



Phylogenomic analyses and distribution of terpene synthases among *Streptomyces*

Lara Martín-Sánchez^{‡1}, Kumar Saurabh Singh^{‡2}, Mariana Avalos^{1,3}, Gilles P. van Wezel^{1,3}, Jeroen S. Dickschat^{*1,4} and Paolina Garbeva^{*1}

Full Research Paper

[Open Access](#)

Address:

¹Department of Microbial Ecology, Netherlands Institute of Ecology (NIOO-KNAW), Droevendaalsesteeg 10, 6708 PB Wageningen, The Netherlands, ²College of Life and Environmental Sciences, Biosciences, University of Exeter, Penryn Campus, Penryn, Cornwall TR10 9FE, United Kingdom, ³Institute of Biology, Leiden University, Sylviusweg 72, 2333 BE Leiden, The Netherlands and ⁴University of Bonn, Kekulé-Institute of Organic Chemistry and Biochemistry, Gerhard-Domagk-Straße 1, 53121 Bonn, Germany

Email:

Jeroen S. Dickschat* - dickschat@uni-bonn.de; Paolina Garbeva* - P.Garbeva@nioo.knaw.nl

* Corresponding author ‡ Equal contributors

Keywords:

biosynthesis; evolution; geosmin; *Streptomyces*; terpenes

Beilstein J. Org. Chem. **2019**, *15*, 1181–1193.

doi:10.3762/bjoc.15.115

Received: 14 March 2019

Accepted: 17 May 2019

Published: 29 May 2019

This article is part of the thematic issue "Terpenes".

Associate Editor: K. N. Allen

© 2019 Martín-Sánchez et al.; licensee Beilstein-Institut.

License and terms: see end of document.

Abstract

Terpene synthases are widely distributed among microorganisms and have been mainly studied in members of the genus *Streptomyces*. However, little is known about the distribution and evolution of the genes for terpene synthases. Here, we performed whole-genome based phylogenetic analysis of *Streptomyces* species, and compared the distribution of terpene synthase genes among them. Overall, our study revealed that ten major types of terpene synthases are present within the genus *Streptomyces*, namely those for geosmin, 2-methylisoborneol, *epi*-isozizaene, 7-*epi*- α -eudesmol, *epi*-cubenol, caryolan-1-ol, cyclooctat-9-en-7-ol, isoafrikanol, pentalenene and α -amorphene. The *Streptomyces* species divide in three phylogenetic groups based on their whole genomes for which the distribution of the ten terpene synthases was analysed. Geosmin synthases were the most widely distributed and were found to be evolutionary positively selected. Other terpene synthases were found to be specific for one of the three clades or a subclade within the genus *Streptomyces*. A phylogenetic analysis of the most widely distributed classes of *Streptomyces* terpene synthases in comparison to the phylogenomic analysis of this genus is discussed.

Introduction

Streptomyces are soil bacteria that belong to the order of actinomycetales and are a rich source of natural products with broad biotechnological interest. Species of this genus have a

remarkable genetic potential to produce a large variety of secondary metabolites with different functions including antibiotics, antifungals, pigments or immunosuppressants [1-3]. These

are compounds of diverse chemical nature such as polyketides, peptides, aminoglycosides or terpenoids [4,5].

Terpenoids are the largest and the most diverse class of natural compounds known to date and include the initial products of terpene synthases and all derivatives made from them in tailoring steps. This very diverse class of organic compounds is best known as plant metabolites. However, recent studies revealed that terpenoids can be produced by all kingdoms of life including bacteria, fungi and protists [6–10]. The ability of an organism to produce terpenoids relies on whether the organism contains terpene synthase genes. Biosynthetically, the production of the different types of terpenes depends on the precursors that these synthases can accommodate: geranyl diphosphate (monoterpenes, C10), farnesyl diphosphate (sesquiterpenes, C15) and geranylgeranyl diphosphate (diterpenes, C20). The biological function of terpenes is best studied for plants where they play important roles in aboveground plant–insect, plant–pathogen and plant–plant interactions [11]. However, terpenes might also play important roles in belowground inter-specific interactions [12]. Terpene synthases are in fact widely distributed among soil microorganisms, and they have been mainly studied in *Streptomyces* species [9,13]. Some volatile terpenes, such as geosmin and 2-methylisoborneol (2-MIB), responsible for the smell of wet soil after rain, have been known for a long time to be produced by *Streptomyces* species [14,15]. Many terpene synthases from *Streptomyces* have been studied and characterised [13]. However, little is known about the distribution of terpene synthase encoding genes among *Streptomyces*. Are terpene synthase genes specific for certain species or randomly distributed among *Streptomyces*?

To address this question, phylogenomic analyses of *Streptomyces* species were performed, using complete genomes available in the NCBI database and compared the distribution of terpene synthase genes among them. Furthermore, we studied whether phylogenetic trees calculated based on the three most abundant terpene synthases in *Streptomyces* represent the evolution of the *Streptomyces* species based on the whole genome-based phylogenetic analyses.

