World Journal of Microbiology and Biotechnology (2020) 36:142
https://doi.org/10.1007/s11274-020-02918-y

REVIEW

Bioinformatics approach to understand nature's unified mechanism of stereo-divergent synthesis of isoprenoid skeletons

Pranav Bhaskar[1] · **Dipti Sareen**[1]

Received: 25 May 2020 / Accepted: 13 August 2020 / Published online: 27 August 2020

Abstract

In isoprenoid metabolism, cyclisation is the important gateway to chemical diversity. Terpene synthase is responsible for the cyclisation of a few universal substrates forming hundreds of often stereo-chemically complex mono- and poly-cyclic terpene hydrocarbons with a broad spectrum of functions in pharmaceuticals, flavours and fragrance industry. Although they are discovered and characterised mainly from plants and fungi, yet only a small share of bacterial terpenes has been investigated so far owing to their low level of expression in wild-type microorganisms. Extensive bacterial genome mining has revealed a treasure trove of terpene synthase genes and their regulated heterologous overexpression has pitched-in to describe the biochemical function of putative genes and sequester new terpene metabolites. This review deals with the modern genome mining techniques and molecular methods, providing more experimental tools for studying the structure and functions of terpenoid metabolites and strongly supports the idea that genome mining is a utile approach in deciphering the terpenoid diversity in bacteria.

Declaration This is to confirm that this manuscript has not been published elsewhere and is not under consideration by any other journal. Both the authors have approved the manuscript with full text, number of figures/tables and agree with its submission to the journal.

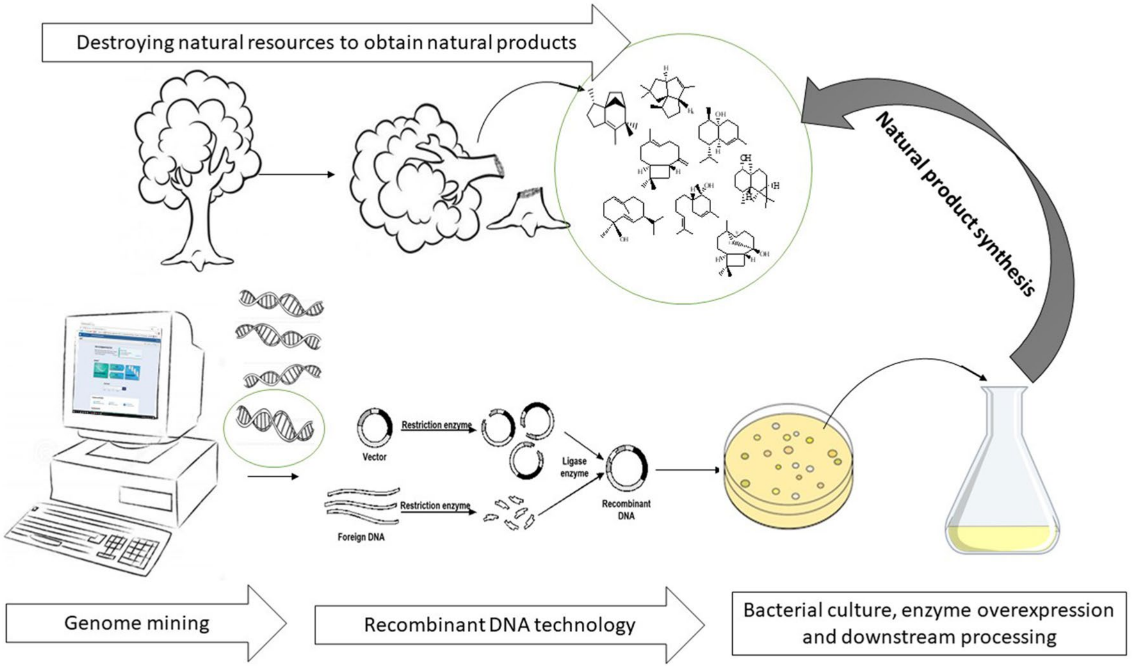
Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11274-020-02918-y) contains supplementary material, which is available to authorized users.

✉ Dipti Sareen
diptsare@pu.ac.in

Pranav Bhaskar
pranavbhaskar@live.com

[1] Department of Biochemistry, Panjab University, Chandigarh 160014, India

Graphic abstract



Keywords Actinobacteria · Bacterial terpenes · Genome mining · Heterologous expression · Terpene synthase

Abbreviations

PKS	Polyketide synthase
NRPS	Non-ribosomal peptide synthetase
TC	Terpene cyclase
BGC	Biosynthetic gene cluster
TPS	Terpene synthase
PS	Pentalenene synthase
IPP	Isopentenyl diphosphate
DMAPP	Dimethylallyl diphosphate
MVA	Mevalonic acid
DOXP	Deoxyxylulose phosphate
GPP	Geranyl diphosphate
FPP	Farnesyl diphosphate
GGPP	Geranyl geranyl diphosphate
SC	Sesquiterpene cyclase
HMM	Hidden Markov model
BLASTp	Basic Local Alignment Sequence Tool for proteins
2-MIB	2-Methylisoborneol
SAM	*S*-Adenosyl-L-methionine
2-MeGPP	2-Methyl geranyl diphosphate
2-MB	2-Methylbornane
LPP	Linalyl diphosphate
GS	Geosmin synthase
EIZS	*epi*-Isozizaene synthase
GC/MS	Gas chromatography mass spectroscopy
CFE	Cell free extract
VOC	Volatile organic compounds
NMR	Nuclear magnetic resonance
DTS	Diterpene synthases
TDP	Terpentetrienedienyl diphosphate
TTE	Terpentetriene
ORF	Open reading frame
CPP	Copalyl diphosphate
Bds	Bonnadiene synthase
Pms	Phomopsene synthase
Sas	Spiroalbatene synthase
CAS	Cembrene A synthase

Introduction

Isoprenoids (terpenoids or terpenes), occupy a special space among the principal metabolites, being the most structurally diverse and the largest group of natural compounds commonly produced in plants (Sacchettini and Poulter 1997). They have a wide range of physiological properties and have a myriad of applications. Isoprenoids comprise natural rubber; flavour compounds like limonene, menthol and camphor; and six drug classes, including steroids, artemisinins, ingenoids, toxoids, cannabinoids and tocopherols (Roach and Shenvi 2015); as well as linear isoprenoid pyrophosphates involved in post-translational protein modification (Knorre et al. 2009).

Terpenoids are, in general, regarded as plant or fungal metabolites; but only a few have been commercially used because of the inefficient and economically unprecedented production in their native hosts. Also, extensive extraction from plants leads to their death and extinction. It is, therefore, commercially formidable to harvest terpenoids from plant sources due to the requirement of large natural resources for obtaining substantial quantities, and also the associated ecological and environmental impacts, making it necessary to find an alternative source of such compounds. Taking taxol, a blockbuster anticancer agent, as an illustrative example, its chemical synthesis has proven difficult and expensive (Wender et al. 1997a, b) due to the structural complexity and hence, a semi-synthetic approach was adopted from the biotechnologically produced taxol precursors (Wuts 1998). However, microbial biosynthesis offers an opportunity for a more reliable, intensive and efficient way of terpenoid production in a renewable manner.

Until recently, the bacteria were wrongly perceived to be incapable of producing terpenoids due to lack of screening strategies for these kinds of complex molecules, primarily because of their miniscule level of production by a few bacterial species. This roadblock of terpenoid discovery in bacteria could be overcome with the development of advanced genome mining platforms for terpene biosynthetic gene clusters (BGCs) like has already been achieved to quite an extent for polyketide synthases (PKSs)/ non-ribosomal peptide synthetases (NRPSs). Due to the co-linearity logistics followed by these modular megasynthases, the discovery of novel PKS/NRPS products was made possible that was aided by the predictive power of the bioinformatics computational tools (Singh et al. 2017). However, no such rules exist for the prediction of the cyclic hydrocarbon backbones produced by terpene cyclases (TCs) (Helfrich et al. 2019).

Historical overview

The knowledge of a small number of odoriferous terpene metabolites of bacterial origin spans various human civilisations. However, the very first report on terpene came in the year 1891 when Berthelot and Andre investigated the characteristic odour of the freshly ploughed soil and found that a vaporescent substance, evidently the cause of the typical musty scent of soil, could be extracted by steam distillation. Gerber and co-workers carried out the first modern attempt to study these volatile compounds from bacteria and presumed that the peculiar odour of the cultures of Actinomycetales, which extensively occur in soil, might be due to the volatile terpenes they produce (Gerber and Lechevalier 1965; Gerber 1967). Gerber also isolated and determined the structure of the Berthelot's volatile substance, geosmin and the methylated monoterpene, 2-methylisoborneol (2-MIB) as along with many different cyclic sesquiterpenes, produced by Streptomycetes (Gerber and Lechevalier 1965; Gerber 1968, 1971, 1973). In 1981, both geosmin and 2-MIB were reported to be isoprenoid derivatives (Bentley and Meganathan 1981). In the years that followed, a variety of terpene hydrocarbons and alcohols, having broad range of biological activities and pharmacological properties, have been isolated from various bacterial sources (Wilkins and Schöller 2009; Citron et al. 2012, 2015; Rabe et al. 2013; Yamada et al. 2015b; Dickschat 2016).

Over 55,000 terpenes have been identified, isolated and characterised; and with every passing decade, their number is consistently increasing (Wang et al. 2018). However, the targeted isolation of terpenes remained highly challenging until the discovery of bacterial pathways encoded by terpene BGCs, the identification of which has been facilitated by the bioinformatics computational tools and the ever-growing bacterial genome sequence database. The sequence information of terpene BGCs was the point of conception for pathway engineering and the thought of applying synthetic biology tools to expand the molecular diversity of these high value metabolites (Helfrich et al. 2019). We hereby review the bacterial terpene cyclases and the BGCs encoding them that have been identified mainly using bioinformatics approach. The biosynthetic insights thus gained have led to the cloning and heterologous production in well characterised hosts and has paved the way for more economical, sustainable and efficient production of high value terpenoids.

Terpene biosynthesis

TCs or terpene synthases (TPSs) are divalent metal-dependent enzymes which catalyse the conversion of linear, achiral isoprenoid diphosphates into diverse structural and stereochemically complex terpene hydrocarbons. Apart from geosmin, a sesquiterpenoid antibiotic, pentalenolactone has been isolated from more than 30 species of *Streptomyces* (Takeuchi et al. 1969; Martin et al. 1970). Its biosynthetic enzyme, pentalenene synthase (PS) was initially isolated from *Streptomyces exfoliatus* UC5319, subsequently cloned and heterologously expressed in *E. coli* (Cane et al. 1994). PS was the first bacterial TPS for which the full sequence was determined and further, comprehensive experiments with stereo-specifically tagged farnesyl diphosphate (FPP) have proven the detailed mechanism and stereochemistry of cyclisation of FPP to pentalenene (Cane et al. 1990). Since then, PS has been used as a reference to mine novel bacterial TPSs. Metabolically, the terpenoid biosynthesis can be divided into three steps (Sacchettini and Poulter 1997) (Fig. 1):

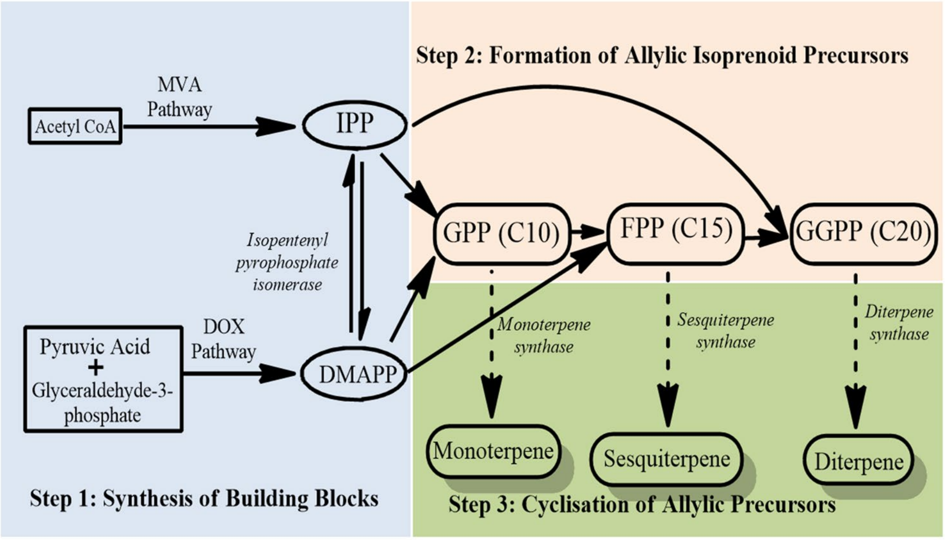


Fig. 1 Schematic representation of the steps of terpenoid biosynthesis. Step 1—Synthesis of 5C building blocks viz. isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) via two individual pathways namely mevalonate (MVA) pathway and non-mevalonate deoxyxylulose (DOX) pathway; Step 2—Successive condensation of IPP and DMAPP to form allylic isoprenoid precursors geranyl diphosphate (GPP), farnesyl diphosphate (FPP) and geranyl geranyl diphosphate (GGPP); Step 3—The enzyme catalysed cyclisation of allylic precursors to form different cyclic terpene metabolites

