Terpene synthases are widely distributed in bacteria

Yuuki Yamada[a], Tomohisa Kuzuyama[b], Mamoru Komatsu[a], Kazuo Shin-ya[c], Satoshi Omura[d,1], David E. Cane[e], and Haruo Ikeda[a,1]

[a]Laboratory of Microbial Engineering, Kitasato Institute for Life Sciences, Kitasato University, Kanagawa 252-0373, Japan; [b]Biotechnology Research Center, University of Tokyo, Tokyo 113-8657, Japan; [c]National Institute of Advanced Industrial Science and Technology, Tokyo 135-0064, Japan; [d]Laboratory of Microbial Engineering, Kitasato Institute for Life Sciences, Kitasato University, Tokyo 108-8461, Japan; and [e]Department of Chemistry, Brown University, Providence, RI 02912-9108

Contributed by Satoshi Omura, November 24, 2014 (sent for review October 14, 2014)

Odoriferous terpene metabolites of bacterial origin have been known for many years. In genome-sequenced *Streptomycetaceae* microorganisms, the vast majority produces the degraded sesquiterpene alcohol geosmin. Two minor groups of bacteria do not produce geosmin, with one of these groups instead producing other sesquiterpene alcohols, whereas members of the remaining group do not produce any detectable terpenoid metabolites. Because bacterial terpene synthases typically show no significant overall sequence similarity to any other known fungal or plant terpene synthases and usually exhibit relatively low levels of mutual sequence similarity with other bacterial synthases, simple correlation of protein sequence data with the structure of the cyclized terpene product has been precluded. We have previously described a powerful search method based on the use of hidden Markov models (HMMs) and protein families database (Pfam) search that has allowed the discovery of monoterpene synthases of bacterial origin. Using an enhanced set of HMM parameters generated using a training set of 140 previously identified bacterial terpene synthase sequences, a Pfam search of 8,759,463 predicted bacterial proteins from public databases and in-house draft genome data has now revealed 262 presumptive terpene synthases. The biochemical function of a considerable number of these presumptive terpene synthase genes could be determined by expression in a specially engineered heterologous *Streptomyces* host and spectroscopic identification of the resulting terpene products. In addition to a wide variety of terpenes that had been previously reported from fungal or plant sources, we have isolated and determined the complete structures of 13 previously unidentified cyclic sesquiterpenes and diterpenes.

terpene synthase | bacteria | heterologous expression

Some 50,000 terpenoid metabolites, including monoterpenes, sesquiterpenes, and diterpenes representing nearly 400 distinct structural families, have been isolated from both terrestrial and marine plants, liverworts, and fungi. In contrast, only a relatively minor fraction of these widely occurring metabolites has been identified in prokaryotes. The first study of bacterial terpenes grew out of an investigation of the characteristic odor of freshly plowed soil reported in 1891 by Berthelot and André (1). Berthelot and André noted that a volatile substance apparently responsible for the typical earthy odor of soil could be extracted from soil by steam distillation. Their attempts to assign a structure to the isolated odor constituent failed;, however, when the neutral alcohol resisted oxidative degradation or other conventional chemical modification. The first modern studies of volatile bacterial terpenes were carried out some 75 years later by Gerber and Lechevalier (2) and Gerber (3–7), who speculated that the characteristic odor of cultures of *Actinomycetales* microorganisms, which are widely distributed in soil, might be caused by volatile terpenes. In addition to determining the structure of Berthelot's geosmin, shown to be a C_{12} degraded sesquiterpene alcohol (and giving it its name, which means earth odor) (2, 3), Gerber (4) also isolated and determined the structures of the methylated monoterpene 2-methylisoborneol as well as several other cyclic sesquiterpenes produced by streptomycetes (5–7). In subsequent years, numerous volatile terpenes have been detected in streptomycetes (8–16). The three most commonly detected streptomycetes terpenoids, geosmin, and 2-methylisoborneol and the tricyclic α,β-unsaturated ketone albaflavenone (Fig. 1) are well-known as volatile odoriferous microbial metabolites. The two terpene alcohols are, in fact, the most frequently found secondary metabolites in actinomycetes (8, 11, 17), filamentous *Cyanobacteria* (18–20), and *Myxobacteria* (21), and they are also produced by a small number of fungi (22–24). The production of 2-methylisoborneol is associated with a characteristic scent, whereas albaflavenone, which was first isolated from cultures of a highly odoriferous *Streptomyces albidoflavus* species, is best described as earthy and camphor-like (25).

Cyclic monoterpene, sesquiterpene, and diterpene hydrocarbons and alcohols are formed by variations of a universal cyclization mechanism that is initiated by enzyme-catalyzed ionization of the universal acyclic precursors geranyl diphosphate (GPP), farnesyl diphosphate (FPP), and geranylgeranyl diphosphate (GGPP) to form the corresponding allylic cations. These parental branched, linear isoprenoid precursors are themselves synthesized by mechanistically related electrophilic condensations of the 5-carbon building blocks dimethylallyl diphosphate and isopentenyl diphosphate. The several thousand known or suspected terpene synthases from plants and fungi have a strongly conserved level of overall amino acid sequence similarity, thus making possible the application of local alignment methods, such as the widely used BLAST algorithm, for the discovery of genes encoding presumptive terpene synthases from plant and fungal sources. Despite the relatively high level of overall sequence conservation, however, assignment of the actual biosynthetic cyclization product of each fungal or plant terpene synthase has remained beyond the reach of available bioinformatic

Significance

Terpenes are generally considered to be plant or fungal metabolites, although a small number of odoriferous terpenes of bacterial origin have been known for many years. Recently, extensive bacterial genome sequencing and bioinformatic analysis of deduced bacterial proteins using a profile based on a hidden Markov model have revealed 262 distinct predicted terpene synthases. Although many of these presumptive terpene synthase genes seem to be silent in their parent microorganisms, controlled expression of these genes in an engineered heterologous *Streptomyces* host has made it possible to identify the biochemical function of the encoded terpene synthases. Genes encoding such terpene synthases have been shown to be widely distributed in bacteria and represent a fertile source for discovery of new natural products.