Results and Discussion

Whole genome-based phylogenetic analyses of *Streptomyces* species

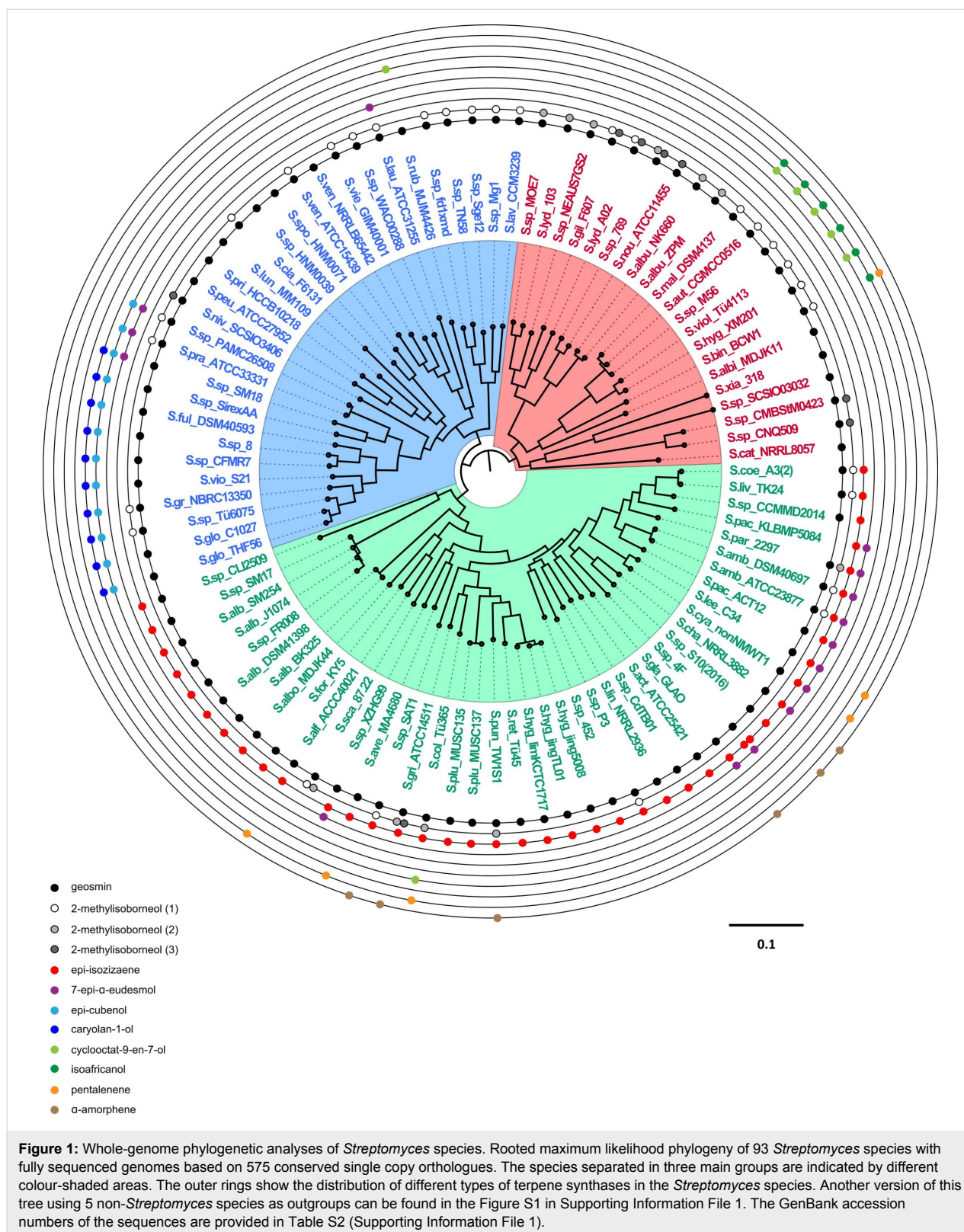
Genome sequences from 93 *Streptomyces* species for which a complete genome was available (represented by a single scaffold and a complete list of annotated protein sequences), were selected to construct a whole genome-based phylogenetic tree (Figure 1). The NCBI database was accessed on September 30th 2018. An orthologues-based approach was adopted to generate a species tree using OrthoFinder. OrthoFinder resulted in a total

of 19980 orthologue groups (Table S1, Supporting Information File 1). A total number of 575 single copy orthologues were further selected for the generation of the species tree. Based on these phylogenetic analyses, the *Streptomyces* species clustered in three different clades (indicated in blue, green and red in Figure 1). This separation into three different clades agrees with the study previously reported by McDonald and Currie, 2017 [16]. Based on phylogenetic analyses of 94 housekeeping genes, they showed a separation of *Streptomyces* species in two major clades and a third group of other lineages.

Distribution of terpene synthases in *Streptomyces*

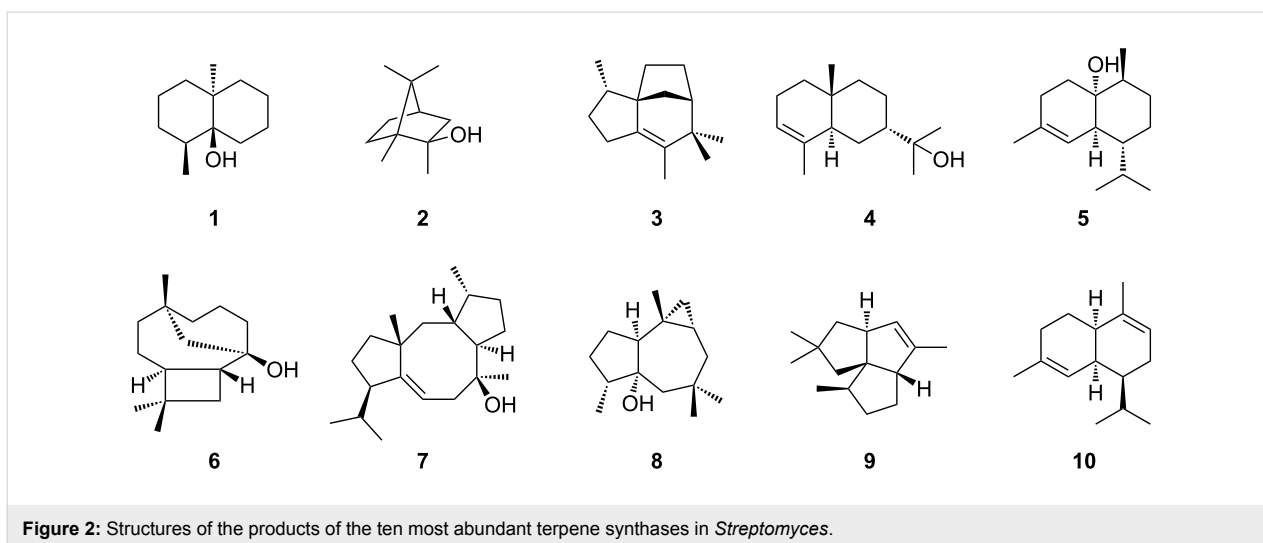
We analysed the distribution of the different types of terpene synthases among *Streptomyces* species with complete genomes. Besides a few rare terpene synthases only occurring in a few or a single species, ten major types of terpene synthases were present among these *Streptomyces* species, including the terpene synthases for geosmin (**1**), 2-methylisoborneol (2-MIB) (**2**), *epi*-isozizaene (**3**), 7-*epi*- α -eudesmol (**4**), *epi*-cubenol (**5**), caryolan-1-ol (**6**), cyclooctat-9-en-7-ol (**7**), isoafricanol (**8**), pentalenene (**9**) and α -amorphene (**10**) (Figure 1 and Figure 2).

The geosmin synthases were the most widely distributed, as they were present in all except one of the *Streptomyces* species (*S. pactum* KLBMP 5084) (Figure 1). This finding suggests that geosmin may have an important ecological function as a chemical signal or as protective specialised metabolite against biotic and abiotic stresses, similarly as the roles played by terpenoids in plants [11]. However, although geosmin was discovered more than 50 years ago [14], its biological or ecological function still remains unclear. *Streptomyces pactum* KLBMP 5084 (the only species included in this study that does not carry geosmin synthases) is an endophytic plant growth-promoting bacterium that provides salt tolerance to the halophytic plant *Limonium sinense* (Plumbaginaceae) [17]. The absence of a geosmin synthase in this bacterium leads us to hypothesise that the role of geosmin may be complemented by the plant host. The only other plant endophyte among the 93 species is *Streptomyces* sp. SAT1 (see Table S9 (Supporting Information File 1) for a list of the isolation sources and habitats of the 93 strains). This strain is an endophyte of the flowering plant *Adenophora trachelioides* from the Campanulaceae family and it does contain a copy of *geoA*, the gene encoding for geosmin synthase. Some species such as *Streptomyces* sp. SirexAA-E harbour a silent geosmin synthase encoding gene in their genomes and do not produce this degraded sesquiterpene under laboratory culture conditions [9]. It will therefore be interesting to investigate whether the geosmin synthase in *Streptomyces* sp. SAT1 is expressed, and to further determine the role of terpenoids in the endophytic life style.