(i) formation of the five carbon (C_5) isoprene units, isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP), by two independent pathways viz Mevalonate (MVA) and Deoxyxylulose phosphate (DOXP) pathway
(ii) successive condensation of IPP and DMAPP and chain elongation resulting into linear isoprenyl diphosphate precursors of varying chain lengths such as geranyl diphosphate or GPP (C_{10}), FPP (C_{15}), geranyl geranyl diphosphate or GGPP (C_{20}), etc. and
(iii) enzyme catalysed cyclisation and modification of linear isoprenoid precursors

This cyclisation mechanism of the linear isoprenoid precursors to (poly)cyclic terpene hydrocarbons or alcohols belongs to the most interesting and complex enzymatic reactions occurring in nature and essentially add to the exquisite chemodiversity to this largest family of natural products.

Mining of novel TPSs

"Genome-mining" has proven to be a good paradigm for the discovery and characterisation of the natural product biosynthetic genes (Corre and Challis 2009; Ikeda 2017). However, the actual biochemical role of a prospect TPS can be attributed only by incubating it with the allylic isoprene precursors and the subsequent sophisticated analysis of the product formed. Most of the bacterial TPSs reported are sesquiterpene cyclases (SCs) (Nakano et al. 2011a), which show no significant amino acid sequence identity with plant and fungal type SCs. However, two characteristic metal ion binding motifs are universally conserved among SCs: the aspartate-rich motif (**DD**XX**D**/E) usually found ~ 80–120 amino acids downstream of the N-terminus of microbial synthases and NSE triad (**N**DXX**S**XX(K/R)**E**), modified in plants to DTE triad (**D**DXX**T**XX(K/R)**E**) (Starks et al. 1997), located 140 ± 5 amino acids downstream of the aspartate-rich motif and these motifs are responsible for the binding of tri-nuclear magnesium cluster (Cane and Ikeda 2012).

On the basis of these motifs, local sequence alignment searches with BLAST algorithm can identify many probable TPSs. Besides this, the profile hidden Markov models (HMM) provide the alternative strategy to identify putative TPSs, which operate by means of codifying the fundamental pattern of functional motifs and search for the proteins with similar function, irrespective of the presence of significant levels of primary sequence similarity (Eddy 1998). The leading profile HMM can discern members of the applicable protein functional families from non-members, with high precision (Finn et al. 2009; Punta et al. 2012). Similar strategy has now been in practice to harvest the probable TPS sequences from the bacterial genome sequence databases. Mining of the increasing number of bacterial genome sequences has already unveiled a treasure trove of novel TPSs but SCs were only reported in *Streptomyces* and *Nostoc* till date. However, recent genome sequencing projects have revealed a large number of potential SC genes and more than 80 uncharacterised single-domain SC homologs were found in a variety of bacterial genomes using a BLASTp search with the PS from *Streptomyces exfoliatus* UC5319 as a query (Nakano et al. 2011b). Many of these TPSs and their homologs show ~ 30% amino acid sequence identity to the known bacterial SCs, indicating that the homologs may catalyse different reactions to generate a variety of terpenoids. The TPSs utilise two main chemical strategies for initial carbocation formation (Lesburg et al. 1997):

i. Ionisation-dependent, or
ii. Protonation-dependent.

Ionisation-dependent TPSs (Class I cyclases)—monoterpene, sesquiterpene and majority of diterpene cyclases—contain an aspartate rich sequence, DDXXD/E and a second sequence i.e., N/DDXXS/TXX(K/R)E. These motifs together bind a trinuclear magnesium cluster with the diphosphate moiety of the substrate (e.g., pentalenene synthase; Fig. 4 Scheme I) that further actuates the cleavage of the substrate pyrophosphate (leaving) group to generate highly reactive allylic carbocations. In contrast, protonation-dependent TPSs (Class II cyclases)—diterpene, triterpene and squalene cyclases—contain a single signature motif DXDD, in which the central aspartate residue is implicated as the proton donor (e.g., terpentetrienedienyl diphosphate synthase; Fig. 5, Scheme I) that triggers initial carbocation formation (Christianson 2006). As class II cyclase leave the diphosphate group retained with their substrate, their products can serve as substrates for further cyclisation by class I cyclase. In both cases, the resulting carbocation undergoes a cascade of cyclisation influenced by a combination of steric and electrostatic forces within the active site pocket to generate the ultimate end-product. The abstraction of a proton or electrophilic attack by water ultimately terminates the cyclisation cascade (Christianson 2006). In fact, the electrostatic environment of the TPS active site can contribute preferably to the formation of stable carbocations; expanding the potential product diversity for the cyclisation of an isoprenoid diphosphate. Terpenoids, therefore, seem to be significantly more widely distributed in nature than has been appreciated previously and hence there is a need to characterise several previously undescribed bacterial TPS homologs.

Expression of bacterial TPS encoding genes

The most probable allylic diphosphate substrate can be identified for the catalytic function of the putative TPS by phylogenetic analysis of the sequences with the characterised TPSs, having known substrates. Also, the actual biochemical role of these putative TPSs is always required to be experimentally established. For this purpose, two types of engineered hosts such as *Escherichia coli* and *Saccharomyces cerevisiae* have been successfully used for heterologous expression and their potential has been demonstrated with success of metabolic pathway engineering, genome editing and developed microbial association for improvement of terpenoid production from bacterial sources (Wang et al. 2018). The in vitro incubation of the recombinant protein with the identified diphosphate substrate has frequently been practised for the evaluation of the putative TPSs (Cane et al. 1994). However, such recombinant proteins expressed heterologously in *E. coli* are frequently obtained as insoluble inclusion bodies (Yamada et al. 2015b). Nonetheless, a few have been successfully refolded to obtain soluble proteins with the required catalytic activity (Cane and Watt 2003; Cane et al. 2006; Takamatsu et al. 2011; Yamada et al. 2015b). For the establishment of the catalytic function of the concerned synthase gene, specially engineered *E. coli* host strains have been developed for the efficient production of terpenoid metabolites (Harada and Misawa 2009), as the normal *E. coli* strains are devoid of appropriate amount of isoprenoid precursors. Thus, the upstream biosynthetic mevalonate pathway gene cluster of *Streptomyces* was engineered in *E. coli* which in turn increased the intracellular concentration of precursors to produce the efficient amount of desired isoprenoids. Moreover, since almost all actinobacteria produce geosmin and/or 2-MIB as terpenoid metabolites, thus they seem to have effective metabolic pools of terpene biosynthesis precursors. Using them, an efficient and versatile alternative heterologous expression system was developed for the synthesis of a broad range of secondary metabolites, including terpenoids.

On the basis of the available genome data of *Streptomyces avermitilis*, the producer of the broadly used anthelmintic agent avermectin, engineered strains of *S. avermitilis* have been developed; which lacks 1.5 Mb of DNA from the left sub-telomeric region harbouring the majority of natural product biosynthetic genes. These large-deletion mutants, referred as "SUKA", like the parent *S. avermitilis* strain, are efficient and robust; and retain the ability of producing biological building blocks and export metabolic end-products (Chou et al. 2010; Komatsu et al. 2010). These are 'a high-quality' hosts for the introduction of individual biosynthetic genes or complete gene clusters from *Streptomyces* and eukaryotic sources, thereby producing otherwise unobtainable, composite biosynthetic intermediates and determining their catalytic roles. SUKA17, one of the SUKA mutants which lack TPS genes, has been used for the heterologous expression of genes encoding the 2-MIB synthase, and the plant codon-optimised amorpha-4,11-diene synthase of *Artemisia annua* and the yield was enhanced by co-transcribing a copy of endogenous *S. avermitilis ptlB* gene encoding FPP synthase, which provides the efficient source of FPP for the synthesis of amorpha-4,11-diene (Komatsu et al. 2010). For the effective expression, the TPS genes along with their native or a heterologous ribosome- binding site, can be incorporated downstream of a suitable substrate GPP (*gps*), FPP (*ptlB*) or GGPP (*crtE*) synthase gene controlled by a constitutively expressed promoter (*rpsJp*) (Komatsu et al. 2010). The combination of using a versatile host SUKA17 and co-expression of the putative TPS gene with the polyprenoid substrate synthase is efficient for the discovery and characterisation of new TPSs.

Terpene synthases characterised so far

To identify TPS genes in bacteria, by the efficient HMM and Pfam strategy, the excavation of huge repertoire of available bacterial genome data has revealed that the majority of them are SCs. However, the identity of some of the putative bacterial TPSs and their biochemical functions were further determined by overexpressing the synthases which consume the allylic precursors to form variety of numerous terpene metabolites as evident from the referred studies which are summarised in Table S1. Based on the phylogenetic analysis of the discovered bacterial TPSs, they are classified into three main categories: monoterpene, sesquiterpene and diterpene synthases (Fig. 1).

Monoterpene synthases

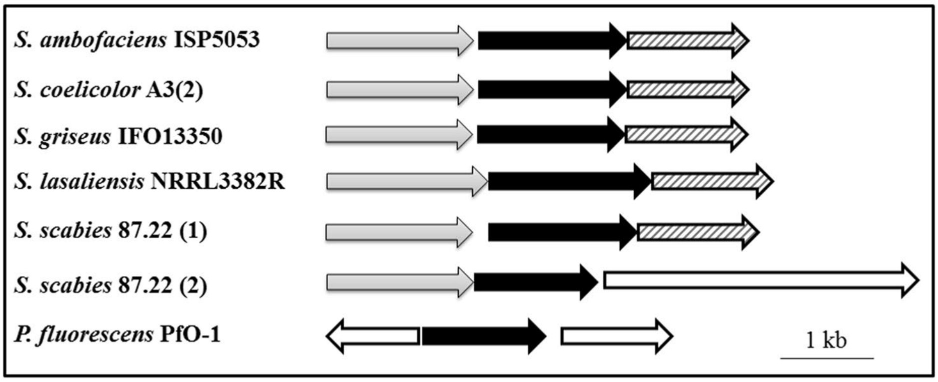
Monoterpenes (C10) are distinctive components of volatile plant oils and are widely used in fragrance and flavour industry and as cooling substances (Sharkey et al. 2008). In wood plants such as conifers, pinene (the most abundant bicyclic monoterpene) is the main constituent of the wood resins, whereas limonene (the most abundant monocyclic monoterpene) is the main component of essential oils from citrus plants, e.g. lemon and orange (Bicas et al. 2009). Monoterpenes, consisting of two linked isoprene (C_5H_8) units, are formed by the enzyme monoterpene synthase catalysed conversion of GPP. The discovery of monoterpene synthases in plants and fungi is more than five decades old story. However, in bacteria, it became possible to discover novel and previously known plant monoterpenes with the advent of genome mining techniques.