Author contributions: Y.Y., D.E.C., and H.I. designed research; Y.Y. and H.I. performed research; Y.Y., T.K., M.K., K.S.-y., S.O., D.E.C., and H.I. analyzed data; and Y.Y., D.E.C., and H.I. wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

[1]To whom correspondence may be addressed. Email: omuras@insti.kitasato-u.ac.jp or ikeda@ls.kitasato-u.ac.jp.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1422108112/-/DCSupplemental.

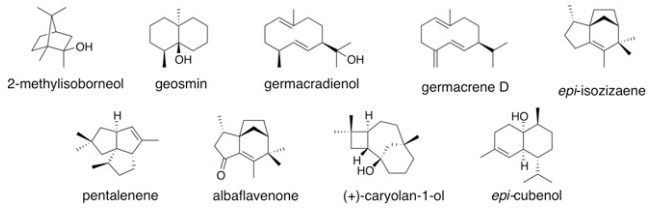

Fig. 1. The structures of the major known terpenes produced by bacteria.

methods. The discovery and biochemical characterization of bacterial terpene synthases represent an even greater challenge, because unlike the plant and fungal enzymes, bacterial terpene synthases not only exhibit no significant overall amino acid sequence similarity to those from plants and fungi but typically display relatively low levels of mutual sequence similarity. To address this challenge, we recently described the successful application of an alternative genome mining strategy for the discovery of previously unidentified bacterial terpene synthases based on the use of hidden Markov models (HMMs) and protein families database (Pfam) searching methods (26). These initial efforts identified a large number of previously unrecognized bacterial terpene synthase candidates, including the discovery of the previously unidentified synthase for the methylated monoterpene 2-methylisoborneol, and led to the heterologous expression of the relevant genes that produce 2-methylisoborneol and 2-methylenebornane from 2-methylgeranyl diphosphate (27). We subsequently refined and expanded the set of HMM parameters using as a reference set exclusively the group of newly predicted bacterial terpene synthases in distinction to the original HMM model (PF03936), which had been based on plant terpene synthases. Using these newly refined parameters, we then succeeded in identifying a previously unrecognized ortholog of 2-methylisoborneol synthase in the cyanobacterium *Pseudanabaena limnetica* str. Castaic Lake (28). Application of this second generation HMM model allowed, in total, the discovery of 140 predicted terpene synthases of bacterial origin.

We now report the development of a third generation HMM model trained by the previously identified 140 bacterial terpene synthases that has expanded the number of predicted bacterial terpene synthases to 262 from within the most complete set of predicted proteins incorporated in the most recent collection of public databases and in-house draft genome sequences of streptomycete microorganisms. Among the newly identified gene sequences, a subset selected by phylogenetic analysis has been expressed in a specially engineered heterologous *Streptomyces* host, and the resultant terpenes have been identified and structurally characterized.

Results and Discussion

Terpenes Derived from Genome-Sequenced *Streptomycetaceae* Microorganisms. To confirm production of terpenoid metabolites in genome-sequenced *Streptomyces* and *Kitasatospora* microorganisms, whole cultures grown on agar medium were extracted with *n*-hexane/methanol and then directly analyzed by GC-MS. The majority of the microorganisms examined produced odoriferous terpenes, including the degraded sesquiterpene alcohol (geosmin) along with the biosynthetic intermediate germacradienol and its cometabolite germacrene D (Fig. 1). The observed terpene product profiles allowed classification of the bacterial strains into three main groups. Group A: *Streptomyces albus* J1074, *Streptomyces avermitilis* American Type Culture Collection (ATCC) 31267, *Streptomyces coelicolor* A3(2), *Streptomyces venezuelae* ATCC 10712, *Kitasatospora setae* KM-6054, *Streptomyces cyaneogriseus* susp. *noncyanogenus* Northern Regional Research Laboratory (NRRL) 15773, *Streptomyces ghanaensis* ATCC 14672, *Streptomyces pristinaespiralis* ATCC 25486, *Streptomyces sviceus* ATCC 29083, and *Streptomyces viridochromogenes* Deutsche Sammlung von Mikroorganismen und ellkulturen GmbH (DSM) 40736 each produced geosmin and its typical cometabolites, whereas *S. albus* and *S. cyaneogriseus* susp. *noncyanogenus* also produced the sesquiterpene albaflavenone in addition to geosmin (Fig. S1). Group B: *Streptomyces griseus*, *Streptomyces roseosporus* NRRL 11379, and *Streptomyces* sp. SirexAA-E produced mainly the sesquiterpene alcohols (+)-caryolan-1-ol and *epi*-cubenol when grown on agar medium (Fig. S1). Although *S. griseus* also harbors genes apparently encoding geosmin synthase and 2-methylisoborneol synthase, the observed level of geosmin production was extremely low, and only trace amounts of 2-methylisoborneol could be detected in liquid culture. Because the sesquiterpenes (+)-caryolan-1-ol, *epi*-cubenol, and geosmin are each formed by cyclization of FPP, the relative yields of these three products might simply reflect competition between the individual synthases for the acyclic common substrate FPP or perhaps more likely, differential expression of the synthases themselves. Group C: extracts of both *Streptomyces clavuligerus* ATCC 27064 and *Streptomyces lactacystinaeus* OM-6519 grown under a variety of culture conditions did not produce any detectable terpenes (Fig. S1). The observed metabolite production profile of *S. pristinaespiralis* ATCC 25486 placed it in an intermediate terpene synthase group, in which both geosmin and selina-4(15),7(11)-diene (Fig. S1) were detected but at relatively low levels. Taken together, these results clearly indicate that the majority of *Streptomyces* species have the ability to produce the 12-carbon degraded sesquiterpene alcohol geosmin, with some strains coproducing either 2-methylisoborneol or albaflavenone.