The first geosmin synthase was characterised from *Streptomyces coelicolor* [18]. Geosmin synthases are composed of two domains that both exhibit the typical highly conserved motifs of

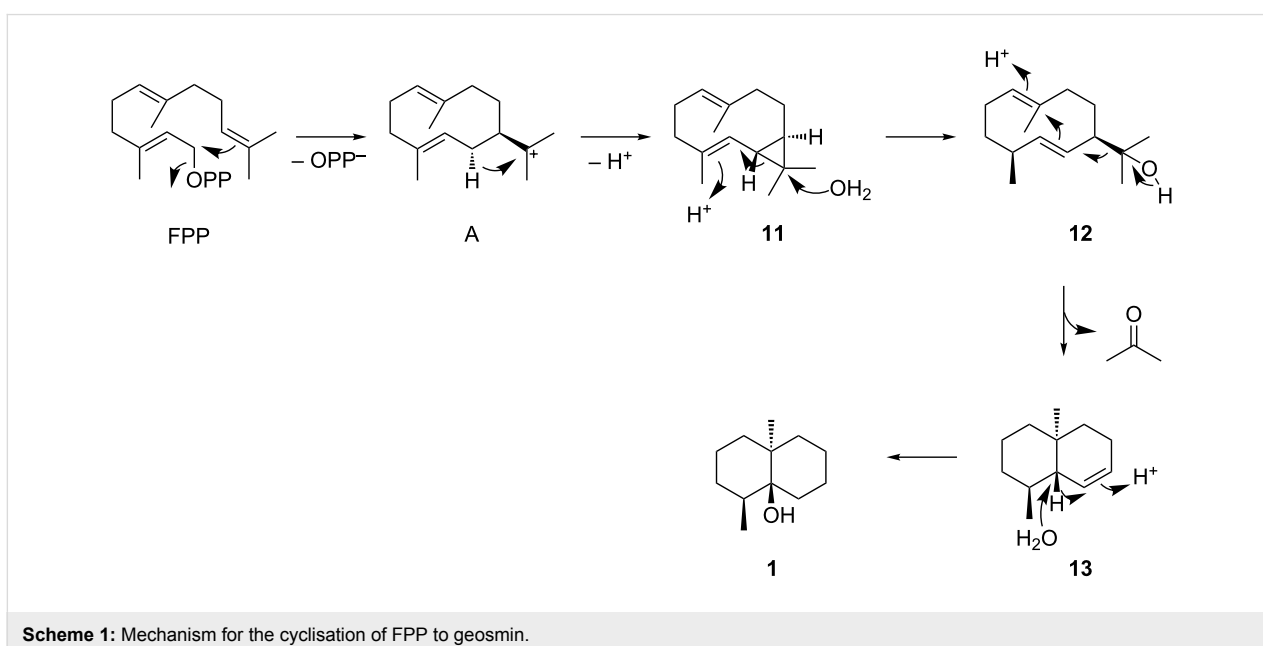
type I terpene synthases, including the aspartate-rich motif, the NSE triad, the pyrophosphate sensor and the RY pair [19–21]. Both domains have a catalytic activity, the N-terminal domain

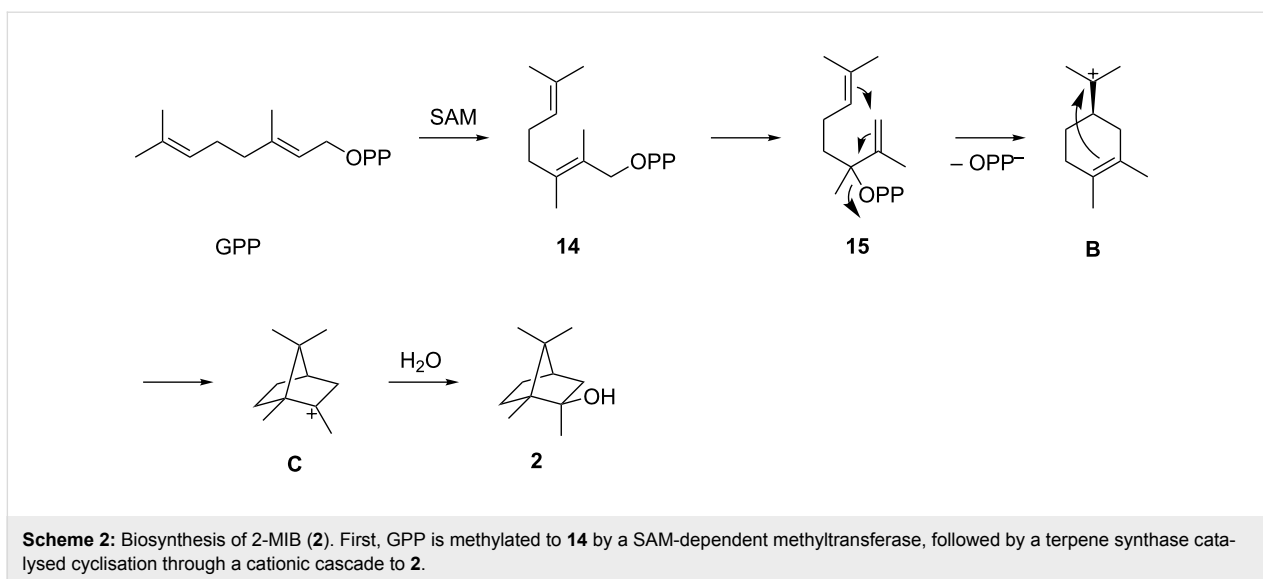


for the conversion of FPP into the intermediate sesquiterpene alcohol (1(10)*E*,5*E*)-germacradien-11-ol (**12**), and the C-terminal domain for its further conversion into geosmin with cleavage of **12** into acetone and the octalin **13** through a retro-Prins fragmentation (Scheme 1) [22–24]. The proposed neutral intermediate isolepidozene (**11**) has so far only been reported from the S233A enzyme variant of geosmin synthase from *S. coelicolor* [18].