Methylisoborneol and methylbornane biosynthesis

2-Methylisoborneol (2-MIB) is considered as a representative bacterial monoterpenoid and has a musty-earthy note of Brie and Camembert cheese flavors (Karahadian et al. 1985). The previously undiscovered synthase for the production of 2-MIB was identified by HMM based genome mining in seven representative microorganisms (*Streptomyces ambofaciens* ISP5053, *Streptomyces coelicolor* A3(2), *Streptomyces griseus* IFO13350, *Streptomyces lasaliensis* NRRL3382R, *Streptomyces scabies* 87.22, *Saccharopolyspora erythraea* NRRL2338, and *Micromonospora olivasterospora* KY11048) (Komatsu et al. 2008). Further testing of heterologously expressed 2-MIB synthases from these microorganisms for the metabolite production revealed that except *M. olivasterospora* which produced 2-methylenebornane (2-MB), all other produced 2-MIB. The *tpc* gene of *S. lasaliensis*, *saml0357* of *S. ambofaciens*, *sgr1269* of *S. griseus* and *sco7700* of *S. coelicolor* which encode 481-aa, 440-aa, 437-aa and 440-aa proteins, respectively, show a very low amino acid sequence similarity (less than 20% identity to PS of *S. exfoliatus*). However, each of these four synthases contains the same variant of the two-highly conserved characteristic motifs of the bacterial TPS (Table S1). The genomic analysis of the tested strains revealed that each of them contains *S*-adenosyl-L-methionine (SAM)-dependent C-methyltransferase gene along with cyclic nucleotide-binding protein encoding gene, located upstream of the 2-MIB synthase gene, in the form of a three-component biosynthetic operon (Fig. 2) (Komatsu et al. 2008). The in vitro formation of 2-MIB (Fig. 3A), on incubating GPP and SAM with the corresponding recombinant C-methyltransferase were found to catalyse SAM-dependent methylation of GPP at C-2, to form methylated acyclic substrate 2-methylgeranyl diphosphate (2-MeGPP) which on further cyclisation generates 2-MIB, catalysed by 2-MIB synthase.

An orthologous gene product, Pfl01_1841 of *Pseudomonas fluorescens* Pf0-1, which has significant sequence similarity to actinomycete 2-MIB synthases, consists of a different variant of the two-characteristic divalent metal binding motifs (Table S1) from that of the four 2-MIB synthase sequences mentioned earlier. An in vitro incubation of recombinant Pfl01_1841 with 2-MeGPP catalysed the generation of monoterpene hydrocarbon 2-methylenebornane

Fig. 2 Gene organisation of the predicted monoterpene cyclase in a 3-component biosynthetic operon. The greyed, blackened and the oblique-lined arrows indicate cyclic nucleotide binding protein, monoterpene cyclase and methyltransferase genes, respectively (Adapted from Komatsu et al. 2008 and modified)



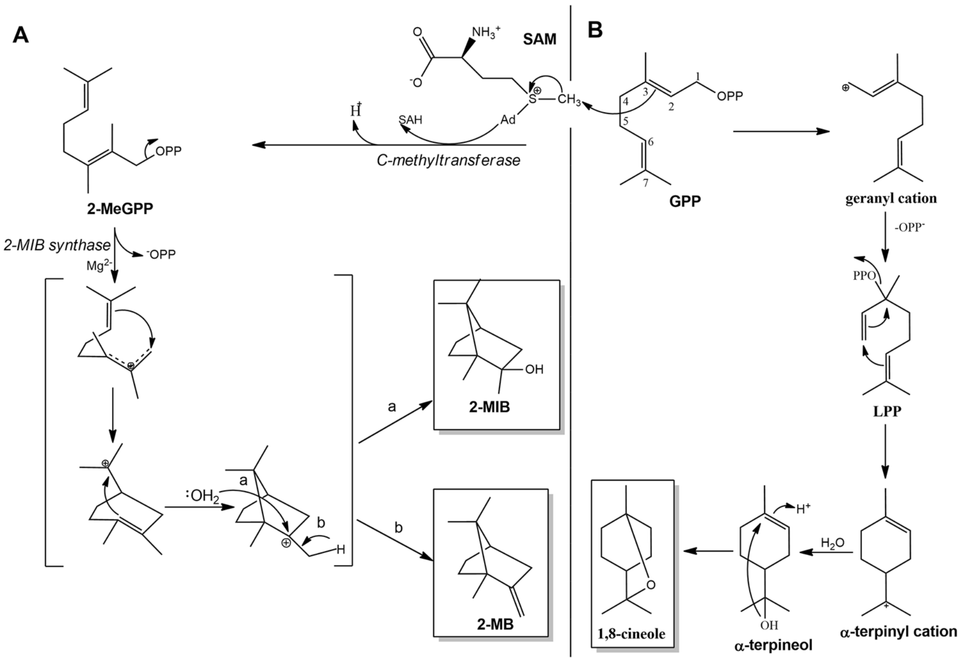


Fig. 3 Cyclisation of Geranyl diphosphates (GPP) by monoterpene synthases. **a** *S*-adenosyl-L-methionine (SAM)-dependent C-methyltransferase mediated synthesis of 2-methylisoborneol (2-MIB) and 2-methylisobornane (2-MB) and **b** Direct biosynthesis of monoterpene 1,8-cineole via α-terpineol intermediate. *LPP* linalyl diphosphate; *MeGPP* methyl geranyl diphosphate

(2-MB; Fig. 3a), due to a final deprotonation step instead of the attack of water in the terpene cyclisation, with no trace of detectable 2-MIB (Chou et al. 2011). The genome of *P. fluorescens* Pf0-1 was checked for the presence of C-methyltransferase gene in the proximity of the 2-MIB synthase gene in an operon, for the synthesis of 2-MIB and was found to be missing (Fig. 2); also the culture extracts were found to be devoid of 2-MB as well (Chou et al. 2011).

The above studies contributed to establish it as a fact that the enzymatic conversion of GPP to a monoterpene hydrocarbon requires an intermediate step of conversion of GPP to 2-MeGPP catalysed by SAM mediated C-methyltransferase. There were no reports of any bacterial monoterpene synthase which could catalyse the direct conversion of GPP to monoterpene without involving the methylation step, till Nakano et al. reported the synthesis of 1,8-cineole in *S. clavuligerus* ATCC27064 (Nakano et al. 2011b).

1,8-cineole, Linalool and Nerolidol biosynthesis

1,8-cineole, also known as eucalyptol, is a monoterpenic oxide and is present in many plant essential oils. It has been reported to possess various therapeutic applications as antimicrobial (Krist et al. 2008), antimalarial (Su et al. 2008), anthelmintic (Shah et al. 2011), cytotoxic, antitumor (Asanova et al. 2003), anti-inflammatory, analgesic (Santos and Rao 2000), hypotensive and smooth muscle relaxant (Lahlou et al. 2002). Seeing its vast therapeutic applications, it has been looked upon in bacterial genome database. The *cnsA* gene of *S. clavuligerus* was found to encode a 330-aa protein which contains the requisite conserved motifs (Table S1). The *cnsA* gene product (annotated as 1, 8-cineole synthase) cyclises GPP to achiral 1, 8-cineole directly (Fig. 3b), without the methylation of GPP, via the chiral intermediate α-terpineol of unknown absolute configuration. The proposed mechanism of direct cyclisation of GPP to a monoterpene requires the substrate ionisation followed by initial isomerisation step to linalyl diphosphate (LPP) by re-attacks of diphosphate to the geranyl cation at C3, which allows the subsequent cyclisation of LPP to the α-terpinyl cation. This cation is a common intermediate towards all cyclic monoterpene hydrocarbons and alcohols, while acyclic products may arise from geranyl cation or LPP. They characterised the second monoterpene synthase *SSCG_05291* gene product from *S. clavuligerus* and christened it *lnsA* (linalool/ nerolidol synthase) (Nakano et al. 2011a). LnsA, a 333-aa protein having 36% sequence identity with the PS of *Streptomyces* sp. UC5319, consists of the universally conserved metal binding motifs (Table S1). Interestingly, it is the only known bacterial TPS which generates acyclic products linalool and nerolidol. These compounds were identified by incubation of the purified enzyme with GPP and FPP, respectively followed by the GC/MS analysis with comparison to authentic samples. This enzyme converts FPP with higher efficiency than GPP. However, both the products were not detected in the liquid culture extracts of wild type *S. clavuligerus*, but were detected in the headspace extracts from *S. clavuligerus* agar plate cultures (Citron et al. 2012). The heterologous expression of the linalool/nerolidol synthase gene in *S. lividans* and *S. avermitilis*, however, resulted in the production of linalool only (Nakano et al. 2011a;

Yamada et al. 2015b), indicating that the enzyme may naturally function as monoterpene synthase.

Sesquiterpene synthases

Sesquiterpenes (C15), the volatile cyclic olefins, have remarkable medicinal properties and have been used as essential components of the pharmacopoeia since ancient times. A large array of structurally diverse variety of (poly) cyclic sesquiterpene skeletons are produced by the metal dependent conversion of FPP catalysed by SC (Cane 1985). Unlike, cyclic monoterpenes which are generated via the α-terpinyl cation, initial cyclisation reaction of FPP has six possible cations (Fig. S1) (Degenhardt et al. 2009), involving either the direct cyclisation to the central double bond of the initially formed farnesyl cation or proceeds via its isomer the nerolidyl cation. The geometric constraints of transoid farnesyl cation undergoes the direct 1,10- or 1,11- cyclisation to form a 10-membered (*E,E*)-germacrenyl cation or a 11-membered (*E,E*)-humulyl cation, respectively. Another type of cyclisation proceeds with the formation of tertiary cisoid nerolidyl cation after isomerisation of C2–C3 double bond by the migration and the release of diphosphate (Davis and Croteau 2000). This cation can undergo cyclisation (Fig. S1) to either the central (at C6) or distal (at C10) double bond forming 1,6- [bisabolyl cation], 1,7- [cycloheptanyl cation], 1,10- [(*E,Z*)-germacrenyl cation] and 1,11- [(*E,Z*)-humulyl cation].

Pentalenone and pentalenolactone biosynthesis

The study of a bacterial SC began with the extensive investigation of pentalenene. It is the hydrocarbon precursor of the pentalenolactone which inhibits glyceraldehyde-3-phosphate dehydrogenase in Gram-negative and Gram-positive bacteria and hence, is antibacterial (Hartmann et al. 1978; Cane and Sohng 1989). PS is the first characterised enzyme of this class, which was partially purified from *S. exfoliatus* UC5319 cell lysate (Cane and Pargellis 1987). Later on, when the complete 9.03 Mb genome of *S. avermitilis* was sequenced, it was found to consist 30 presumptive secondary metabolite BGCs (Omura et al. 2001). It includes a 13.4 kb BGC containing 13 unidirectionally transcribed ORFs for the biosynthesis of pentalenolactone. One of them is *sav2998*, encoding a 336-aa protein (PtlA) that shows 76% sequence identity to the previously known PS of *S. exfoliatus*. This cluster has been characterised, both genetically and biochemically, as an orthologous PS (Tetzlaff et al. 2006). The purified recombinant PtlA was shown to catalyse the Mg^{2+}-dependent cyclisation of FPP in vitro which in general initiates with the formation of *(E,E)*-humulyl cation (Fig. S1). The *(E,E)*-humulyl cation undergoes 1,2-hydride shift (Fig. 4, Scheme I), either directly or by a series of protonation and deprotonation to form protoilludyl cation. The cyclisation proceeds with the dyotropic rearrangements, a process in which two σ-bonds simultaneously migrate intramolecularly, to finally yield pentalenene.

Geosmin biosynthesis

The biosynthesis of geosmin has also been investigated in great detail. Two groups independently characterised the gene (sco6073) encoding geosmin synthase (GS) in *S. coelicolor* A3(2) (Cane and Watt 2003; Gust et al. 2003). The 2181-bp *sco*6073 gene encodes a 726-aa protein in which the N-terminal (319 aa) and C-terminal (339 aa) halves show about 30% identity to PS from *S. exfoliatus*. Similarly, the 2178-bp *sav*2163 gene of *S. avermitilis* encodes a 725-aa protein in which the N-terminal (363 aa) and C-terminal (329 aa) halves each exhibit 27% identity (Cane et al. 2006). Also, the protein showed 77% identity to the previously known *S. coelicolor* A3(2) GS (SCO6073). The N-terminal domain of both the synthases contain the characteristic conserved Mg^{2+}-binding motifs, the acidic amino acid-rich motif -**DD**HFLE- and the NSE triad -**ND**LF**S**YQR**E**-, whereas the C-terminal contains an unusual acidic amino acid-rich motif, -**DD**YYP- and a canonical -**ND**VF**S**YQK**E**- sequence in both synthases (Table S1). The N-terminal domain has been shown to catalyse cyclisation of FPP to produce germacradienol and germacrene D, whereas the C-terminal domain has been demonstrated to catalyse 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 undergoes a proton-initiated hydration and hydride-shift to form geosmin (Jiang and Cane 2008). As a matter of fact, the sequence of geosmin synthase is highly conserved all over a wide variety of bacterial genera, with more than 75 presumed proteins from actinobacteria, myxobacteria, and cyanobacteria exhibiting 45–99% sequence identity to the *S. coelicolor* A3 (2) SCO6073 (Cane and Ikeda 2012).