Bioinformatic Screening for Bacterial Terpene Synthases. All terpene synthases of bacterial, fungal, and plant origin possess a pair of characteristic conserved metal-binding domains consisting of an acidic amino acid-rich motif [D/N)DXX(D/E) or DDXXXE] located within 80–120 (bacterial and fungal terpene synthases) or 230–270 aa (plant terpene synthases) of the N terminus and an Asn/Ser/Glu triad (NxxxSxxxE) found some 140 aa downstream, closer to the C terminus. Furthermore, plant synthases contain an additional characteristic conserved N-terminal domain (PF01397) of 252 aa of completely unknown biochemical function that forms an α-barrel structure (29). Indeed, the latter domain is more highly conserved than the actual downstream functional terpene synthase region, facilitating ready detection of generic plant terpene synthase sequences by application of the BLAST algorithm for local sequence alignment to deduced protein sequences from plant protein databases. By contrast, bacterial terpene synthases not only lack the characteristic conserved N-terminal domain (PF01397) of plant synthases but also, the overall amino acid sequences of common microbial terpene synthases, such as geosmin synthase, *epi*-isozizaene synthase, and 2-methylisoborneol synthase, show very low levels of overall sequence similarity to any other known bacterial terpene synthase. Our first successful trial using HMMs and Pfam searching revealed the existence of several monoterpene synthases of bacterial origin, including 2-methylisoborneol synthase. Using a second generation HMM generated from 41 bacterial terpene synthases that had been identified in the first round of bacterial database screening revealed an expanded pool of over 140 candidate terpene synthases. In the meantime, the number of bacterial genome sequences has continued to grow at an increasing rate. We have now constructed a third generation HMM trained by over 140 terpene synthases that has then been used for a Pfam search of the most recent bacterial protein sequences available in the public databases as well as in-house draft genome sequences.

Of a total of 8,759,463 predicted bacterial proteins, 262 proteins were selected as possible terpene synthases, with E-value ranges from 5.3×10^{-7} to 6.9×10^{-207} [by contrast, screening of the expanded databases with the second generation HMM profile (30) resulted in E-value ranges for the same proteins of 1.1×10^{-1} to 2.9×10^{-258}]. The candidate terpene synthase protein sequences harvested in the new search were then aligned

and subjected to phylogenetic analysis based on the resultant sequence alignment (Fig. 2). Not unexpectedly, the predicted terpene synthases were mainly found in *Actinomycetales* microorganisms, reflecting presumably the significant numbers of *Actinomycetales* genome sequences in the current public genome databases. Additional predicted terpene synthases were also found in gram-negative bacteria belonging to the orders Myxococcales, Oscillatoriales, Nostocales, Burkholderiales, Herpetosiphonales, Rhizobiales, Chlamydiales, Flavobacteriales, Chromatiales, Ktedonobacterales, Sphingobacteriales, and Pseudomonadales. Notably, no predicted terpene synthases were evident in the genome data obtained from both low-GC content gram-positive bacteria (phylum Firmicutes) and microorganisms belonging to the superkingdom Archaea.

Application of the first generation HMM (PF03936) to 1,922,990 bacterial proteins had originally revealed 41 predicted terpene synthases that were classified by phylogenetic analysis into three main groups consisting of monoterpene, sesquiterpene, and diterpene synthases (27). By contrast, the phylogenetic tree derived from the newly identified 262 bacterial terpene synthases uncovered by the third generation HMM model is considerably more complex, exhibiting a much higher degree of branching, thus indicating that bacterial terpene synthases are far more structurally diverse than previously indicated. The majority of 262 presumptive terpene synthases is most likely sesquiterpene synthases. Within this group, two major clades correspond to geosmin synthases and *epi*-isozizaene synthases located in upper sector (blue-shaded area) and right sector (orange-shaded area), respectively, in Fig. 2. A third major clade that corresponds to 2-methylisoborneol synthases is evident in the left sector in Fig. 2, green-shaded area. Geosmin synthases are found in almost all *Actinomycetales*, except for the geosmin nonproducing species *S. roseosporus*. Three geosmin nonproducing species (*S. clavuligerus*, *S. lactacystinaeus*, and *Streptomyces* sp. SirexAA-E) each harbor presumptively silent genes that are predicted by phylogenetic analysis to encode geosmin synthases SCLAV_0159, SLT2_649, and SACTE_0245, respectively (Fig. 2). The observed inability to produce geosmin might simply reflect extremely low levels of geosmin gene expression in each microorganism or competing consumption of the common precursor FPP by the synthesis of both (+)-caryolan-1-ol and *epi*-cubenol in *Streptomyces* sp. SirexAA-E. Several *Streptomyces* species harbor an orthologous *epi*-isozizaene synthase (orange-shaded clade in Fig. 2). Two *Streptomyces* species, *S. albus* and *S. cyaneogriseus* subsp. *noncyanogenus*, that each produced albaflavenone along with geosmin (Fig. S1) harbor orthologous *epi*-isozizaene synthases XNR_1580 and SCYA_00859, respectively.

The sesquiterpenoid antibiotic pentalenolactone is a commonly occurring metabolite that has been isolated from more than 30 species of *Streptomyces* (31, 32). The first step in pentalenolactone biosynthesis is the cyclization of FPP to pentalenene, the parent hydrocarbon of the pentalenolactone family of metabolites. Pentalenene synthase (ADO85594) of *Streptomyces exfoliatus* UC5319 was, in fact, the first characterized, cloned, and sequenced terpene synthase from *Streptomyces* (33, 34). The genome sequence of *S. avermitilis* has been shown to harbor a biosynthetic gene cluster for a closely related metabolite neopentalenolactone. This cluster has been genetically and biochemically characterized in detail, including the demonstration of the presence of an orthologous pentalenene synthase (SAV_2998) (35, 36). Pentalenene synthases of known sequence are restricted to a relatively small clade within a branch distance of 0.1 K_{aa} (yellow-shaded right sector in Fig. 2) that includes SBI_09679 (*Streptomyces bingchenggensis*), SCYA_02107 (*S. cyaneogriseus* subsp. *noncyanogenus*), and B446_29695 (*Streptomyces collinus* Tü 365). The genes encoding the latter three orthologous pentalenene synthases are each found at analogous positions within biosynthetic gene clusters in their parent