The second most widely distributed terpene synthases are the 2-MIB synthases (Figure 1). As discussed below, the phylogenetic analysis of 2-MIB synthases classifies these enzymes into three different groups. This distribution is also indicated in Figure 1 (white, light gray and dark gray circles). The 2-MIB

synthases are present in members of all the three clades from the whole genome phylogenetic tree (Figure 1), but are most abundant in members of the clade depicted in red. These terpene synthases catalyse a unique cyclisation reaction utilizing the modified substrate 2-methyl-GPP to form 2-MIB (**2**) [25,26]. An *S*-adenosylmethionine (SAM) dependent methyl transferase is responsible for the methylation of GPP into 2-methyl-GPP (**14**, Scheme 2). Its isomerisation to **15** allows for a cyclisation via the cationic intermediates **B** and **C** to **2**. Genes encoding for SAM-dependent methyl transferases were found forming a cluster together with the 2-MIB synthase in several *Streptomyces* species [26,27]. Besides the C-terminal domain typical of class I terpene synthases, these enzymes contain an additional proline-rich N-terminal domain that



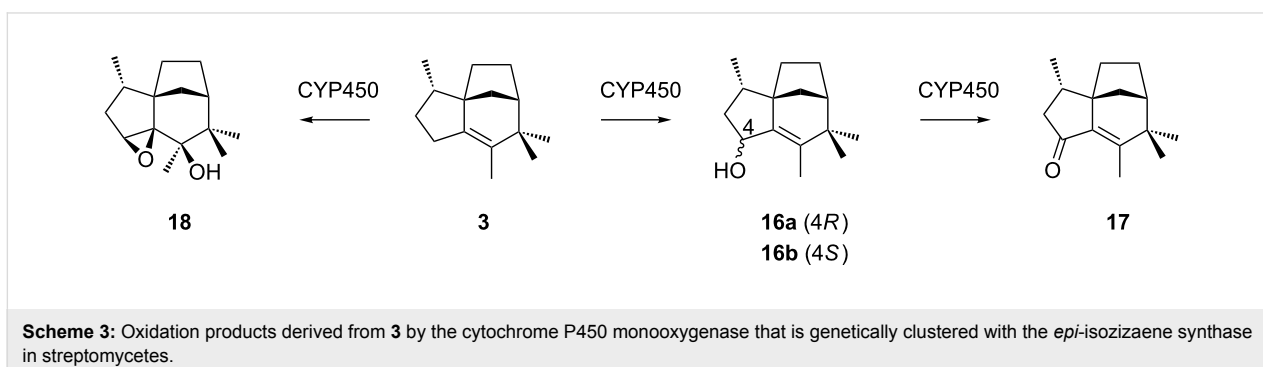


appears to be disordered in the crystal structure of 2-MIB synthase. The function of this domain is unknown, but it is conserved in most 2-MIB synthases and not present in any other terpene synthase [20,28].

epi-Isozizaene (**3**) is a tricyclic sesquiterpene precursor of the antibiotic albaflavenone (**17**) (Scheme 3) [29]. Furthermore, both enantiomers of the corresponding alcohols (*R*)- and (*S*)-albaflavenol (**16ab**) and the epoxide 4 β ,5 β -epoxy-2-*epi*-zizaan-6 β -ol (**18**) are known oxidation products that are all made by a cytochrome P450 monooxygenase [10,29] that is genetically clustered with the *epi*-isozizaene synthase for the cyclisation of FPP to **3** [30]. These enzymes are the most widespread sesquiterpene synthases in bacteria, and their coding genes are present in the genomes of more than 100 of the sequenced *Streptomyces* species [13]. Interestingly, *epi*-isozizaene synthases are only present in members of one clade (indicated as the green clade) in the phylogenetic analyses shown in Figure 1 and occur in almost all species of this clade with one exception (*S. scabiei* 87.22), suggesting an (unknown) ecological function of **3** or one of its oxidation products for strepto-

mycetes of this clade for their adaption to a specific ecological niche.

7-*epi*- α -Eudesmol (**4**) synthases are mostly present in a small group of species within the phylogenomic clade depicted in green in Figure 1, with some exceptions (*S. laurentii* ATCC 31255, *Streptomyces* sp PAMC 26508, *S. pratensis* ATCC 33331, *Streptomyces* sp_SM18 and *Streptomyces* sp. XZHG99, Figure 1). These exceptions may indicate horizontal gene transfer of the genes encoding for these enzymes. The sesquiterpene 7-*epi*- α -eudesmol synthase from *S. viridochromogenes* DSM 40736 has been chemically characterised in vivo by heterologous expression in *E. coli* BL21 and identification of the product in culture headspace extracts by GC–MS [31]. Compound **4** was also isolated from in vitro incubations of FPP with the recombinant enzyme and its optical rotation was shown to be opposite to the material from *Eucalyptus* [32], but the absolute configuration remains unknown. Production of this sesquiterpene by *S. viridochromogenes* DSM 40736 has not been observed [31], but **4** was occasionally reported from other streptomycetes encoding a 7-*epi*- α -eudesmol synthase [33,34].



epi-Cubenol (**5**) and caryolan-1-ol (**6**) synthases almost always occur together in one strain. We found only two examples of a strain that has a gene for caryolan-1-ol synthase but not for *epi*-cubenol synthase. These enzymes were found only in a sub-branch of closely related *Streptomyces* species from the blue clade and not present in members of any other phylogenomic group (Figure 1). Both enzymes have been identified and characterised in *S. griseus* NBRC 13350 [35,36] and their enzymatic mechanisms for the cyclisation of FPP have been investigated [35,37–39].

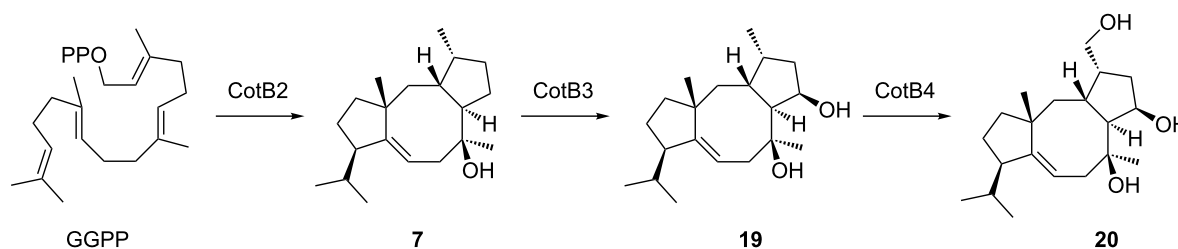
Cyclooctat-9-en-7-ol (**7**) and isoafrikanol (**8**) synthases are mainly characteristic for a group of very closely related species in the phylogenomic clade depicted in red in Figure 1, with two exceptions, *S. rubrolavendulae* MJM4426 and *S. collinus* Tü 365, members of the other two phylogenomic clades that also present a cyclooctat-9-en-7-ol synthase. Cyclooctat-9-en-7-ol synthase (CotB2) from *S. melanosporofaciens* was the first bacterial type I diterpene cyclase characterised [40] and its crystal structure was the first of a diterpene cyclase of bacterial origin reported [41]. Isoafrikanol synthases were first noticed in *S. violaceusniger* and *S. rapamycinicus* based on the presence of **8** in culture headspace extracts as a major sesquiterpene [34,42], followed by the biochemical characterisation of the recombinant enzyme from *Streptomyces malaysiensis* [43]. The diterpene **7** is a precursor to the lysophospholipase inhibitor cyclooctatin (**20**) formed by the action of two genetically clustered cytochrome P450 monooxygenases CotB3 and CotB4 (Scheme 4) [40,44], while no derivatives from **8** are currently known.