For mining and characterising the novel secondary metabolites, the correctness of the genomic sequence in the public database is very essential as was critically analysed by Giglio et al. in their studies involving the model cyanobacterium *Nostoc punctiforme* PCC73102 (Giglio et al. 2008). They discovered a GS homolog which encodes a hypothetical protein (ZP_00109187), having 55% sequence similar to the N-terminal of SCO6073 and contains the universally preserved motifs too (Table S1). However, the encoded protein has only 630-aa, which is approximately 100-aa shorter than SCO6073. They claimed that the hypothetical protein ZP_00109187 encoded by *npun*02003620 to be a truncated one. Although, the C-terminal half harbours a typical –DDYFP- motif, but the essential NSE triad was found missing due to which, during catalytic

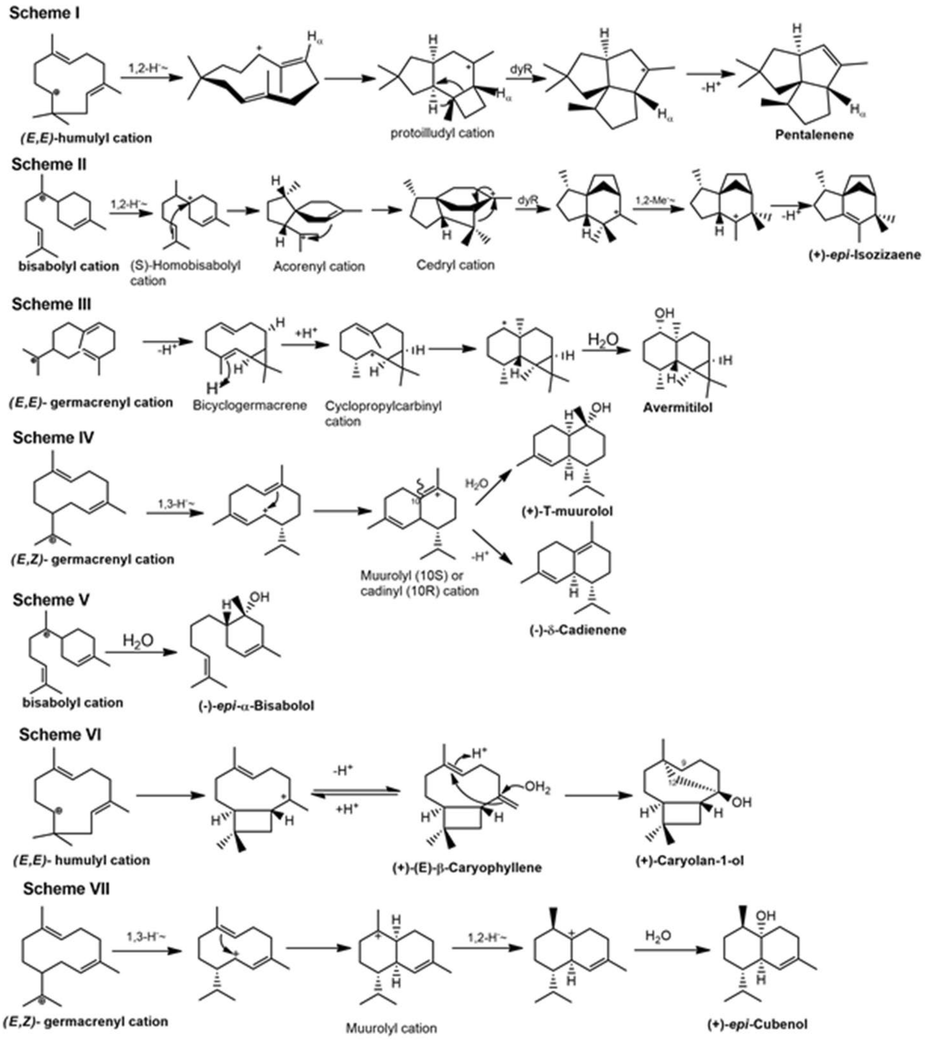


Fig. 4 Biosynthesis of sesquiterpene metabolites from a common substrate farnesyl pyrophosphate (FPP) via the intermediate cations. Scheme I: Biosynthetic mechanism of conversion of FPP proceeds via *(E,E)*-humulyl cation followed by 1,2-hydride shift and a series of protonation, deprotonation and a dyotropic rearrangement (dyR) to yield pentalenene in *S. avermitilis*; Scheme II: Production of *epi*-isozizaene in *S. avermitilis* from FPP via (R)-bisabolyl cation proceed through 1,2-hydride migration, subsequent cyclisation and dyR and final deprotonation; Scheme III: Avermitilol synthesis in *S. avermitilis* proceeds via a series of protonation and deprotonation of (R)-*(E,E)*-germacrenyl cation and finally the capture of water yields avermitilol; Scheme IV: Biosynthesis of (−)-δ-cadinene and (+)-T-muurolol via (S)-*(E,Z)*-germacrenyl cation which proceeds through 1,3-hydride migration to form an intermediate cation which cyclizes to form muurolyl (10S) cation which forms (+)-T-muurolol by the attack of water and (+)-T-muurolyl (10S) or cadinyl (10R) cation forms (−)-δ-cadinene by deprotonation, in *S. clavuligerus*; Scheme V: (S)-bisabolyl cation captures water to form (−)-*epi*-α-bisabolol, in *S. citrocolor*; Scheme VI: *(E,E)*-humulyl cation undergoes a series of protonation-deprotonation to form (+)-(E)-β-caryophyllene which forms (+)-caryolan-1-ol, in *S. griseus,* by the attack of water; Scheme VII: (S)-*(E,Z)*-germacrenyl cation undergoes 1,3-hydride migration and subsequent cyclisation to muurolyl cation, which on 1,2-hydride shift and attack of water forms (+)-*epi*-cubenol in *S. griseus*

cyclisation of FPP to germacradienol, it does not form geosmin. The ostensive truncation was demonstrated to be the consequence of a single, but imperative, sequencing mistake in the published DNA (Giglio et al. 2008). So, a correct complete sequence was published for a 753-aa protein (WP_012409287), which contained the C-terminal NSE/DTE triad –**ND**IF**S**YQK**E**– that was missing in the previous published sequence. Further the correct protein was cloned, expressed and incubated with FPP in vitro which successfully resulted in the production of geosmin (Giglio et al. 2008).

epi-Isozizaene and albaflavenone biosynthesis

Among all the genome sequenced bacteria so far, *epi*-isozizaene synthase (EIZS) is the most commonly found SC. The biochemical function of EIZS was first studied in *S. coelicolor* A3(2) and it was found to catalyse the committed step for the biosynthesis of an α, β- unsaturated tricyclic sesquiterpene ketone, albaflavenone. After fourteen years of its discovery in *S. albidoflavus* (Gurtler et al. 1994), the enzyme responsible for its biosynthesis from *epi*-isozizaene could be established only due to advancements in genome mining techniques. As a part of two-gene operon with *sco5222*, the *S. coelicolor* genome contains a EIZS gene *sco5222* which encodes a 361-aa protein (SCO5222) having 24% identity to PS of *S. exfoliatus* with two common conserved Mg^{2+}-binding motifs (Table S1) (Lin et al. 2006). The recombinant EIZS was shown to catalyse the cyclisation of FPP which proceeds via (R)-nerolidyl cation and the (R)-bisabolyl cation (Fig. S1) that forms the (S)-homobisabolyl cation in a 1,2-hydride shift (Fig. 4, Scheme II). A subsequent cyclisation yields the acorenyl cation and cedryl cation which undergoes dyotropic rearrangement, 1,2-methyl migration and final deprotonation to generate (+)-*epi*-isozizaene. The genomic analysis of *sco5222* gene revealed that it shares four nucleotides ATGA transcriptional overlap at its 3′ end with *sco5223*, a cytochrome P450 CYP-170A1 gene. Also GC/MS analysis of extracts from *S. coelicolor* culture constituted the time-dependent production of *epi*-isozizaene, and its oxidation products (4R)- and (4S)- albaflavenols and antibiotic albaflavenone, catalysed by the purified recombinant SCO5223. However, the disruption of *sco5223* gene resulted in the non-formation of all the three oxidation products (Zhao et al. 2008), indicating that *sco5223* gene product cytochrome P450 CYP170A1 (also known as albaflavenone monooxygenase) is responsible for cytochrome P450 mediated oxidation of *epi*-isozizaene to finally generate the antibiotic albaflavenone.

The orthologous protein SAV_3032 of bacterium *S. avermitilis* was also reported to catalyse the cyclisation of FPP to *epi*-isozizaene (Takamatsu et al. 2011). The *sav3032* gene also shares the ATGA transcriptional overlap at its 3′ end with the gene *sav3031*, which encodes for the cytochrome P450 CYP 170A2 protein, analogous to *S. coelicolor* SCO5223 in a two-gene operon. *sav3032* encodes 363-amino acid protein (SAV_3032) which exhibit 82% sequence identity with SCO5223 and 25% with that to PS of *S. exfoliatus*. It harbours the same variant of universally conserved metal binding motifs as SCO5222 (Table S1). Purified recombinant SAV_3032 was shown to catalyse the Mg^{2+}-dependent cyclisation of FPP to *epi*-isozizaene, although *epi*-isozizaene or any of its oxidation products could not be detected in the wild type culture of *S. avermitilis*. However, when *sav3032* was introduced in mutant SUKA16 under the control of rpsJ promoter, the transformant thus formed was able to produce *epi*-isozizaene. Co-expression of *sav3032* gene with *sav3031* resulted in the production of (4S)- and (4R)- albaflavenols, albaflavenone and a previously unknown epoxy-alcohol (Takamatsu et al. 2011). The biosynthetic two-gene operon for the production of albaflavenone has been found to be extremely conserved and is extensively distributed as orthologous pairs of translationally coupled SC gene to the downstream cytochrome P450 gene, which shows 56–100% identity apparent in 10 species of streptomycetes; also the production of *epi*-isozizaene and/or albaflavenone has been successfully achieved in most of them.

Avermitilol bioysnthesis

S. avermitilis has been known for the production of commonly used anthelminthic macrolide avermectin (Omura et al. 2001). In addition to the genes encoding for PS (*sav2998*) (Tetzlaff et al. 2006), GS (*sav*2163) (Cane et al. 2006) and EIZS (*sav*3032) (Takamatsu et al. 2011), it also contains avermitilol synthase (*sav76*) gene which encodes 335-aa protein (SAV76) having 35% identity with PS of *S. exfoliatus* harbouring conserved characteristic metal binding motifs (Table S1). Although the extracts of wild type of *S. avermitilis* was devoid of avermitilol, but the purified recombinant protein SAV76 expressed in *E. coli,* on incubation with FPP resulted in the major enzymatic product as tricyclic sesquiterpene alcohol, avermitilol (85%) and its isomer viridiflorol (3%) as well as germacrenes A (1%) and B (5%) (Chou et al. 2010). To effect the production of avermitilol, they introduced *sav76* gene in mutant SUKA17 under control of rpsJ (*sav4295*) promoter but surprisingly, the production of avermitilol decreased to 15% alongwith the generation of 67% avermitilone (a ketone derivative of avermitilol), 10% germacrene A, 5% germacrene B and 2% viridiflorol. The proposed mechanism for the biosynthesis of avermitilol proceeds from FPP via (R)-*(E,E)*-germacrenyl cation (Fig. S1) which upon deprotonation (Fig. 4, Scheme III) forms the neutral intermediate bicyclogermacrene. Bicyclogermacrene on reprotonation yields the cyclopropylcarbinyl cation which cyclises into an intermediate secondary cation. Upon final attack of water, this secondary cation forms the avermitilol.