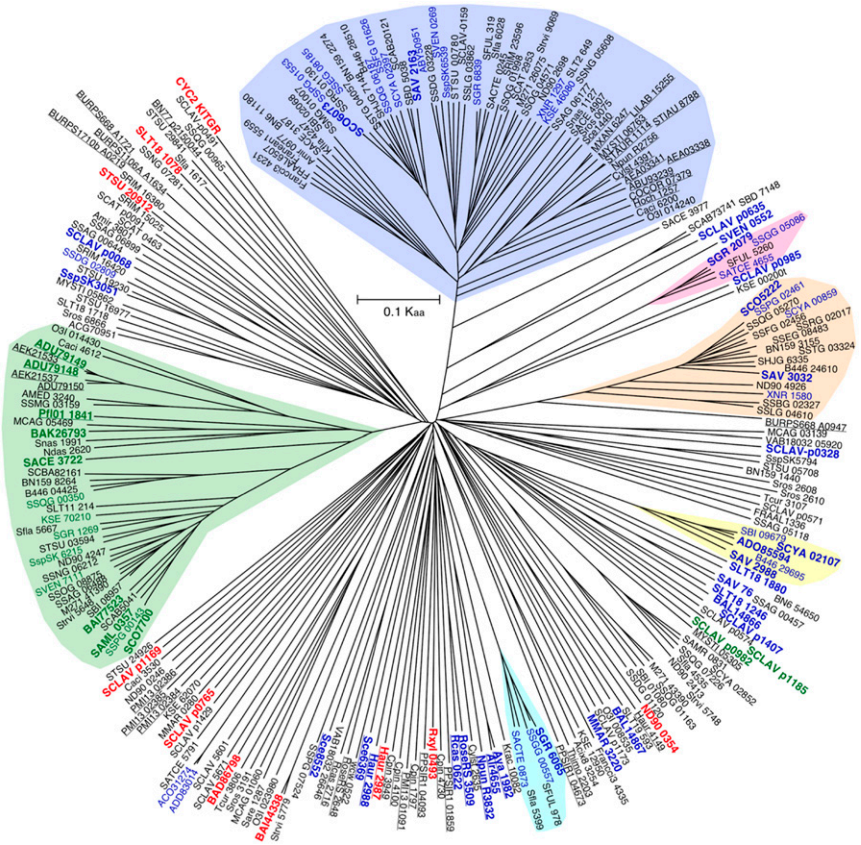

Fig. 2. Phylogenetic analysis of presumptive terpene synthases from bacterial databases. Monoterpene, sesquiterpene, and diterpene synthases are indicated in green-, blue-, and red-colored characters, respectively. The biochemical functions of terpene synthases written in bold and underlined were confirmed by these studies and published data, and they were found from gram-negative bacteria. The blue, orange, and green zones indicate geosmin, *epi*-isozizaene, and 2-methylisoborneol synthases, respectively. Locus tags or accession numbers are described in *SI Materials and Methods*.

microorganism that correspond to either the pentalenolactone (*pen*) cluster of *S. exfoliatus* (37) or the neopentalenolactone (*ptl*) cluster of *S. avermitilis* (35). Heterologous expression in an engineered *Streptomyces* host of the *pen*-like biosynthetic gene cluster of *S. cyaneogriseus* subsp. *noncyanogenus* (AB981710) resulted in production of pentalenolactone (38). The *pen* clusters of *S. exfoliatus*, *S. bingchenggensis*, and *S. cyaneogriseus* subsp. *noncyanogenus* each contain two cytochrome P450 genes, including orthologs of PenM, the P450 that catalyzes the last step in pentalenolactone biosynthesis (a rare oxidative rearrangement) (39). However, the corresponding *ptl* biosynthetic gene clusters of *S. avermitilis* and *S. collinus* both lack a homolog of PenM, suggesting that *S. collinus* most likely produces neopentalenolactone, which has been established for *S. avermitilis*. These results support the inference that presumptive terpene synthases that are located within a clade within the branch distance of 0.1 K_{aa} most likely catalyze formation of a common terpene cyclization product.

Two additional sesquiterpene synthases, SGR_2079 (upper-right, red-shaded clade in Fig. 2) and SGR_6065 (lower light blue-shaded clade in Fig. 2), of *S. griseus* have previously been shown to be (+)-caryolan-1-ol synthase (40) and *epi*-cubenol synthase (41), respectively. The former clade also contains SSGG_05086 (*S. roseosporus*) and SACTE_4655 (*Streptomyces* sp. SirexAA-E), whereas the latter includes the orthologous *epi*-cupenol synthases SSGG_00557 and SACTE_0873. Indeed, we have shown that these two microorganisms produce both (+)-caryolan-1-ol and *epi*-cubenol (Fig. S1).

Heterologous Expression of Genes Encoding Bacterial Terpene Synthases. The harvesting of more than 260 presumptive terpene synthases from bacterial genome data using a third generation HMM profile has revealed that the respective terpene synthase genes are widely distributed in this kingdom. Among these synthases, the general biochemical function can be provisionally assigned with reasonable certainty by detailed phylogenetic analysis based on sequence alignment with characterized terpene synthases (as described above). Nonetheless, the assignment of the actual biochemical function of each presumptive synthase, including the identification of both the relevant acyclic allylic diphosphate substrate and the resultant cyclization product, still requires direct experimental verification. In many cases, such biochemical characterization has been affected by direct incubation of purified recombinant enzymes with the appropriate acyclic allylic diphosphate substrate in the presence of a divalent cation (usually Mg^{2+}) and identification of the resultant enzymatic cyclization products. Such in vitro methods, however, while providing high quality data, are subject to a number of commonly encountered drawbacks. Although many known terpenes can be readily identified by GC-MS analysis and compared with extensive libraries of MS spectra and retention indices, spectroscopic identification of new or previously characterized compounds that are not in the MS databases requires significantly larger quantities of terpene products that are difficult to obtain in a typical enzymatic incubation. Moreover, expression in *Escherichia coli* of heterologous bacterial terpene synthases frequently results in the formation of insoluble inclusion bodies. Thus, although recombinant *S. avermitilis* pentalenene synthase (SAV_2998) could be expressed in *E. coli* in soluble form, the three remaining *S. avermitilis* terpene synthases [geosmin synthase (SAV_2163) (42), *epi*-isozizaene synthase (SAV_3032) (43), and avermitilol synthase (SAV_76) (44)] were all obtained as inclusion bodies when expressed in *E. coli*. In these cases, the geosmin and *epi*-isozizaene synthases could fortunately be refolded to obtain soluble proteins that retained the requisite terpene synthase activity, whereas soluble avermitilol synthase could be expressed in *E. coli* using a synthetic gene with codons optimized for expression in *E. coli*.