Pentalenene (**9**) and α -amorphene (**10**) synthases are the least abundant terpene synthases in *Streptomyces* species, each present in only 6 species (Figure 1). They are mostly present in members of the phylogenomic clade depicted in green in Figure 1, except for one case, *S. bingchenggensis* BCW1, but within the green clade their distribution is scattered and the number of identified genes for these enzymes is too low to draw conclusions on their occurrence in *Streptomyces*. The pental-

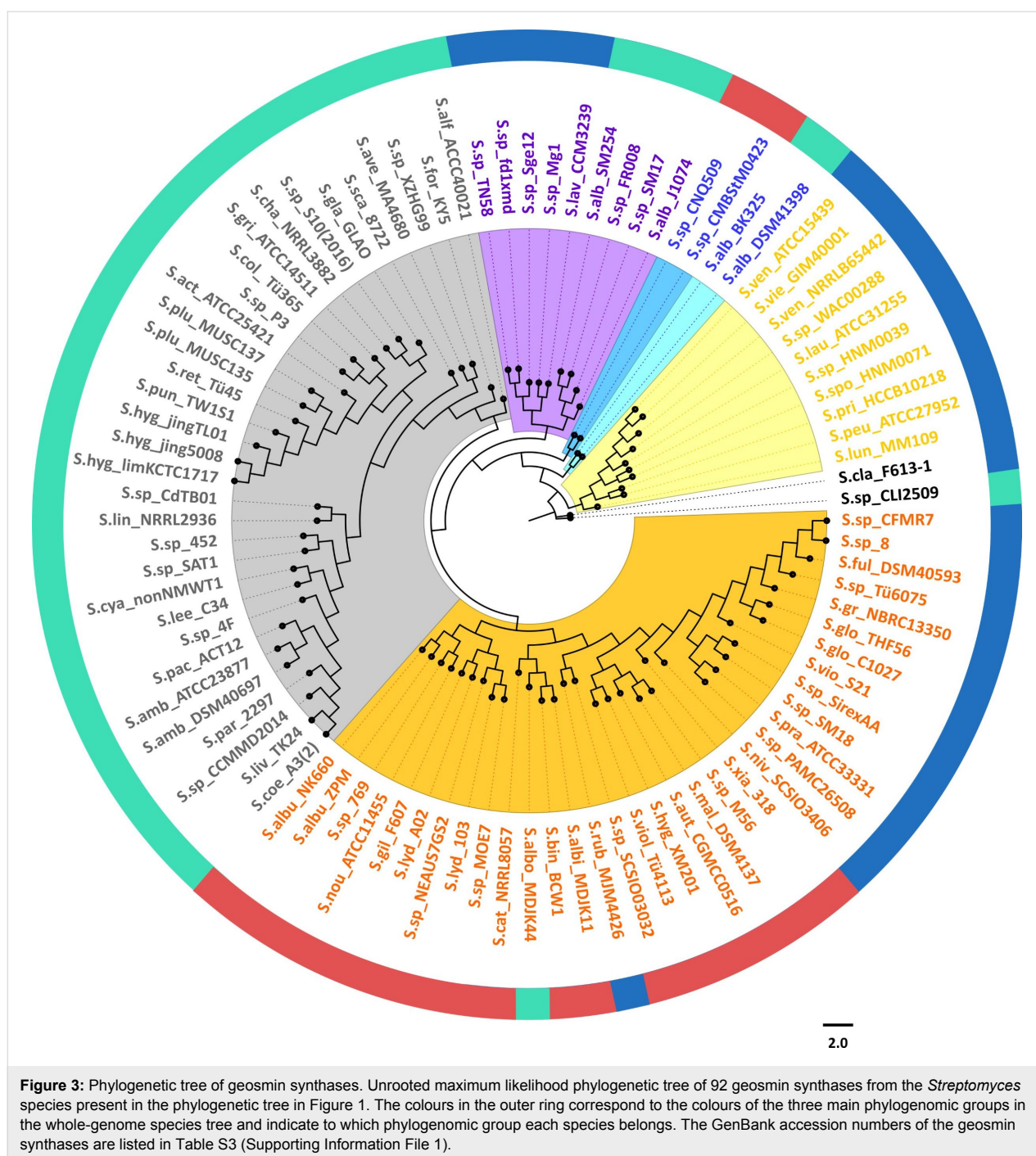
ene synthase from *S. exfoliatus* was the first characterised bacterial terpene synthase [45,46]. Its crystal structure was also the first reported for a bacterial terpene synthase [47]. Pentalenene synthase catalyses the cyclisation of FPP into pentalenene, which is the first step in the biosynthesis of the antibiotic pentalenolactone. This mechanism has been extensively studied and involves the initial ionisation of the substrate FPP and the formation of a humulyl cation as an intermediate in the biosynthesis of pentalenene [45,46,48,49], while the later steps of the cyclisation cascade were subject to revision based on the findings of quantum chemical calculations [50,51]. The α -amorphene synthase from *S. viridochromogenes* DSM 40736 was characterised by heterologous expression in *E. coli* BL21 [31] and by in vitro experiments with the purified enzyme [32].

Phylogenetic analysis of geosmin synthases

In order to determine if the geosmin synthases co-evolved with the *Streptomyces* species a phylogenetic tree was constructed with the geosmin synthases of all the species present in the full genome tree. As seen in Figure 3, the geosmin synthases separated into different clades. These clades do not fully correspond with specific phylogenomic groups from the genome-based analyses. Most of the geosmin synthases of the green and red phylogenetic clade in the whole genome-based tree of Figure 1 grouped together into one clade. The enzymes from the blue phylogenetic clade in the genome-based tree were the most scattered. All these results may point to the occurrence of horizontal gene transfer within the genus *Streptomyces*. However, if bacteria from other taxonomic groups such as myxobacteria and cyanobacteria and their geosmin synthases are included in a phylogenetic analysis, it can be seen that the geosmin synthase amino acid sequences from distantly related organisms clearly fall into distant clades [33]. Therefore, these results could also be interpreted as evidence for a rapid evolution of secondary metabolite genes to create new natural products with beneficial ecological functions for the producing organism. While many streptomycetes produce geosmin as a major metabolite of their bouquets of volatiles, the number and amounts of geosmin synthase side products associated with it



Scheme 4: Biosynthesis of cyclooctatin (**20**) from **7**.