T-Muurolol and δ-cadinene biosynthesis

S. clavuligerus ATCC 27,064 genome comprises a remarkably large number of about 20 bioinformatically identified TPS genes, out of which (−)-δ-cadinene synthase (*sscg_02150*) and (+)-T-muurolol synthase (*sscg_03688*) are the only two characterised SC genes from this strain (Hu et al. 2011). The gene *sscg_02150*, which encoded a predicted protein of 327-aa, was lined downstream by a putative cytochrome P450 and upstream by a predicted methyl transferase being unidirectionally transcribed and an oppositely transcribed oxidoreductase in the operon. Similarly, the *sscg_03688* gene encoded a predicted 418-aa protein flanked immediately upstream and downstream by two cytochrome P450s being unidirectionally transcribed and a downstream methyl transferase. It was speculated that these putative cytochrome P450s and methyl transferases encoded within the two operons probably catalyse an undeciphered sequence of oxidations and methylations to form the sesquiterpene metabolites (T-muurolol and δ-cadinene). Both TPSs display the two highly conserved metal binding motifs (Table S1) and were found to catalyse the cyclisation of FPP to the widely occurring plant sesquiterpenes (−)-δ-cadinene and (+)-T-muurulol, respectively. These SCs from *S. clavuligerus* show only 32% sequence homology and are also distant apart in the phylogenetic tree of bacterial terpene cyclases (Rabe et al. 2016), but they have quite similar biosynthesis mechanism (Hu et al. 2011). Both biosynthesis reactions initiate via (S)-*(E,Z)*-germacrenyl cation (Fig. S1), which upon 1,3-hydride shift form an intermediate cation (Fig. 4, Scheme IV). The final attack of water on muurolyl cation, formed due to the cyclisation of the intermediate cation, lead to the production of (+)-T-muurolol. The final deprotonation of either muurolyl or cadinyl cation produces (−)-δ-cadinene. Also, the -N**D**LM**T**VDK**E**- motif was found to contain DTE, similar to that of plant SCs instead of NSE residues commonly found in bacterial (−)-δ-cadinene synthases (Hu et al. 2011).

β-Caryophyllene and caryolan-1-ol biosynthesis

Another plant sesquiterpene, β-caryophyllene was found to occur in the essential oils of clove, cannabis and rosemary and its oxide show anti-fungal activity against dermatophytes in the treatment of onychomycosis (Yang et al. 2000). Also, it was found to be released from the maize leaves and roots in response to damage by the attack of different herbivore larvae (Köllner et al. 2008), thereby, contributing to the plant's defense against herbivores by attracting entomopathogenic nematodes. Seeing its application in agriculture and pharmaceutical, its production was looked upon in bacterial genomes. A gene encoding (+)-caryolan-1-ol synthase (*sgr_2079* or *gcoA*) from *S griseus* NBRC13350 was identified by Nakano et al. (2011a, b, c, d). GcoA (335-aa) exhibits 29% identity with the PS of *S. exfoliates* and was found to contain both the highly conserved Mg^{2+}-binding motifs (Table S1). When the recombinant enzyme was incubated with FPP, it yielded a small quantity of β-caryophyllene with a major product (+)-caryolan-1-ol. Several β-caryophyllene synthases were identified from plants but have not been reported for the production of (+)-caryolan-1-ol. Thus, GcoA was the first characterised caryolan-1-ol synthase from bacteria, catalysing the first cyclisation of FPP to β-caryophyllene (Fig. 4, Scheme VI) followed by a second cyclisation to form (+)-caryolan-1-ol. (−)-caryolan-1-ol chemical synthesis method is reported (Barton et al. 1952) but the biochemical synthesis of (+)-caryolan-1-ol, that too in bacteria, is unprecedented. Interestingly, in most cases, (+)-caryolan-1-ol synthase occurs simultaneously with the (+)-*epi*-cubenol synthase (Dickschat 2016).

epi-Cubenol biosynthesis

epi-Cubenol is an inhibitor of DNA polymerase and the extracts from the essential oil of the waste wood chips of Japanese cedar, prepared by steam distillation, was found to suppress the growth of a fungi *Trichophyton rubrum*, which causes the tinea disease (Takao et al. 2012). In an another study, the cell free extract (CFE) and a purified (+)-*epi*-cubenol synthase from the liverwort *Heteroscyphus planus* was found to catalyse the conversion of FPP to yield (+)-*epi*-cubenol (Nabeta et al. 1997). Besides plants, it was also reported to be present in the crude extracts of certain *Streptomyces* sps. For instance, (+)-*epi*-cubenol synthase was originally isolated from *Streptomyces* sp. LL-B7 by biochemical assay-guided methods, that was known to produce cadin-4-en-1-ol, an enantiomer of (+)-*epi*-cubenol (Gerber 1971). Later, Cane et al. confirmed the presence of *epi*-cubenol synthase in the CFE obtained from *Streptomyces* sp. LL-B7 by incubating CFE with tritiated FPP to give tritiated (+)-*epi*-cubenol (Cane et al. 1993). Considering its antifungal application and based on previous reports of *Streptomyces* sps. containing *epi*-cubenol synthases, bacterial genomes were screened out for its homologs. Soon after identifying the (+)-caryolan-1-ol synthase gene from *S. griseus* NBRC13350, a second gene (*sgr_6065* or *gecA*) from this organism was heterologously expressed in *S. lividans* which was identified as (+)-*epi*-cubenol synthase by incubation of the recombinant enzyme with FPP (Nakano et al. 2012). GecA is a 358-aa protein, having 28% sequence identity with the PS of *S. exfoliatus* and has both the conserved bacterial TPS motifs in the active site (Table S1). The optical rotatory power of the (+)-*epi*-cubenol obtained from *S. griseus* was also determined and it indicated to be the same enantiomer as in the liverwort *H. planus.* The biosynthesis of (+)-*epi*-cubenol proceeds via (S)-*(E,Z)*-germacrenyl cation (Fig. S1) which undergoes 1,3-hydride shift (Fig. 4, Scheme VII) to form an intermediate cation followed by subsequent cyclisation to yield muurolyl cation. The muurolyl cation then undergoes 1,2-hydride shift and a final attack of water to form (+)-*epi*-cubenol.

Germacradien-4-ol and *epi*-α-bisabolol biosynthesis

Two single domain SC homologs SC1 and SC2 were identified in *S. citricolor* NBRC13005 by genome mining approach, which were heterologously expressed in *S. lividans* (Nakano et al. 2011c). The culture extracts from the heterologous host was analysed by GC/MS which detected the presence of one sesquiterpene alcohol in each case. Both TPSs SC1 (316-aa) and SC2 (313-aa) display 32% and 29% identity with the PS of *S. exfoliatus*, respectively and contains the requisite metal binding motifs (Table S1). They were found to catalyse the cyclisation of FPP to the sesquiterpene alcohol metabolites namely (−)-germacradien-4-ol and (-)-*epi*-α-bisabolol, respectively. The mechanism of biosynthesis of (−)-germacradien-4-ol starts with the formation of (R)-*(E,E)*-germacrenyl cation (Fig. S1) and proceeds by the 1,3 hydride shift to form an intermediate cation which captures the water to form (−)-germacradien-4-ol, whereas the biosynthesis of (−)-*epi*-α-bisabolol (Fig. 4, Scheme V) occurs by the attack of water to the (S)-bisabolyl cation. α-Bisabolol, also known as levomenol, is a monocyclic sesquiterpene alcohol and is primarily found as the constituent of the essential oil from German chamomile, *Matricaria recutita* and *Myoporum crassifolium* plants (Kamatou and Viljoen 2010). (−)-*epi*-α-Bisabolol is known to have anti-inflammatory and anti-microbial properties and has been used for hundreds of years in cosmetics because of its perceived skin healing properties.

Sodorifen biosynthesis

Recently, genome mining of the newly sequenced rhizobacterium *Serratia plymuthica* WS3236 led to the identification of a 4.6 kb putative sodorifen BGC harbouring four individual genes (*sodA*, *sodB*, *sodC* and *sodD*), which are responsible for the production of sodorifen in *S. plymuthica* (Duell et al. 2019). *S. plymuthica* is known to emit several VOCs having sodorifen as one of the major metabolites. It is a polymethylated hydrocarbon and has been shown to prevent the fungal and plant growth (Kai et al. 2010; Von Reuß et al. 2010). The first two genes *sod*A and *sod*B are homologs of MEP pathway genes, IPP isomerase and 1-deoxy-D-xylulose-5-phosphate synthase (DXP synthase), respectively and are responsible for the supply of additional FPP. The other two genes *sod*C (SAM-dependent C-methyltransferase) and *sod*D (Sodorifen cyclase) are the necessary genes and their gene products are required to catalyse the production of sodorifen from its precursor FPP. By using direct pathway cloning, the sodorifen BGC from *S. plymuthica* was integrated into a vector under the control of one single promoter, upstream of *sod*A and heterologously expressed into *E. coli* with low VOC background. Heterologous expression under the control of tetracycline inducible Ptet_o promoter led to the synthesis of sodorifen which on comparison with other cultures have been found to be 26-fold high.

Also, Rabe et al. (2016) screened out and characterised 6 putative analogous TPSs (-)-α-amorphene synthase and (−)-7-*epi*-α-eudesmol synthase from *S. viridochromogenes*, (+)-T-muurolol synthase form *Roseiflexus castenholzii*, (+)-4-*epi*-cubebol synthase from *Streptosporangium roseum,* (−)-γ-cadinene synthase from *Chitinophaga pinensis* and (−)-(E)-β-caryophyllene synthase from *Saccharothrix expanaensis* by performing phylogenetic analysis, the gene cloning by homologous recombination in yeast, followed by heterologous expression in *E. coli*. The incubation experiments of purified recombinant enzymes with ^{13}C-labelled topoisomers of FPP enabled the thorough understanding of the cyclisation mechanisms of these six bacterial tepene cyclases by one and two dimensional NMR spectroscopic analysis.

Diterpene synthases (DTSs)

Around ~ 12,000 diterpenoids are known and most of them are produced by plants and fungi while diterpenoids of bacterial origin are rarely known (Smanski et al. 2012). Diterpenes (C20) are biosynthesised from the linear isoprene precursor GGPP and are the precursors for numerous biologically important compounds such as taxol, terpentecin, platensimycin, etc., conferring anti-microbial and anti-tumor action. Their unique structures and diverse bioactivities have attracted much attention from chemists and biologists in recent past, but cloning, characterising and engineering the respective genes of diterpenoid pathway in higher organisms remain challenging. Access to the bacterial genome sequences and recent bioinformatics advancements have revealed the prolific biosynthetic potential for diterpenoids in bacteria, especially actinobacteria (Daum et al. 2009). The discovery of gibberellins from the bacteria *Rhizobium phaseoli* (Atzorn et al. 1988) laid down the foundation stone for the discovery of bacterial diterpenoids.

Terpentecin biosynthesis

Terpentetrienedienyl diphosphate (TDP) synthase (Cyc1), a class II DTS and terpentetriene (TTE) synthase (Cyc2), a class I DTS were the first two characterised bacterial DTSs reported for the biosynthesis of diterpenoid antibiotic terpentecine, from *Streptomyces griseolosporeus* MF730-N6 (currently referred as *Kitasatospora griseola*) (Hamano et al. 2002). Although Cyc1 shows less than 30% sequence similarity to the N-terminal halves of characterised eukaryotic DTSs, the availability of the signature DXDD motif confirms its bioinformatics-based functional attribution, which was subsequently validated experimentally in vitro. Cyc2 was promptly identified on the basis of its sequence homology to previously known bacterial class I DTSs, and presence of the characteristic DDXXD and NSE/DTE motifs (Hamano et al. 2002). On analysing the genome sequence of *S. griseolosporeus* MF730-N6, it was observed that mevalonate pathway gene cluster is present downstream to a set of five genes which is vitalfor the biosynthesis of terpentecine. Apart from these five genes, the cluster possessed 9 more genes, out of which two genes *orf11* and *orf12* encoded a 499-aa Class II synthase (having 29% sequence identity to *ent*-kaurene synthase of fungus *Phaeosphaeria* sp.) and a 311-aa long Class I synthase (with 25% identity to *S. exfoliatus* pentalenene synthase). Multiple sequence alignment of ORF11 with the eukaryotic DTSs revealed that the motif QXXDGSW, which was supposed to stabilise an intermediate cation during the process of cyclisation and the motif DXDDTA, which chelates the divalent metal ion in order to mediate substrate binding, were found to be conserved. The inferred class I ORF12 protein consisted of the universally conserved pair of divalent metal-binding motifs, the aspartate rich motif –**DDRWD-** and the triad –**NDYYSWGR E**-. On functional analysis of the two purified recombinant enzymes, it was found that their products, Cyc1 and Cyc2 are essential for the conversion of GGPP into TDP by Cyc1 followed by the formation of TTE (Fig. 5, Scheme I), the precursor of terpentecin.