We have reported an alternative experimental system based on a versatile heterologous expression system that uses an engineered *S. avermitilis* host from which the native natural product biosynthetic genes have been deleted. The resultant mutant host has been shown to be particularly suited for the production of a wide variety of natural products (27, 38, 45, 46). This versatile engineered host system, which has been stripped of all of its native terpene synthase genes and is, thus, unable to produce endogenous terpenoid metabolites that might otherwise interfere with production and analysis of heterologously expressed terpenes and their derivatives, is ideally suited to the generation of exogenous terpenoid metabolites in quantities sufficient for their rigorous spectroscopic analysis. Using this method, individual genes encoding monoterpene, sesquiterpene, or diterpene synthases, including their native or heterologous ribosome-binding sequences, can be inserted downstream of the appropriate genes for GPP synthase (*gps*), FPP synthase (*fps*), or GGPP synthase (*ggs*) under the control of a strong, constitutively expressed promoter (*rpsJp*) harbored in the integrating vector pKU1021 (46). Coexpression of the gene for the relevant polyprenyl diphosphate synthase (Fig. S2) significantly enhances the production of the target terpenoid metabolite(s). The resulting levels of terpene metabolite production have been found to be independent of the arrangement of the genes for the polyprenyl diphosphate synthase and the terpene synthase in the engineered biosynthetic expression system (Fig. S2).

We have been especially interested in determining the biochemical function of presumptive terpene synthases that are classified by phylogenetic analysis into isolated clades that do not contain any other synthases of previously assigned function. This group includes the substantial number of predicted terpene synthases of the apparently terpene nonproducing strains *S. lactacystinaeus* and *S. clavuligerus*, the latter of which harbors at least 17 distinct synthases. Of special interest are also terpene synthases other than the geosmin synthase and 2-methylisoborneol synthase found in gram-negative bacteria, such as *Cyanobacteria* and *Proteobacteria*, because only a handful of terpene synthases have previously been characterized from these organisms. Table S1 summarizes the terpenoid metabolites that we have now identified by expression of heterologous terpene synthases in the engineered *S. avermitilis* host. Almost all transformants produced terpenoid metabolites, among which sesquiterpenes were the major products along with a variety of monoterpenes and diterpenes as minor metabolites. Interestingly, a minority of such transformants failed to produce detectable terpenes using any of three different expression vectors. It is, of course, possible that such synthases are intrinsically dysfunctional; alternatively, they may have altogether different catalytic functions. The terpenes that were produced by heterologous expression all accumulated in the mycelium, except for 1,8-cineole, which was generated by expression of *sclav_p0982* using the vector pKU1021*gps*. SCLAV_p0982 has previously been shown directly by in vitro incubation to catalyze the cyclization of GPP to 1,8-cineole (47). Expression of *sclav_p0982* in the heterologous *Streptomyces* host established that the expressed terpene synthase generates not only 1,8-cineole but also, β-pinene and camphene. Similar results were also obtained by heterologous expression of *sgr_6065*. SGR_6065 has previously been characterized as *epi*-cubenol synthase by in vitro analysis (41), which was illustrated in the production profile (Fig. S1). Expression of *sgr_6065* in the heterologous *Streptomyces* host established that SGR_6065 generates not only *epi*-cubenol but also, a number of mechanistically related sesquiterpenoid metabolites, including α-muurolene and cadinadienes (Table S1). SCLAV_p0068 has previously been characterized as (+)-T-muurolol synthase by both enzymatic reaction using the recombinant protein and in vivo expression in an engineered *E. coli* host (48). Consistent with these results, we have found that expression of *sclav_p0068* in the heterologous *S. avermitilis* host also produced the same sesquiterpene alcohol. Curiously, coexpression of the native upstream *S. clavuligerus* gene encoding a cytochrome P450 (SCLAV_p0067) along with *sclav_p0068*

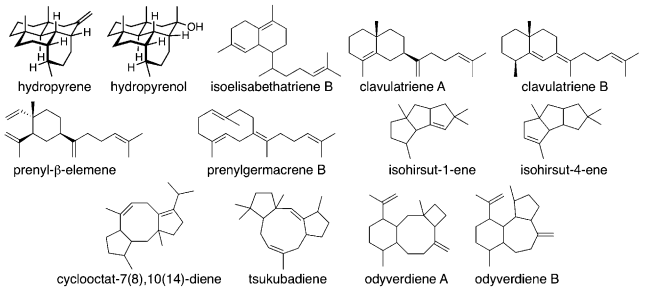

Fig. 3. The structures of newly identified sesquiterpenes and diterpenes produced by heterologous expression of genes encoding terpene synthases derived from *Streptomyces* microorganisms in engineered *S. avermitilis* SUKA22.

led to the production of the structurally and mechanistically unrelated sesquiterpene (−)-drimenol. This result suggests that SCLAV_p0067 may somehow alter the cyclization of FPP mediated by SCLAV_p0068, but the cyclization mechanism has not yet been proven. More than one-half of presumptive terpene synthases harbored by the terpene nonproducing *S. clavuligerus* could be expressed in the heterologous *Streptomyces* host, giving rise to identifiable monoterpene, sesquiterpene, and diterpene hydrocarbons and alcohols. Similarly, when several presumptive terpene synthases from the apparently terpene nonproducing *S. lactacystinaeus* were heterologously expressed in *S. avermitilis*, one-half of these were biochemically functional, which was evidenced by the production of a variety of sesquiterpene and diterpene products. Interestingly, the majority of the presumptive terpene synthases derived from gram-negative bacteria segregated into isolated clades. Two such sesquiterpene synthases, Alr4685 and Npun_R3832, of *Nostoc* sp. PCC 7120 and *Nostoc punctiforme* PCC 73102, respectively, have previously been characterized in vitro (49). Consistent with these observations, heterologous *Streptomyces* expression of the corresponding sesquiterpene synthases gave rise to formation of sesquiterpene products.