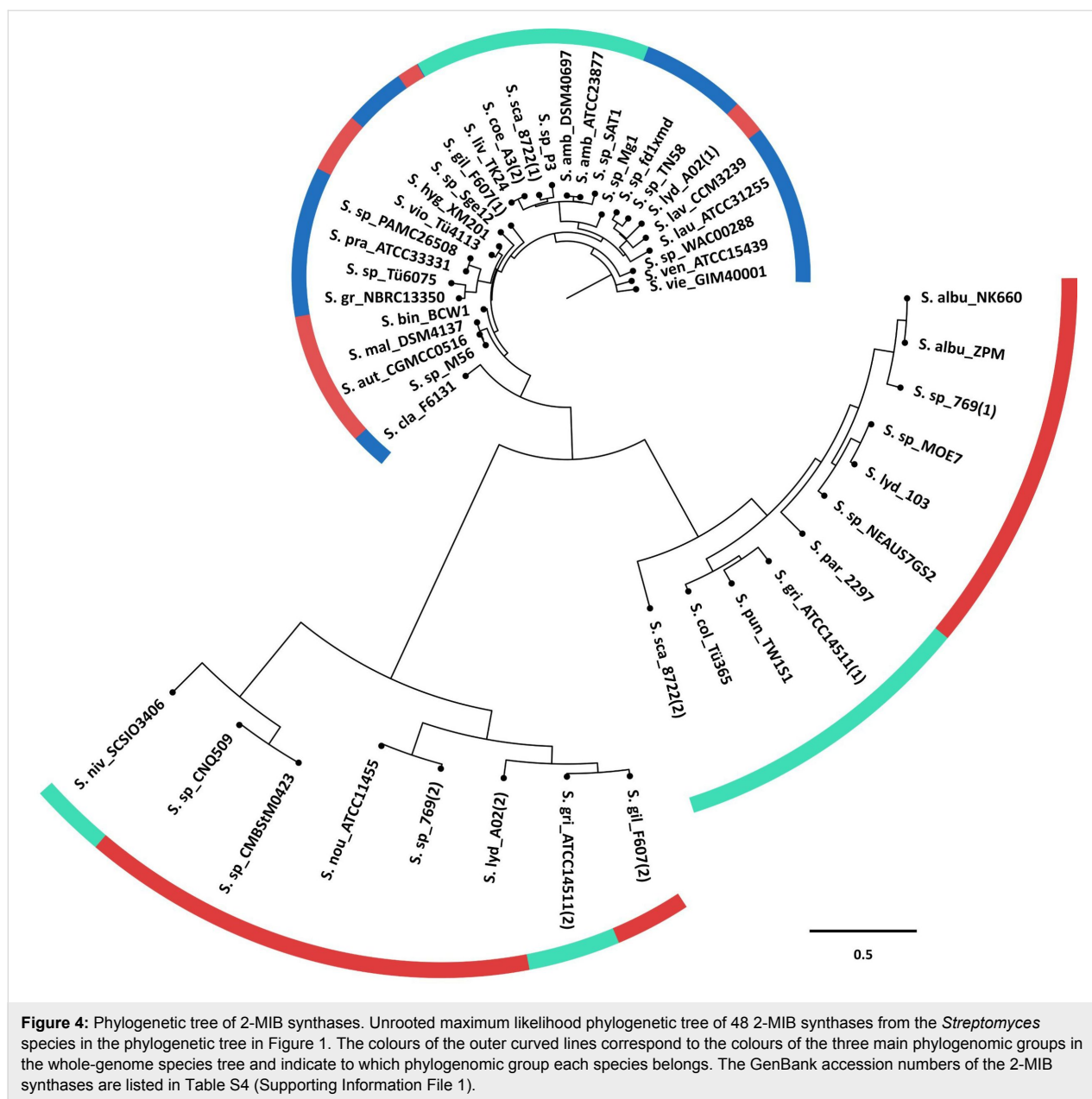


can vary [33,34], possibly as a result of an evolution of enzyme function.

Phylogenetic analysis of 2-MIB synthases

To gain insights into the evolution of the 2-MIB synthases a phylogenetic analysis of all the enzymes present in the *Streptomyces* species analysed in our study were performed (Figure 4). The phylogenetic tree of the 2-MIB synthases shows a clear separation into three clades (also indicated in Figure 1: group 1,

white circles, representing the major clade on the top of Figure 4; group 2, light grey circles, representing the clade on the bottom right; group 3, dark grey circles, representing the clade on the bottom left). Two of them are relatively distant from each other and even more so from the third clade where most species cluster together. This separation does not correspond with the separation observed based on the whole genome phylogenomic analyses. Only some of the enzymes that cluster together belong to species from the same phylogenomic group.