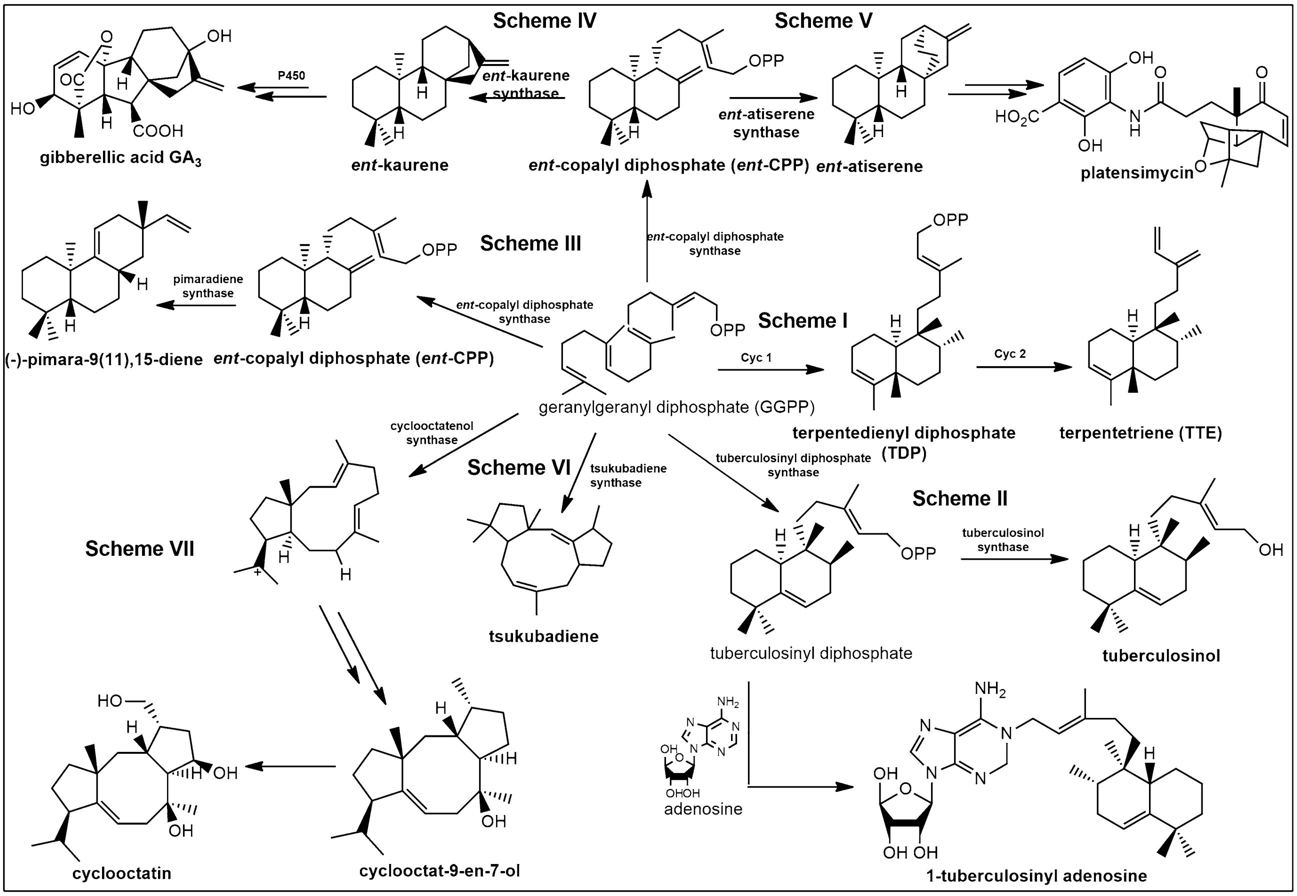


Fig. 5 Biosynthesis of diterpene metabolites from a common substrate geranylgeranyl diphosphate (GGPP) by class I and class II DTS enzymes. Scheme I: Cyclisation of GGPP to form terpentedienyl diphosphate (TDP) and further to terpentetriene (TTE) in *Kitasatospora griseola* catalysed by cyclase 1 (class II) and cyclase 2 (class I) respectively; Scheme II: In *M. tuberculosis* H37, the cyclisation of GGPP to form tuberculosinyl diphosphate and a second cyclisation of tuberculosinyl diphosphate to tuberculosinol catalysed by tuberculosinyl diphosphate synthase (class II) and tuberculosinol synthase (class I) respectively; Scheme III–V: Cyclisation of GGPP to form *ent*-copalyl diphosphate is the first committed step in the biosynthesis of pimaradiene in *Streptomyces* sp. KO-3988 catalysed by *ent*-copalyl diphosphate synthase; in the formation of gibberellic acid GA_3 via *ent*-kaurene, in *B. japonicum* USDA110, and in the synthesis of platensimycin via *ent*-atiserene in *S. platensis* (MA7327 and MA7339); Scheme VI: Tsukubadiene synthase (class I) enzyme catalyses the cyclisation of GGPP to tsukubadiene in *S. tsukubaensis* NRRL 18,488; Scheme VII: Generation of cyclooctat-9-en-7-ol by the cyclisation of GGPP to form an initial carbocation catalysed by class I DTS enzyme cyclooctat-9-en-7-ol synthase (CotB2) from *S. melanosporofaciens* MI614-43F2

Tuberculosinol and isotuberculosinol biosynthesis

Genome sequence analysis of *Mycobacterium tuberculosis* H37Rv, revealed the presence of a set of two DTS genes, a bacterial class II DTS, identified to be halimadienyl (or tuberculosinyl) diphosphate synthase (Rv3377c) which catalyses the bicyclisation of GGPP to produce tuberculosinyl diphosphate (Fig. 5, Scheme II) and a class I tuberculosinol synthase (Rv3378c) that catalyses the subsequent ionization mediated rearrangement of tuberculosinyl diphosphate to form tuberculosinol and isotuberculosinol (Nakano and Hoshino 2009; Nakano et al. 2011d). Rv3377c and Rv3378c gene products harbour the essential metal binding aspartate rich motifs DIDD and DDLLD respectively, but Rv3378c lacks the conserved NSE/DTE motif. Naturally, class I enzyme produces a terpene nucleoside, 1-tuberculosinyladenosine, by the nucleophilic attack of adenosine on the tuberculosinyl carbocation formed during initial ionisation step. Interestingly, both the genes are only found in the virulent species of *M. tuberculosis*, the leading cause of TB disease and the tuberculosinyladenosine has been hypothesised to be providing the pathogenicity to the bacteria (Nakano et al. 2005).

ent-Copalyl biosynthesis

A class II DTS, *ent*-copalyl diphosphate (*ent*-CPP) synthase was also identified from *Streptomyces* sp. KO-3988 (ORF2) (Ikeda et al. 2007), *Bradyrhizobium japonicum* USDA110 (Morrone et al. 2009), and three strains of *S. platensis* (MA7327, MA7339 and CB00739) (Smanski et al. 2011; Rudolf et al. 2016) and all share the characteristic DXDD motif. Out of the five genes implicated in the biosynthesis of viguiepinol, *orf2* and *orf3*, lie upstream of the MVA pathway gene cluster in *Streptomyces* sp. KO-3988 (Ikeda et al. 2007). *orf2* encodes a 511-aa protein with 32% sequence identity with *ent*-CPP synthase of *Gibberella fujikuroi* and *orf3* encodes a 295-aa protein, both harbours an aspartate rich motif -**DD**HVE- and the triad-**ND**LAT**FERE**-, having no sequence homology to any known TPS. Recombinant ORF2 corresponds to a class II synthase, catalyses the Mg^{2+}-dependent cyclisation of GGPP to copalyl diphosphate while ORF3, which was deduced to be class I synthase, in turn converts *ent*-CPP to pimara-9(11),15-diene (Fig. 5, Scheme III). The *ent*-CPP in *B. japonicum* is converted to *ent*-kaurene, the precursor for the production of gibberellin (gibberellic acid GA_3) (Fig. 5, Scheme IV). In *S. platensis*, *ent*-CPP is either converted to *ent*-kaurene or to *ent*-atiserene, a precursor for the mammalian fatty acid synthase inhibitors, platensimycin (Fig. 5, Scheme V) and platensin. The crystal structure of *ent*-CPP synthase from *S. platensis* CB00739 was solved and the docking studies revealed the catalytically active conformation of GGPP (Rudolf et al. 2016).

Tsukubadiene and cyclooctatin biosynthesis

Class I DTS genes were also identified by genome mining approach in *S. tsukubaensis* NRRL 18,488 (Yamada et al. 2015b) and *S. melanosporofaciens* MI614-43F2 (Kim et al. 2009). Tsukubadiene synthase from *S. tsukubaensis* was reported to catalyse the cyclisation of GGPP to tsukubadiene (Fig. 5, Scheme VI). Both the genes show no significant sequence similarity (< 13%) with the known class I DTSs. Although, tsukubadiene synthase consists of both the universally conserved bivalent metal binding domains, but on the other hand, the class I DTS gene *cotB2* in *S. melanosporofaciens,* encodes for a 307-aa protein having only the NSE/DTE triad -**N**DFY**S**YDR**E**- and lacks the aspartate rich DDXXD motif. The *cotB2* gene is present in the cyclooctatin biosynthetic gene cluster with *cotB1* encoding GGPP synthase, whereas *cotB3* and *cotB4* each encodes for a cytochrome P450. GGPP on incubation with recombinant CotB2 undergoes Mg^{2+} dependent cyclisation to form cyclooctat-9-en-7-ol (Fig. 5, Scheme VII), precursor for the lysophospholipase inhibitor, cyclooctatin.

Bonnadiene, phomposene and cembrene A biosynthesis

Two DTS genes WP_030432512 and WP_030431358 were identified by genome mining approach in a tropical soil actinomycete *Allokutzneria albata* DSM44149 and the heterologously expressed products subjected to in-depth mechanistic studies which helped to characterise these genes to be Bonnadiene synthase (Bds) and Phomopsene synthase (Pms) (Lauterbach et al. 2018). The enzyme Bds converted GGPP into a diterpene hydrocarbon bonnadiene, having a new spirotricyclic skeleton. However, the second enzyme Pms was found to catalyse the 1,11- and 10,14-cyclisation of GGPP into a mixture of compounds: a fungal diterpene phomopsene, the new compound allokutznerene and a previously known diterpene, spiroviolene. In an another study by the same group, two DTS genes Sas (WP_030426588) and CAS (WP_030430753) were identified in the same actinobacteria *A. albata* (Rinkel et al. 2018). Both the genes were cloned into the expression vector *p*YE-Express by homologous recombination in yeast, which were further expressed in *E. coli* and the purified protein was incubated with GGPP. First gene product was found to convert GGPP into a diterpene hydrocarbon, spiroalbatene, having a new spirotetracyclic skeleton and thus the catalysing enzyme was named spiroalbatene synthase (Sas). Second enzyme was found to catalyse the cyclisation of GGPP to cembrene A and thus characterised as cembrene A synthase (CAS).