The majority of the terpene hydrocarbons that we have identified by heterologous expression of bacterial terpene synthases in the engineered *Streptomyces* host is known compounds that have previously been isolated from fungal, plant, or in some cases, bacterial sources. Significantly, however, we have also isolated and determined the structures of 13 previously unidentified cyclic terpenes, none of which have previously been described from fungal, plant, or bacterial sources (Fig. 3). Three such presumptive terpene synthases (SCLAV_p0765, SCLAV_p1169, and SCLAV_1407) from the terpene nonproducing organism *S. clavuligerus* turn out to have unique catalytic activities. Thus, the diterpene alcohol hydropyrenol as well as the diterpene hydrocarbons hydropyrene and isoelisabethatriene B were all isolated from the standard heterologous *Streptomyces* host carrying the *sclav_p0765* gene. The former two metabolites are unique tetracyclic diterpenes with a parent skeleton that has not previously been reported in the same or modified form, whereas isoelisabethatriene B is a previously undescribed isomer of the known sea plume metabolite isoelisabethatriene (50). In like manner, none of the diterpene hydrocarbons clavulatrienes A and B, prenyl-β-elemene, and prenylgermacrene B, all of which are produced by a heterologous *Streptomyces* host carrying *sclav_p1169*, have been previously described, whereas the cometabolite lobophytumin C has previously been reported only as a soft coral metabolite (51). The sesquiterpene hydrocarbon β-elemene is known to be a common artifact of GC-MS analysis arising from thermal Cope rearrangement of germacrene A at elevated temperature. Prenyl-β-elemene is, thus, likely to be an artifact arising from rearrangement of the primary cyclization product prenylgermacrene. Although the terpene synthases SCLAV_p1407 of *S. clavuligerus* and SLT18_1880 of *S. lactacystinaeus* were classified into different clades, expression of each gene in the heterologous *Streptomyces* host produced isomeric linear triquinane sesquiterpene hydrocarbons that were similar to the known fungal cucumin sesquiterpenes and that differed from each other only in the position of one of the double bonds (52). The production of linear triquinane sesquiterpenes by bacteria has not been reported to date. Another presumptive terpene synthase (SLT18_1078) from *S. lactacystinaeus* was found to catalyze the cyclization of GGPP to cyclooctat-7(8),10(14)-diene. The closely related diterpene alcohol cyclooctat-9-ene-7-ol has been reported to be formed by in vitro cyclization of GGPP catalyzed by CotB2 derived from the cyclooctatine-producing *Streptomyces melanosporofaciens* MI614-43F2 (53). The three diterpene hydrocarbons (tsukubadiene and odyverdienes A and B) produced by heterologous *Streptomyces* hosts carrying *stsu_20912* and *nd90_0354*, respectively, are each unique structures with parent diterpene skeletons that have not been found in any organism to date. In a few cases, the low levels of heterologous metabolite production have so far prevented precise assignment of previously unidentified structures when these did not match the GC-MS properties of known compounds in the MS databases. Complete assignment of these previously unidentified structures will require improvements in the levels of production in the heterologous expression system.

In 2008, we first described the results of a Pfam search of predicted proteins in the current bacterial genome databases using an HMM that was based on a group of aligned protein sequences of plant terpene synthases, resulting in the discovery of a large number of previously unrecognized bacterial terpene synthases (27). Thereafter, another Pfam search of the expanding bacterial genome databases using a second generation HMM based exclusively on 41 newly revealed bacterial terpene synthases uncovered 140 presumptive terpene synthases (30). We have now extended these results using a newly trained third generation HMM, which was generated from the previously recognized 140 presumptive terpene synthases to thereby reveal 262 presumptive terpene synthases of 8,759,463 predicted proteins from the combined public and in-house databases of bacterial genome sequences. Although the majority of the newly recognized terpene synthases is associated with *Actinomycetales* microorganisms, a minor but still significant fraction was found in gram-negative bacteria, mostly belonging to *Cyanobacteria* and *Myxobacteria*. Genes encoding numerous presumptive terpene synthases have been efficiently expressed in a heterologous host derived from an engineered *S. avermitilis* from which the endogenous natural product biosynthetic genes had been deleted, leading to the production and identification of numerous terpene hydrocarbons and alcohols, most of which had not previously been observed in bacterial cultures. Although the majority of the terpenes produced were identical to known fungal or plant terpenes, to date, we have already identified and structurally characterized 13 previously unidentified cyclic terpenes, including 2 sesquiterpenes and 11 diterpenes, using the method of heterologous *Streptomyces* expression. These studies suggest that genes encoding terpene synthases are widely distributed in bacteria and that these terpene synthase genes represent a fertile source for discovery of new natural products.

Materials and Methods

Bacterial Strains and Plasmid Vectors. *S. venezuelae* ATCC 10712 [Japan Collection of Microorganisms (JCM) 4526], *S. ghanaensis* ATCC 14672 (JCM 4963), *S. pristinaespiralis* ATCC 25486 (JCM 4507), *S. sviceus* ATCC 29083 (JCM 4929), and *S. viridochromogenes* DSM 40736 (JCM 4977) were obtained from the culture collection of the RIKEN Bioresource Center. *S. cyaneogriseus* subsp. *noncyanogenus* NRRL 15773 and *S. roseosporus* NRRL 11379 were from the Agriculture Research Service Culture Collection (NRRL). *Streptomyces* sp. SirexAA-E was distributed by Cameron Currie, University of Wisconsin, Madison, WI. Other *Streptomycetaceae* microorganisms were used in our laboratory stock cultures. *Streptomyces* integrating vectors carrying *attPlint* of bacteriophage ϕC31 [pKU1021*gps* (AB982124), pKU1021*fps* (AB982125), and pKU1021*ggs* (AB982126)] were used for the expression of genes encoding monoterpene, sesquiterpene, and diterpene synthases in *S. avermitilis* SUKA22. Cloning and transformation of terpene synthase

MICROBIOLOGY

genes are described in *SI Materials and Methods* and primers for PCR amplification are listed in Table S2.

