compound, furaquinocin A: Identification and heterologous expression of the gene cluster. *Journal of Bacteriology*, 188, 1236–1244.
- Kawasaki, T., Kuzuyama, T., Kuwamori, Y., Matsuura, N., Itoh, N., Furihata, K., et al. (2004). Presence of copalyl diphosphate synthase gene in an actinomycete possessing the mevalonate pathway. *The Journal of Antibiotics*, 57, 739–747.
- Kim, S.-Y., Zhao, P., Igarashi, M., Sawa, R., Tomita, T., Nishiyama, M., et al. (2011). Cloning and heterologous expression of the cyclooctatin biosynthetic gene cluster afford a diterpene cyclase and two P450 hydroxylases. *Chemistry and Biology*, 16, 736–743.
- Komatsu, M., Tsuda, M., Omura, S., Oikawa, H., & Ikeda, H. (2008). Identification and functional analysis of genes controlling biosynthesis of 2-methylisoborneol. *Proceedings of the National Academy of Sciences of the United States of America*, 105, 7422–7427.
- Komatsu, M., Uchiyama, T., Omura, S., Cane, D. E., & Ikeda, H. (2010). Genome-minimized *Streptomyces* host for the heterologous expression of secondary metabolism. *Proceedings of the National Academy of Sciences of the United States of America*, 107, 2646–2651.
- Lesburg, C. A., Zhai, G., Cane, D. E., & Christianson, D. W. (1997). Crystal structure of pentalenene synthase: Mechanistic insights on terpenoid cyclization reactions in biology. *Science*, 277, 1820–1824.

- Lin, X., Hopson, R., & Cane, D. E. (2006). Genome mining in *Streptomyces coelicolor*: Molecular cloning and characterization of a new sesquiterpene synthase. *Journal of the American Chemical Society*, 128, 6022–6023.
- MacMillan, J., & Beale, M. H. (1999). Diterpene biosynthesis. In D. E. Cane (Ed.), *Comprehensive natural products chemistry: Isoprenoids including carotenoids and steroids*, Vol. 2, (pp. 217–243). Amsterdam, Netherlands: Elsevier.
- Martin, V. J., Pitera, D. J., Withers, S. T., Newman, J. D., & Keasling, J. D. (2003). Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. *Nature Biotechnology*, 21, 796–802.
- Moody, S. C., Zhao, B., Lei, L., Nelson, D. R., Mullins, J. G., Waterman, M. R., et al. (2012). Investigating conservation of the albaflavenone biosynthetic pathway and CYP170 bifunctionality in streptomycetes. *FEBS Journal*, 279, 1640–1649.
- Nakano, C., Horinouchi, S., & Ohnishi, Y. (2011). Characterization of a novel sesquiterpene cyclase involved in (+)-caryolan-1-ol biosynthesis in *Streptomyces griseus*. *The Journal of Biological Chemistry*, 286, 27980–27987.
- Nakano, C., Kim, H.-K., & Ohnishi, Y. (2011a). Identification of the first bacterial monoterpene cyclase, a 1,8-cineole synthase, that catalyzes the direct conversion of geranyl diphosphate. *Chembiochem: A European Journal of Chemical Biology*, 12, 1988–1991.
- Nakano, C., Kim, H.-K., & Ohnishi, Y. (2011b). Identification and characterization of the linalool/nerolidol synthase from *Streptomyces clavuligerus*. *Chembiochem: A European Journal of Chemical Biology*, 12, 2403–2407.
- Nakano, C., Kudo, F., Eguchi, T., & Onishi, Y. (2011). Genome mining reveals two novel bacterial sesquiterpene cyclases: (-)-Germacradien-4-ol and (-)-*epi*- α -bisabolol synthases from *Streptomyces citricolor*. *Chembiochem: A European Journal of Chemical Biology*, 12, 2271–2275.
- Nakano, C., Okamura, T., Sato, T., Dairi, T., & Hoshino, T. (2005). *Mycobacterium tuberculosis* H37Rv3377c encodes the diterpene cyclase for producing the halimane skeleton. *Chemical Communications (Cambridge, England)*, 8, 1016–1018.
- Nawrath, T., Dickschat, J. S., Muller, R., Jiang, J., Cane, D. E., & Schulz, S. (2008). Identification of (8S,9S,10S)-8,10-dimethyl-1-octalin, a key intermediate in the biosynthesis of geosmin in bacteria. *Journal of the American Chemical Society*, 130, 430–431.
- Pollack, F. C., & Berger, R. G. (1996). Geosmin and related volatiles in bioreactor-cultured *Streptomyces citreus* CBS 109.60. *Applied and Environmental Microbiology*, 62, 1295–1299.
- Sacchettini, J. C., & Poulter, C. D. (1997). Biochemistry: Creating isoprenoid diversity. *Science*, 277, 1788–1789.
- Schöller, C. G., Gürtler, H., Redersen, R., Molin, S., & Wilkins, K. (2002). Volatile metabolites from actinomycetes. *Journal of Agricultural and Food Chemistry*, 50, 2612–2615.
- Seo, M.-J., Zhu, D., Endo, S., Ikeda, H., & Cane, D. E. (2011). Genome mining in *Streptomyces*. Elucidation of the role of Baeyer-Villiger monooxygenases and non-heme iron-dependent dehydrogenase/oxygenases in the final steps of the biosynthesis of pentalenolactone and neopentalenolactone. *Biochemistry*, 50, 1739–1754.
- Smanskia, M. J., Yu, Z., Casper, J., Lin, S., Peterson, R. M., Chen, Y., et al. (2011). Dedicated *ent*-kaurene and *ent*-atiserene synthases for platensimycin and platencin biosynthesis. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 13498–13503.
- Sonnhammer, E. L., Eddy, S. R., Birney, E., Bateman, A., & Durbin, R. (1998). Pfam: Multiple sequence alignments and HMM-profiles of protein domains. *Nucleic Acids Research*, 26, 320–322.
- Takamatsu, S., Lin, X., Nara, A., Komatsu, M., Cane, D. E., & Ikeda, H. (2011). Characterization of a silent sesquiterpenoid biosynthetic pathway in *Streptomyces avermitilis* controlling *epi*-isozizaene and albaflavenone biosynthesis and isolation of a new oxidized *epi*-isozizaene metabolite. *Microbial Biotechnology*, 4, 184–191.

- Tetzlaff, C. N., You, Z., Cane, D. E., Takamatsu, S., Omura, S., & Ikeda, H. (2006). A gene cluster for biosynthesis of the sesquiterpenoid antibiotic pentalenolactone in *Streptomyces avermitilis*. *Biochemistry*, 45, 6179–6186.
- Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, 22, 4673–4680.
- Wang, C. M., & Cane, D. E. (2008). Biochemistry and molecular genetics of the biosynthesis of the earthy odorant methylisoborneol in *Streptomyces coelicolor*. *Journal of the American Chemical Society*, 130, 8908–8909.
- Wilkins, K., & Schöller, C. (2009). Volatile organic metabolites from selected *Streptomyces* strains. *Actinomycetologica*, 23, 27–33.