Conclusions

Extensive bacterial genome sequencing and genome data mining has facilitated the elicitation of a wealth of hidden natural products. The advancement of bioinformatics and biotechnology helped to characterise several natural product BGCs, which encodes several key enzymes that catalyse the production of terpene metabolites. However, bioinformatics analysis alone could not possibly attribute a particular biochemical function to majority of newly identified TPS genes. The inability to detect one terpene in a wild-type bacterium might plainly contemplate, tremendously low levels of terpene gene expression or vying utilisation of common allylic precursor by the synthesis of other different terpenes. The catalytic function of these putative TPS genes could be ascertained either by subsequent cloning and expression in *E. coli* or by expression in a particularly modified heterologous *Streptomyces* host, followed by spectroscopic identification of the newly generated metabolic products. The actual catalytic function of the presumptive TPS can be determined by incubation of the recombinant synthase with the allylic precursors. Thus, the combinatorial strategy of applying bioinformatic tools, genetic engineering and mechanistic enzymology would continually enhance the knowledge and could unveil the interrelatedness among primary sequence of TPS, functional domain organisation, protein structure and cyclisation process. An extensive bacterial genome data is available today and this has led to the synthesis of several terpene compounds with huge industrial applications. The above findings elaborate our canonical understanding of the structural elements required for the catalysis and also guide the mechanistic studies for the discovery of novel bacterial terpene synthases, but there are many dark areas too in this interesting class of enzymes for further exploration.

Acknowledgements PB acknowledges the independent BSR-RFSMS fellowship (F.7-209/2009(BSR)), awarded by the University Grants Commission (UGC). The financial assistance received from Department of Science and Technology-Promotion of University Research and Scientific Excellence (DST-PURSE) and University Grants Commission-Special Assistance Programme (UGC-SAP) (DRS Phase-I) is duly acknowledged. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions PB and DS conceived of the presented idea for the article. PB performed the literature search and data analysis. PB wrote the manuscript; DS provided critical feedback and helped to shape the manuscript.

Compliance with ethical standards

Conflict of interest No potential conflict of interest is being reported by the authors.

References

Asanova ZK, Suleimenov EM, Atazhanova GA et al (2003) Biological activity of 1,8-cineole from levant wormwood. Pharm Chem J 37:28–30. https://doi.org/10.1023/A:1023699012354

Atzorn R, Crozier A, Wheeler CT, Sandberg G (1988) Production of gibberellins and indole-3-acetic acid by *Rhizobium phaseoli* in relation to nodulation of *Phaseolus vulgaris* roots. Planta 175:532–538. https://doi.org/10.1007/BF00393076

Barton DHR, Bruun T, Lindsey AS (1952) Sesquiterpenoids. Part II. Tricyclic derivatives of caryophyllene. J Chem Soc. https://doi.org/10.1039/jr9520002210

Bentley R, Meganathan R (1981) Geosmin and methylisoborneol biosynthesis in streptomycetes. FEBS Lett 125:220–222

Bicas JL, Dionísio AP, Pastore GM (2009) Bio-oxidation of terpenes: an approach for the flavor Industry. Chem Rev 109:4518–4531. https://doi.org/10.1021/cr800190y

Cane DE (1985) Isoprenoid biosynthesis. stereochemistry of the cyclization of allylic pyrophosphates. Acc Chem Res 18:220–226. https://doi.org/10.1021/ar00115a005

Cane DE, He X, Kobayashi S et al (2006) Geosmin biosynthesis in *Streptomyces avermitilis*. Molecular cloning, expression, and mechanistic study of the germacradienol/geosmin synthase. J Antibiot (Tokyo) 59:471–479. https://doi.org/10.1038/ja.2006.66

Cane DE, Ikeda H (2012) Exploration and mining of the bacterial terpenome. Acc Chem Res 45:463–472. https://doi.org/10.1021/ar200198d

Cane DE, Oliver JS, Harrison PM et al (1990) Biosynthesis of pentalenene and pentalenolactone. J Am Chem Soc 112:4513–4524. https://doi.org/10.1021/ja00167a059

Cane DE, Pargellis C (1987) Partial purification and characterization of pentalenene synthase. Arch Biochem Biophys 254:421–429. https://doi.org/10.1016/0003-9861(87)90120-2

Cane DE, Sohng J-K (1989) Inhibition of glyceraldehyde-3-phosphate dehydrogenase pentalenolactone: kinetic and mechanistic studies. Arch Biochem Biophys 270:50–61

Cane DE, Sohng J, Lamberson CR et al (1994) High-level expression in *Escherichia coli* of a terpenoid cyclase from. Biochemistry 33:5846–5857

Cane DE, Tandon M, Prabhakaran PC (1993) Epicubenol synthase and the enzymatic cyclization of farnesyl diphosphate. J Am Chem Soc 115:8103–8106. https://doi.org/10.1021/ja00125a029

Cane DE, Watt RM (2003) Expression and mechanistic analysis of a germacradienol synthase from *Streptomyces coelicolor* implicated in geosmin biosynthesis. Proc Natl Acad Sci USA 100:1547–1551. https://doi.org/10.1073/pnas.0337625100

Chou WKW, Fanizza I, Uchiyama T 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 132:8850–8851. https://doi.org/10.1021/ja103087w

Chou WKW, Ikeda H, Cane DE (2011) Cloning and characterization of Pfl_1841, a 2-methylenebornane synthase in *Pseudomonas fluorescens* PfO-1. Tetrahedron 67:6627–6632. https://doi.org/10.1016/j.tet.2011.05.084

Christianson DW (2006) Structural biology and chemistry of the terpenoid cyclases. Chem Rev 106:3412–3442. https://doi.org/10.1021/cr050286w

Citron CA, Barra L, Wink J, Dickschat JS (2015) Volatiles from nineteen recently genome sequenced actinomycetes. Org Biomol Chem 13:2673–2683. https://doi.org/10.1039/c4ob02609h

Citron CA, Gleitzmann J, Laurenzano G et al (2012) Terpenoids are widespread in actinomycetes: a correlation of secondary metabolism and genome data. ChemBioChem 13:202–214. https://doi.org/10.1002/cbic.201100641

Corre C, Challis GL (2009) New natural product biosynthetic chemistry discovered by genome mining. Nat Prod Rep 26:977–986. https://doi.org/10.1039/b713024b

Dairi T, Hamano Y, Kuzuyama T et al (2001) Eubacterial diterpene cyclase genes essential for production of the isoprenoid antibiotic terpentecin. J Bacteriol 183:6085–6094. https://doi.org/10.1128/JB.183.20.6085-6094.2001

Daum M, Herrmann S, Wilkinson B, Bechthold A (2009) Genes and enzymes involved in bacterial isoprenoid biosynthesis. Curr Opin Chem Biol 13:180–188. https://doi.org/10.1016/j.cbpa.2009.02.029

Davis EM, Croteau R (2000) Cyclization enzymes in the biosynthesis of monoterpenes, sesquiterpenes, and diterpenes. In: Leeper FJ, Vederas JC (eds) Topics in current chemistry. Springer, Berlin, pp 53–95

Degenhardt J, Köllner TG, Gershenzon J (2009) Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants. Phytochemistry 70:1621–1637. https://doi.org/10.1016/j.phytochem.2009.07.030

Dickschat JS (2016) Bacterial terpene cyclases. Nat Prod Rep 33:87–110. https://doi.org/10.1039/c5np00102a

Duell ER, D'Agostino PM, Shapiro N et al (2019) Direct pathway cloning of the sodorifen biosynthetic gene cluster and recombinant generation of its product in *E. coli*. Microb Cell Fact 18:1–11. https://doi.org/10.1186/s12934-019-1080-6

Eddy SR (1998) Profile hidden Markov models. Bioinformatics 14:755–763. https://doi.org/10.1093/bioinformatics/14.9.755

Finn RD, Mistry J, Tate J et al (2009) The Pfam protein families database. Nucleic Acids Res 38:211–222. https://doi.org/10.1093/nar/gkp985

Gerber NN (1967) Geosmin, an earthy-smelling substance isolated from actinomycetes. Biotechnol Bioeng 9:321–327. https://doi.org/10.1002/bit.260090305

Gerber NN (1968) Geosmin, from microorganisms, is trans-1, 10-dimethyl-trans-9-decalol. Tetrahedron Lett 9:2971–2974. https://doi.org/10.1016/S0040-4039(00)89625-2

Gerber NN (1971) Sesquiterpenoids from actinomycetes: cadin-4-ene-1-ol. Phytochemistry 10:185–189

Gerber NN (1973) Volatile lactones from *Streptomyces*. Tetrahedron Lett 14:771–774. https://doi.org/10.1016/S0040-4039(01)95708-9

Gerber NN, Lechevalier HA (1965) Geosmin, earthy-smelling substance isolated from actinomycetes. Appl Microbiol 13:935–938

Giglio S, Jiang J, Saint CP et al (2008) Isolation and characterization of the gene associated with geosmin production in cyanobacteria. Environ Sci Technol 42:8027–8032. https://doi.org/10.1021/es801465w

Gurtler H, Pedersen R, Anthoni U et al (1994) Albaflavenone, a sesquiterpene ketone with a zizaene skeleton produced by a streptomycete with a new rope morphology. J Antibiot (Tokyo) 47:434–439

Gust B, Challis GL, Fowler K et al (2003) PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. Proc Natl Acad Sci 100:1541–1546. https://doi.org/10.1073/pnas.0337542100

Hamano Y, Kuzuyama T, Itoh N et al (2002) Functional analysis of eubacterial diterpene cyclases responsible for biosynthesis of a diterpene antibiotic, terpentecin. J Biol Chem 277:37098–37104. https://doi.org/10.1074/jbc.M206382200

Harada H, Misawa N (2009) Novel approaches and achievements in biosynthesis of functional isoprenoids in *Escherichia coli*. Appl Microbiol Biotechnol 84:1021–1031. https://doi.org/10.1007/s00253-009-2166-6

Hartmann S, Neeff J, Heer U, Mecke D (1978) Arenaemycin (pentalenolactone): a specific inhibitor of glycolysis. FEBS Lett 93:339–342. https://doi.org/10.1016/0014-5793(78)81135-1

Helfrich EJN, Lin GM, Voigt CA, Clardy J (2019) Bacterial terpene biosynthesis: Challenges and opportunities for pathway engineering. Beilstein J Org Chem 15:2889–2906. https://doi.org/10.3762/bjoc.15.283

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 18:32–37. https://doi.org/10.1016/j.chembiol.2010.11.008

Ikeda C, Hayashi Y, Itoh N et al (2007) Functional analysis of eubacterial ent-copalyl diphosphate synthase and pimara-9(11),15-diene synthase with unique primary sequences. J Biochem 141:37–45. https://doi.org/10.1093/jb/mvm004

Ikeda H (2017) Natural products discovery from micro-organisms in the post-genome era. Biosci Biotechnol Biochem 81:13–22. https://doi.org/10.1080/09168451.2016.1248366

Ikeda H, Ishikawa J, Hanamoto A et al (2003) Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. Nat Biotechnol 21:526–531. https://doi.org/10.1038/nbt820

Jiang J, Cane DE (2008) Geosmin biosynthesis. Mechanism of the fragmentation—rearrangement in the conversion of Germacradienol to Geosmin. J Am Chem Soc 130:428–429. https://doi.org/10.1021/ja077792i.Geosmin

Kai M, Crespo E, Cristescu SM et al (2010) *Serratia odorifera*: analysis of volatile emission and biological impact of volatile compounds on *Arabidopsis thaliana*. Appl Microbiol Biotechnol 88:965–976. https://doi.org/10.1007/s00253-010-2810-1

Kamatou GPP, Viljoen AM (2010) A review of the application and pharmacological properties of α-bisabolol and α-bisabolol-rich oils. JAOCS, J Am Oil Chem Soc 87:1–7. https://doi.org/10.1007/s11746-009-1483-3

Karahadian C, Josephson DB, Lindsay RC (1985) Volatile compounds from *Pencillium* sp. contributing musty-earthy notes to brie and camembert cheese flavors. J Agric Food Chem 33:339–343. https://doi.org/10.1021/jf00063a005

Kim SY, Zhao P, Igarashi M et al (2009) Cloning and heterologous expression of the cyclooctatin biosynthetic gene cluster afford a diterpene cyclase and two P450 hydroxylases. Chem Biol 16:736–743. https://doi.org/10.1016/j.chembiol.2009.06.007

Knorre DG, Kudryashova NV, Godovikova TS (2009) Chemical and functional aspects of posttranslational modification of proteins. Acta Naturae 1:29–51. https://doi.org/10.32607/20758251-2009-1-3-29-51