Bacterial Cultivation and Analysis of Terpenoid Metabolites. The bacterial growth conditions for the isolation of terpenes have been described previously (30, 31). The *n*-hexane extracts were directly subjected to GC-MS analysis [Shimadzu GC-17A; 70 eV, electron ionization, positive ion mode; 30 m × 0.25 mm InterCap 5 capillary column [diphenyl (5%)-dimethyl (95%) polysiloxane; GL Sciences] using a temperature program of ∼50–280 °C and a temperature gradient of 20 °C/min]. Compounds produced were evaluated by comparison with the mass spectra of the corresponding reference compounds in the National Institute of Standards and Technology (NIST)/United States Environmental Protection Agency (EPA)/National Institutes of Health (NIH) MS Library (2014 version). The generation of previously unidentified terpene products is described in *SI Materials and Methods*, whereas the NMR structure elucidation of all previously unidentified compounds is described in another work (54). The large-scale preparation of terpenoid metabolites is described in *SI Materials and Methods*.

ACKNOWLEDGMENTS. We thank C. Currie (University of Wisconsin) for providing a strain of *Streptomyces* sp. SirexAA-E. This work was supported by a Research Grant-in-Aid for "Project focused on developing key technology of discovering and manufacturing drug for next-generation treatment and diagnosis" from the Ministry of Economy, Trade and Industry of Japan (to K.S.-y. and H.I.), US National Institutes of General Medical Sciences Grant GM30301 (to D.E.C.), and a Research Grant-in-Aid for Scientific Research on Innovative Areas from the Ministry of Education, Culture, Sports, Science and Technology of Japan (to H.I.).