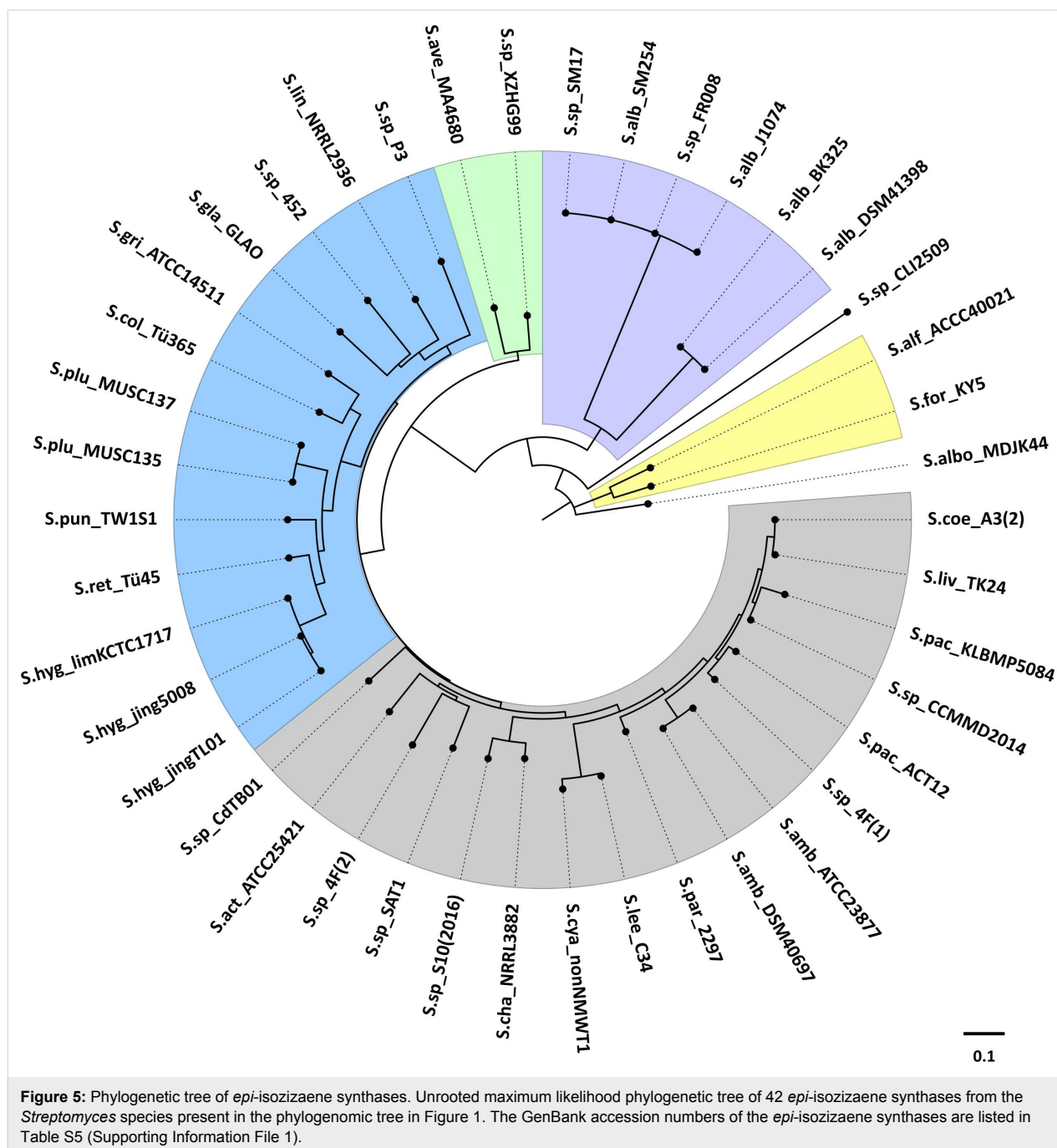


This indicates that the evolution of these enzymes does not correspond to the evolution of the *Streptomyces* species, and that a different force is driving how these enzymes evolved.

Phylogenetic analysis of *epi*-isozizaene synthases

epi-Isozizaene synthases are terpene synthases belonging only to a specific phylogenomic group of *Streptomyces* species. The phylogenetic analysis presented in Figure 5 shows two clades containing most of *epi*-isozizaene synthases and three other minor clades. Not all the enzymes are clustering in the same way as their containing species based on the whole-genome phylogenetic analyses. For example, *S. pactum* ACT12 *epi*-iso-

zizaene synthase clusters together with that of *Streptomyces* sp. 4F, while these two species were present in different branches in the full-genome-based phylogenomic tree. *Streptomyces* sp. 4F clustered together with *S. qaidamensis* S10(2016) and *S. chartreusis* NRRL 3882 in the phylogenomic tree. However, a second *epi*-isozizaene synthase present in *Streptomyces* sp. 4F clusters together with that of *Streptomyces* sp. SAT1, while these two species were located in separate clades of the phylogenomic tree. The occurrence of two genes for terpene synthases with putatively the same function may more strongly point to horizontal gene transfer events. Other cases include the *epi*-isozizaene synthases from *Streptomyces* sp. 452, *S. glaucescens* GLAO, *S. lincolnensis* NRRL 2936 and *Strepto-*



myces sp. P3 that group together with other enzymes, different to those belonging to species located in their same clade in the whole-genome phylogenomic analyses. This indicates also that some of these terpene synthases have evolved independently of the evolution of the *Streptomyces* species.

Phylogeny of terpene synthases does not correspond to species-level taxonomy

The comparison of the *Streptomyces* species whole genome-based phylogenetic tree and the three terpene synthase trees

shows that not all three comparing phylogenies are congruent. All *Streptomyces* strains included in this study carry at least one copy of *geoA*, with one exception. However, the topology of the geosmin synthase tree is not in harmony with the species tree and only some tips of the trees are conserved (Figure S3, Supporting Information File 1). The topological incongruence is even higher for *epi*-isozizaene and 2-MIB synthase trees (Figures S4 and S5, Supporting Information File 1). Tree reconstruction artefacts cannot explain these incongruences because all phylogenies obtained good statistical support. These data

support horizontal gene transfers of terpene synthase genes in *Streptomyces*, but could also point to secondary metabolite genes as being less conserved than housekeeping (primary metabolism) genes. Rapid evolution of secondary metabolism can lead to new natural products with advanced ecological functions in specific ecological niches. If horizontal gene transfer is indeed of high importance, one intriguing question would be why there are almost no *Streptomyces* strains with two or more genes for geosmin synthases and *epi*-isozizaene synthases. This could be explained by the rapid loss of genetic information after uptake of redundant information. It may also reflect the mechanism of integration of the incoming genetic information into the chromosome of the target organism by homologous recombination within identical or highly similar nucleotide sequences. In this study, we searched for the minimal number of events that are required to reconcile the terpene synthase trees with the species tree by performing NOTUNG analyses [52] (for detailed explanations cf. Supporting Information File 1, pp. 33–35). The analyses indicated that the discrepancies between the terpene synthase trees and the species tree, can be explained by horizontal gene transfer of the genes encoding for terpene synthases.

Conclusion

Overall, this study confirmed that *Streptomyces* species divide in three phylogenomic groups, based here on their whole genomes. Analysis of the distribution of the ten most abundant classes of terpene synthases in *Streptomyces* led to the surprising result that some terpene synthases are restricted to one phylogenomic group or even a subgroup which may point to a specific ecological function of the terpene for the respective group of organisms. The phylogenetic analyses of terpene synthases are not congruent with the phylogenomic analyses. Hence, the evolution of these enzymes does not correspond to the evolution of the *Streptomyces* species, possibly pointing to horizontal gene transfers as an important mechanism involved in the distribution of terpene synthase genes.

In this study, we focused on the distribution and evolution of terpenes synthases among *Streptomyces* species. It would be interesting in follow-up studies to assess the distribution and evolution of these genes among other bacteria, fungi, protists and plants. In addition, a deeper knowledge of the ecological function of terpenes in bacteria and in the interaction with their environment is highly desired.

Experimental

Streptomyces genomes selection

Genomes with whole sequences available in the NCBI database (thus not partial sequences) were included. Custom shell scripts (<https://github.com/kumarsaurabh20/>

[distribution_of_terpene_synthases](#)) were used to filter and download the nucleotide and protein sequences of all complete genomes including an annotation file in GFF format. The 93 selected sequences and their accession numbers are listed in Table S2 (Supporting Information File 1).

Construction of orthologous gene families

Sequence data of the proteins from the 93 *Streptomyces* species described above were collected. After removing sequences shorter than 50 amino acids, a total of 171,033 sequences were used to construct orthologous gene families using OrthoFinder – v 2.2.6 [53] applying the default setting (BLASTp e-value cutoff = $1e^{-5}$; MCL inflation I = 1.5). Using single-copy orthologues a species tree was inferred from unrooted gene trees that were constructed from all single copy genes using the STAG algorithm and the species tree was rooted using the STRIDE algorithm [54]. Both tools are available as core utilities in the OrthoFinder pipeline.