Köllner TG, Held M, Lenk C et al (2008) A maize (E)-β-caryophyllene synthase implicated in indirect defense responses against herbivores is not expressed in most American maize varieties. Plant Cell 20:482–494. https://doi.org/10.1105/tpc.107.051672

Komatsu M, Tsuda M, Omura S et al (2008) Identification and functional analysis of genes controlling biosynthesis of 2-methylisoborneol. Proc Natl Acad Sci USA 105:7422–7427. https://doi.org/10.1073/pnas.0802312105

Komatsu M, Uchiyama T, Omura S et al (2010) Genome-minimized *Streptomyces* host for the heterologous expression of secondary metabolism. Proc Natl Acad Sci USA 107:2646–2651. https://doi.org/10.1073/pnas.0914833107

Krist S, Sato K, Glasl S et al (2008) Antimicrobial effect of vapours terpeniol, (R)-(-)-linalool, carvacrol, (S)-(-)-perillaldehyde and 1,8-cineole on airborne microbes using a room diffuser. Flavour Fragr J 23:353–356

Lahlou S, Figueiredo AF, Magalhães PJC, Leal-Cardoso JH (2002) Cardiovascular effects of 1,8-cineole, a terpenoid oxide present in many plant essential oils, in normotensive rats. Can J Physiol Pharmacol 80:1125–1131. https://doi.org/10.1139/y02-142

Lauterbach L, Rinkel J, Dickschat JS (2018) Two bacterial diterpene synthases from *Allokutzneria albata* produce bonnadiene, phomopsene, and allokutznerene. Angew Chemie Int Ed 57:8280–8283. https://doi.org/10.1002/anie.201803800

Lesburg CA, Zhai G, Cane DE, Christianson DW (1997) Crystal structure of pentalenene synthase: mechanistic insights on terpenoid cyclization reactions in biology. Science (80-) 277:1820–1824. https://doi.org/10.1126/science.277.5333.1820

Lin X, Hopson R, Cane DE (2006) Genome mining in *Streptomyces coelicolor* : molecular cloning and characterization of a new sesquiterpene synthase. J Am Chem Soc 128:6022–6023. https://doi.org/10.1021/ja061292s

Martin DG, Slomp G, Mizsak S et al (1970) The structure and absolute configuration of pentalenolactone (PA 132). Tetrahedron Lett 56:4901–4904

Medsker LL, Jenkins D, Thomas JF, Koch C (1969) Odorous compounds in natural waters 2-exo-hydroxy-2-methylbornane, the major odorous compound produced by several actinomycetes. Environ Sci Technol 3:476–477. https://doi.org/10.1021/es60028a008

Morrone D, Chambers J, Lowry L et al (2009) Gibberellin biosynthesis in bacteria: Separate ent-copalyl diphosphate and ent-kaurene synthases in *Bradyrhizobium japonicum*. FEBS Lett 583:475–480. https://doi.org/10.1016/j.febslet.2008.12.052

Nabeta K, Fujita M, Komuro K et al (1997) In vitro biosynthesis of cadinanes by cell-free extracts of cultured cells of *Heteroscyphus planus*. J Chem Soc - Perkin Trans 1:2065–2070. https://doi.org/10.1039/a701145h

Nakano C, Hoshino T (2009) Characterization of the Rv3377c gene product, a type-B diterpene cyclase, from the *Mycobacterium tuberculosis* H37 genome. ChemBioChem 10:2060–2071. https://doi.org/10.1002/cbic.200900248

Nakano C, Kim HK, Ohnishi Y (2011a) Identification and characterization of the linalool/nerolidol synthase from *Streptomyces clavuligerus*. ChemBioChem 12:2403–2407. https://doi.org/10.1002/cbic.201100501

Nakano C, Kim HK, Ohnishi Y (2011b) Identification of the first bacterial monoterpene cyclase, a 1,8-cineole synthase, that catalyzes the direct conversion of geranyl diphosphate. ChemBioChem 12:1988–1991. https://doi.org/10.1002/cbic.201100330

Nakano C, Kudo F, Eguchi T, Ohnishi Y (2011c) Genome mining reveals two novel bacterial sesquiterpene cyclases: (-)-germacradien-4-ol and (-)-epi-α-bisabolol synthases from *Streptomyces citricolor*. ChemBioChem 12:2271–2275. https://doi.org/10.1002/cbic.201100418

Nakano C, Okamura T, Sato T et al (2005) *Mycobacterium tuberculosis* H37Rv3377c encodes the diterpene cyclase for producing the halimane skeleton. Chem Commun. https://doi.org/10.1039/b415346d

Nakano C, Ootsuka T, Takayama K et al (2011d) Characterization of the Rv3378c gene product, a new diterpene synthasefor producing tuberculosinol and (13R, S)-isotuberculosinol (nosyberkol), from the *Mycobacterium tuberculosis* H37Rv genome. Biosci Biotechnol Biochem 75:75–81. https://doi.org/10.1271/bbb.100570

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) 65:551–558. https://doi.org/10.1038/ja.2012.68

Omura S, Ikeda H, Ishikawa J et al (2001) Genome sequence of an industrial microorganism *Streptomyces avermitilis*: Deducing the ability of producing secondary metabolites. Proc Natl Acad Sci USA 98:12215–12220. https://doi.org/10.1073/pnas.211433198

Pollak FC, Berger RG (1996) Geosmin and related volatiles in bioreactor-cultured *Streptomyces citreus* CBS 109.60. Appl Environ Microbiol 62:1295–1299

Punta M, Coggill PC, Eberhardt RY et al (2012) The Pfam protein families database. Nucleic Acids Res 40:D290–D301. https://doi.org/10.1093/nar/gkr1065

Rabe P, Citron CA, Dickschat JS (2013) Volatile terpenes from actinomycetes: a biosynthetic study correlating chemical analyses to genome data. ChemBioChem 14:2345–2354. https://doi.org/10.1002/cbic.201300329

Rabe P, Schmitz T, Dickschat JS (2016) Mechanistic investigations on six bacterial terpene cyclases. Beilstein J Org Chem 12:1839–1850. https://doi.org/10.3762/bjoc.12.173

Rinkel J, Lauterbach L, Rabe P, Dickschat JS (2018) Two diterpene synthases for spiroalbatene and cembrene A from *Allokutzneria albata*. Angew Chemie Int Ed 57:3238–3241. https://doi.org/10.1002/anie.201800385

Roach JJ, Shenvi RA (2015) Supramolecular catalysis: terpenes in tight spaces. Nat Chem 7:187–189. https://doi.org/10.1038/nchem.2191

Rudolf JD, Bin DL, Cao H et al (2016) Structure of the ent-Copalyl diphosphate synthase PtmT2 from *Streptomyces platensis* CB00739, a bacterial type II diterpene synthase. J Am Chem Soc 138:10905–10915. https://doi.org/10.1021/jacs.6b04317

Sacchettini JC, Poulter CD (1997) Creating lsoprenoid diversity. Science (80-) 277:1788–1789

Santos FA, Rao VSN (2000) Antiinflammatory and antinociceptive effects of 1,8-cineole a terpenoid oxide present in many plant essential oils. Phyther Res 14:240–244. https://doi.org/10.1002/1099-1573(200006)14:4<240:AID-PTR573>3.0.CO;2-X

Schöller CEG, Gürtler H, Pedersen R et al (2002) Volatile metabolites from actinomycetes. J Agric Food Chem 50:2615–2621. https://doi.org/10.1021/jf0116754

Shah AJ, Gilani AH, Abbas K et al (2011) Studies on the chemical composition and possible mechanisms underlying the antispasmodic and bronchodilatory activities of the essential oil of *Artemisia maritima* L. Arch Pharm Res 34:1227–1238. https://doi.org/10.1007/s12272-011-0801-0

Sharkey TD, Wiberley AE, Donohue AR (2008) Isoprene Emission from Plants: Why and How. Ann Bot 101:5–18. https://doi.org/10.1093/aob/mcm240

Singh M, Chaudhary S, Sareen D (2017) Review: non-ribosomal peptide synthetases: identifying the cryptic gene clusters and decoding the natural product. J Biosci 42(1):175–187

Smanski MJ, Peterson RM, Huang SX, Shen B (2012) Bacterial diterpene synthases: New opportunities for mechanistic enzymology and engineered biosynthesis. Curr Opin Chem Biol 16:132–141. https://doi.org/10.1016/j.cbpa.2012.03.002

Smanski MJ, Yu Z, Casper J et al (2011) Dedicated ent-kaurene and ent-atiserene synthases for platensimycin and platencin biosynthesis. Proc Natl Acad Sci USA 108:13498–13503. https://doi.org/10.1073/pnas.1106919108

Starks CM, Back K, Chappell J, Noel JP (1997) Structural basis for cyclic terpene biosynthesis by. Science (80-) 277:1815–1820

Su V, King D, Woodrow I et al (2008) *Plasmodium falciparum* growth is arrested by monoterpenes from eucalyptus oil. Flavour Fragr J 23:315–318. https://doi.org/10.1002/ffj.1880

Takamatsu S, Lin X, Nara A 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 4:184–191. https://doi.org/10.1111/j.1751-7915.2010.00209.x

Takao Y, Kuriyama I, Yamada T et al (2012) Antifungal properties of Japanese cedar essential oil from waste wood chips made from used sake barrels. Mol Med Rep 5:1163–1168. https://doi.org/10.3892/mmr.2012.821

Takeuchi S, Ogawa Y, Yonehara H (1969) The structure of pentalenolactone (PA-132). Tetrahedron Lett 10:2737–2740. https://doi.org/10.1016/S0040-4039(01)88256-3

Tetzlaff CN, You Z, Cane DE et al (2006) A gene cluster for biosynthesis of the sesquiterpenoid antibiotic pentalenolactone in *Streptomyces avermitilis*. Biochemistry 45:6179–6186. https://doi.org/10.1021/bi060419n

Von Reuß SH, Kai M, Piechulla B, Francke W (2010) Octamethylbicyclo[3.2.1]octadienes from the rhizobacterium *Serratia odorifera*. Angew Chemie - Int Ed 49:2009–2010. https://doi.org/10.1002/anie.200905680

Wang C, Liwei M, Park J-B et al (2018) Microbial platform for terpenoid production: *Escherichia coli* and yeast. Front Microbiol 9:1–8. https://doi.org/10.3389/fmicb.2018.02460

Wender PA, Badham NF, Conway SP et al (1997a) The pinene path to taxanes. 5. Stereocontrolled synthesis of a versatile taxane precursor. J Am Chem Soc 119:2755–2756. https://doi.org/10.1021/ja9635387

Wender PA, Badham NF, Conway SP et al (1997b) The pinene path to taxanes. 6. A concise stereocontrolled synthesis of taxol. J Am Chem Soc 119:2757–2758. https://doi.org/10.1021/ja963539z

Wendt KU, Schulz GE, Corey EJ, Liu DR (2000) Enzyme mechanisms for polycyclic triterpene formation. Angew Chem Int Ed 39:2812–2833

Wilkins K, Schöller C (2009) Volatile organic metabolites from selected *Streptomyces* strains. Actinomycetologica 23:27–33. https://doi.org/10.3209/saj.saj230202

Wuts PG (1998) Semisynthesis of taxol. Curr Opin Drug Discov Devel 1:329–337

Yamada Y, Arima S, Nagamitsu T et al (2015a) Novel terpenes generated by heterologous expression of bacterial terpene synthase genes in an engineered *Streptomyces* host. J Antibiot (Tokyo) 68:385–394. https://doi.org/10.1038/ja.2014.171

Yamada Y, Kuzuyama T, Komatsu M et al (2015b) Terpene synthases are widely distributed in bacteria. Proc Natl Acad Sci 112:857–862. https://doi.org/10.1073/pnas.1422108112

Yang D, Michel L, Chaumont JP, Millet-Clerc J (2000) Use of caryophyllene oxide as an antifungal agent in an in vitro experimental model of onychomycosis. Mycopathologia 148:79–82. https://doi.org/10.1023/A:1007178924408

Zhao B, Lin X, Lei L et al (2008) Biosynthesis of the sesquiterpene antibiotic albaflavenone in *Streptomyces coelicolor* A3(2). J Biol Chem 283:8183–8189. https://doi.org/10.1074/jbc.M710421200