1. Berthelot M, André G (1891) Sur l'odeur proper de la terre. *C R Acad Sci* 112(1):598–599.
2. Gerber NN, Lechevalier HA (1965) Geosmin, an earthly-smelling substance isolated from actinomycetes. *Appl Microbiol* 13(6):935–938.
3. Gerber NN (1967) Geosmin, an earthy-smelling substance isolated from actinomycetes. *Biotechnol Bioeng* 9(3):321–327.
4. Gerber NN (1969) A volatile metabolite of actinomycetes, 2-methylisoborneol. *J Antibiot (Tokyo)* 22(10):508–509.
5. Gerber NN (1971) Sesquiterpenoids from Actinomycetes: Cadin-4en-1-ol. *Phytochemistry* 10(1):185–189.
6. Gerber NN (1972) Sesquiterpenoids from actinomycetes. *Phytochemistry* 11(1):385–388.
7. Gerber NN (1973) Volatile lactones from *Streptomyces*. *Tetrahedron Lett* 14(10):771–774.
8. Pollak FC, Berger RG (1996) Geosmin and related volatiles in bioreactor-cultured *Streptomyces citreus* CBS 109.60. *Appl Environ Microbiol* 62(4):1295–1299.
9. Jáchymová J, Votruba J, Víden I, Rezanka T (2002) Identification of *Streptomyces* odor spectrum. *Folia Microbiol (Praha)* 47(1):37–41.
10. Wilkins K, Schöller C (2009) Volatile organic metabolites from selected *Streptomyces* strains. *Actinomycetologica* 23(2):27–33.
11. Schöller CE, Gürtler H, Pedersen R, Molin S, Wilkins K (2002) Volatile metabolites from actinomycetes. *J Agric Food Chem* 50(9):2615–2621.
12. Dickschat JS, Wenzel SC, Bode HB, Müller R, Schulz S (2004) Biosynthesis of volatiles by the myxobacterium *Myxococcus xanthus*. *ChemBioChem* 5(6):778–787.
13. Dickschat JS, Bode HB, Wenzel SC, Müller R, Schulz S (2005) Biosynthesis and identification of volatiles released by the myxobacterium *Stigmatella aurantiaca*. *ChemBioChem* 6(11):2023–2033.
14. Dickschat JS, Martens T, Brinkhoff T, Simon M, Schulz S (2005) Volatiles released by a *Streptomyces* species isolated from the North Sea. *Chem Biodivers* 2(7):837–865.
15. Schulz S, Dickschat JS (2007) Bacterial volatiles: The smell of small organisms. *Nat Prod Rep* 24(4):814–842.
16. Rabe P, Citron CA, Dickschat JS (2013) Volatile terpenes from actinomycetes: A biosynthetic study correlating chemical analyses to genome data. *ChemBioChem* 14(17):2345–2354.
17. Gerber NN (1979) Volatile substances from actinomycetes: Their role in the odor pollution of water. *CRC Crit Rev Microbiol* 7(3):191–214.
18. Izaguirre G, Hwang CJ, Krasner SW, McGuire MJ (1982) Geosmin and 2-methylisoborneol from cyanobacteria in three water supply systems. *Appl Environ Microbiol* 43(3):708–714.
19. Jüttner F (1987) Volatile organic substances. *The Cyanobacteria*, eds Fay B, Baalen CV (Elsevier, Amsterdam), pp 453–469.
20. Wu J-T, Jüttner F (1988) Differential partitioning of geosmin and 2-methylisoborneol between cellular constituents in *Oscillatoria tenuis*. *Arch Microbiol* 150(6):580–583.
21. Dickschat JS, et al. (2007) Biosynthesis of the off-flavor 2-methylisoborneol by the myxobacterium *Nannocystis exedens*. *Angew Chem Int Ed Engl* 46(43):8287–8290.
22. Mattheis JP, Roberts RG (1992) Identification of geosmin as a volatile metabolite of Penicillium expansum. *Appl Environ Microbiol* 58(9):3170–3172.
23. Börjesson T, Stöllman U, Schnürer J (1993) Off odors compounds produced by molds on oatmeal agar: Identification and relation to other growth characteristics. *J Agric Food Chem* 41(11):2104–2111.
24. Larsen TO, Frisvad JC (1995) Characterization of volatile metabolites from 47 *Penicillium* taxa. *Mycol Res* 99(10):1153–1166.
25. Gürtler H, et al. (1994) Albaflavenone, a sesquiterpene ketone with a zizaene skeleton produced by a streptomycete with a new rope morphology. *J Antibiot (Tokyo)* 47(4):434–439.
26. Finn RD, et al. (2010) The Pfam protein families database. *Nucleic Acids Res* 38(Database issue):D211–D222.
27. Komatsu M, Tsuda M, Omura S, Oikawa H, Ikeda H (2008) Identification and functional analysis of genes controlling biosynthesis of 2-methylisoborneol. *Proc Natl Acad Sci USA* 105(21):7422–7427.
28. Giglio S, Chou WKW, Ikeda H, Cane DE, Monis PT (2011) Biosynthesis of 2-methylisoborneol in cyanobacteria. *Environ Sci Technol* 45(3):992–998.
29. Whittington DA, et al. (2002) Bornyl diphosphate synthase: Structure and strategy for carbocation manipulation by a terpenoid cyclase. *Proc Natl Acad Sci USA* 99(24):15375–15380.
30. Yamada Y, Cane DE, Ikeda H (2012) Diversity and analysis of bacterial terpene synthases. *Natural Product Biosynthesis by Microorganisms and Plants, Part A. Methods in Enzymology*, ed Hopwood DA (Elsevier Inc/Academic Press, New York), Vol 515, pp 123–166.
31. Koe BK, Sobin BA, Celmer WD (1956–1957) PA 132, a new antibiotic. I. Isolation and chemical properties. *Antibiot Annu* 1956-1957(1956-1957):672–675.
32. Takahashi S, Takeuchi M, Arai M, Seto H, Otake N (1983) Studies on biosynthesis of pentalenolactone. V isolation of deoxypentalenylglucuron. *J Antibiot (Tokyo)* 36(3):226–228.
33. Cane DE, Abell C, Tillman AM (1984) Pentalenene biosynthesis and the enzymatic cyclization of farnesyl pyrophosphate. Proof that the cyclization is catalyzed by a single enzyme. *Bioorg Chem* 12(4):312–328.
34. Lesburg CA, Zhai G, Cane DE, Christianson DW (1997) Crystal structure of pentalenene synthase: Mechanistic insights on terpenoid cyclization reactions in biology. *Science* 277(5333):1820–1824.
35. Tetzlaff CN, et al. (2006) A gene cluster for biosynthesis of the sesquiterpenoid antibiotic pentalenolactone in *Streptomyces avermitilis*. *Biochemistry* 45(19):6179–6186.
36. Jiang J, 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(27):6431–6440.
37. Seo M-J, Zhu D, Endo S, Ikeda H, Cane DE (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(10):1739–1754.
38. Ikeda H, Shin-ya K, Omura S (2014) Genome mining of the *Streptomyces avermitilis* genome and development of genome-minimized hosts for heterologous expression of biosynthetic gene clusters. *J Ind Microbiol Biotechnol* 41(2):233–250.
39. Zhu D, Seo M-J, Ikeda H, Cane DE (2011) Genome mining in *streptomyces*. Discovery of an unprecedented P450-catalyzed oxidative rearrangement that is the final step in the biosynthesis of pentalenolactone. *J Am Chem Soc* 133(7):2128–2131.
40. Nakano C, Horinouchi S, Ohnishi Y (2011) Characterization of a novel sesquiterpene cyclase involved in (+)-caryolan-1-ol biosynthesis in *Streptomyces griseus*. *J Biol Chem* 286(32):27980–27987.
41. Nakano C, Tezuka T, Horinouchi S, Ohnishi Y (2012) Identification of the SGR6065 gene product as a sesquiterpene cyclase involved in (+)-epicubenol biosynthesis in *Streptomyces griseus*. *J Antibiot (Tokyo)* 65(11):551–558.
42. Cane DE, 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. *J Antibiot (Tokyo)* 59(8):471–479.
43. Takamatsu S, et al. (2011) Characterization of a silent sesquiterpenoid biosynthetic pathway in *Streptomyces avermitilis* controlling *epi*-isozizaene albaflavenone biosynthesis and isolation of a new oxidized *epi*-isozizaene metabolite. *Microb Biotechnol* 4(2):184–191.
44. Chou WKW, et al. (2010) Genome mining in *Streptomyces avermitilis*: Cloning and characterization of SAV_76, the synthase for a new sesquiterpene, avermitilol. *J Am Chem Soc* 132(26):8850–8851.
45. Komatsu M, Uchiyama T, Omura S, Cane DE, Ikeda H (2010) Genome-minimized *Streptomyces* host for the heterologous expression of secondary metabolism. *Proc Natl Acad Sci USA* 107(6):2646–2651.
46. Komatsu M, et al. (2013) Engineered *Streptomyces avermitilis* host for heterologous expression of biosynthetic gene cluster for secondary metabolites. *ACS Synth Biol* 2(7):384–396.
47. Nakano C, Kim H-K, Ohnishi Y (2011) Identification of the first bacterial monoterpene cyclase, a 1,8-cineole synthase, that catalyzes the direct conversion of geranyl diphosphate. *ChemBioChem* 12(13):1988–1991.
48. Hu Y, Chou WKW, Hopson R, Cane DE (2011) Genome mining in *Streptomyces clavuligerus*: Expression and biochemical characterization of two new cryptic sesquiterpene synthases. *Chem Biol* 18(1):32–37.
49. Agger SA, Lopez-Gallego F, Hoye TR, Schmidt-Dannert C (2008) Identification of sesquiterpene synthases from *Nostoc punctiforme* PCC 73102 and *Nostoc* sp. strain PCC 7120. *J Bacteriol* 190(18):6084–6096.
50. Kohl AC, Kerr RG (2003) Pseudopterosin biosynthesis: Aromatization of the diterpene cyclase product, elisabethatriene. *Mar Drugs* 1(1):54–65.
51. Li L, et al. (2011) Diterpenes from the Hainan soft coral *Lobophytum cristatum* Tixier-Durivault. *J Nat Prod* 74(10):2089–2094.
52. Hellwig V, et al. (1998) New triquinane-type sesquiterpenoids from *Macrocystidia cucumis* (Basidiomycetes). *Eur J Org Chem* 1998(1):73–79.
53. Kim S-Y, et al. (2009) Cloning and heterologous expression of the cyclooctatin biosynthetic gene cluster afford a diterpene cyclase and two p450 hydroxylases. *Chem Biol* 16(7):736–743.
54. Yamada Y, et al. (2014) Novel terpenes generated by heterologous expression of bacterial terpene synthase genes in an engineered *Streptomyces* host. *J Antibiot (Tokyo)*, in press.