Phylogenetic analyses

Phylogenetic analyses on three different terpene synthases (geosmin synthases, 2-MIB synthases and *epi*-isozizaene synthases) were performed. Protein and nucleotide sequences were extracted from the *Streptomyces* genomes based on their distribution. Phylogenetic trees were generated using the protein dataset. Sequences were aligned with Mafft version 7.313 [55] using default parameters including *--auto* and *--inputorder*. All the alignments were trimmed for gaps and ambiguously aligned regions with BMGE – v 1.12 [56] using default parameters. For phylogenetic analyses, ProtTest – v 3.1.2 [57] was used to evaluate all evolutionary models under a AIC and BIC criterion. Maximum likelihood analyses were performed in RAxML – v 8.2.12 [58] under JTT+I+G (PROTGAMAMALG) model with rapid bootstrapping of 1000 replicates. GenBank accessions for each sequence are shown in Tables S3 to S5 in Supporting Information File 1.

Molecular evolution analysis

The coding DNA sequence (CDS) of the three terpene synthase genes (coding for geosmin synthases, 2-MIB synthases and *epi*-isozizaene synthases) in the 93 *Streptomyces* species were collected and aligned with Mafft version 7.313 using default parameters. Geneious – v 9.1 [59] was used to correct frame shifts and premature stop codons. Scripts published in [60] were used to generate codon-based alignments. We used HyPhy instance [61] to perform molecular evolution analysis. To test if positive selection occurred on a proportion of branches in the terpene synthase trees, the SLAC [62] model was used which is an improved version of the commonly used branch-site model. To test the hypothesis that individual sites have been subjected to episodic, positive or diversifying selection, site-specific model

FUBAR [63] was used. Additionally, aBSREL [64] model was used to infer nonsynonymous (dN) and synonymous (dS) substitution rates on a per-site basis for a given coding sequence alignment and corresponding phylogeny. The treefix-DTL (duplication-transfer-loss) software, version 1.0.2 [64], was applied to fix the topology of each terpene synthase tree under default settings with an alpha value of 0.05 for the paired-site test and the model closest to PROTGAMMALGF available via treefix-DTL (PROTGAMMAJTTF) as RAxML substitution model. To reconstruct the types and numbers of the evolutionary events that explain the discrepancies (if any) between the final topologies, NOTUNG version 2.9 [52] was run under default settings (modified weight parameters edge weight = 0.9; duplication weight = 2.0; transfer weight = 3.0; losses weight = 1.0) except for the permission of horizontal transfers and the use of a DTL cost matrix of 2-3-1, corresponding to default costs used by treefix-DTL.

Supporting Information

Supporting Information File 1

Additional figures and tables.

[<https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-15-115-S1.pdf>]

Acknowledgements

This work was supported by NWO ALWOP.178 grant. This is publication 6728 of the NIOO-KNAW.

ORCID® IDs

Lara Martín-Sánchez - <https://orcid.org/0000-0001-5514-4712>

Kumar Saurabh Singh - <https://orcid.org/0000-0001-8352-5897>

Gilles P. van Wezel - <https://orcid.org/0000-0003-0341-1561>

Jeroen S. Dickschat - <https://orcid.org/0000-0002-0102-0631>

References

- Barka, E. A.; Vatsa, P.; Sanchez, L.; Gaveau-Vaillant, N.; Jacquard, C.; Klenk, H.-P.; Clément, C.; Ouhdouch, Y.; van Wezel, G. P. *Microbiol. Mol. Biol. Rev.* **2016**, *80*, No. 1. doi:10.1128/mmlr.00019-15
- Bentley, S. D.; Chater, K. F.; Cerdeño-Tárraga, A.-M.; Challis, G. L.; Thomson, N. R.; James, K. D.; Harris, D. E.; Quail, M. A.; Kieser, H.; Harper, D.; Bateman, A.; Brown, S.; Chandra, G.; Chen, C. W.; Collins, M.; Cronin, A.; Fraser, A.; Goble, A.; Hidalgo, J.; Hornsby, T.; Howarth, S.; Huang, C.-H.; Kieser, T.; Larke, L.; Murphy, L.; Oliver, K.; O'Neill, S.; Rabinowitsch, E.; Rajandream, M.-A.; Rutherford, K.; Rutter, S.; Seeger, K.; Saunders, D.; Sharp, S.; Squares, R.; Squares, S.; Taylor, K.; Warren, T.; Wietzorrek, A.; Woodward, J.; Barrell, B. G.; Parkhill, J.; Hopwood, D. A. *Nature* **2002**, *417*, 141–147. doi:10.1038/417141a
- Bérdy, J. *J. Antibiot.* **2005**, *58*, 1–26. doi:10.1038/ja.2005.1
- Nett, M.; Ikeda, H.; Moore, B. S. *Nat. Prod. Rep.* **2009**, *26*, 1362–1384. doi:10.1039/b817069j
- Medema, M. H.; Kottmann, R.; Yilmaz, P.; Cummings, M.; Biggins, J. B.; Blin, K.; de Bruijn, I.; Chooi, Y. H.; Claessen, J.; Coates, R. C.; Cruz-Morales, P.; Duddela, S.; Dusterhus, S.; Edwards, D. J.; Fewer, D. P.; Garg, N.; Geiger, C.; Gomez-Escribano, J. P.; Greule, A.; Hadjithomas, M.; Haines, A. S.; Helfrich, E. J. N.; Hillwig, M. L.; Ishida, K.; Jones, A. C.; Jones, C. S.; Jungmann, K.; Kegler, C.; Kim, H. U.; Kötter, P.; Krug, D.; Masschelein, J.; Melnik, A. V.; Mantovani, S. M.; Monroe, E. A.; Moore, M.; Moss, N.; Nützmann, H.-W.; Pan, G.; Pati, A.; Petras, D.; Reen, F. J.; Rosconi, F.; Rui, Z.; Tian, Z.; Tobias, N. J.; Tsunematsu, Y.; Wiemann, P.; Wyckoff, E.; Yan, X.; Yim, G.; Yu, F.; Xie, Y.; Aigle, B.; Apel, A. K.; Balibar, C. J.; Balskus, E. P.; Barona-Gómez, F.; Bechthold, A.; Bode, H. B.; Borriss, R.; Brady, S. F.; Brakhage, A. A.; Caffrey, P.; Cheng, Y.-Q.; Clardy, J.; Cox, R. J.; De Mot, R.; Donadio, S.; Donia, M. S.; van der Donk, W. A.; Dorrestein, P. C.; Doyle, S.; Driessen, A. J. M.; Ehling-Schulz, M.; Entian, K.-D.; Fischbach, M. A.; Gerwick, L.; Gerwick, W. H.; Gross, H.; Gust, B.; Hertweck, C.; Höfte, M.; Jensen, S. E.; Ju, J.; Katz, L.; Kaysser, L.; Klassen, J. L.; Keller, N. P.; Kormanec, J.; Kuipers, O. P.; Kuzuyama, T.; Kypides, N. C.; Kwon, H.-J.; Lautru, S.; Lavigne, R.; Lee, C. Y.; Linqun, B.; Liu, X.; Liu, W.; Luzhetskyy, A.; Mahmud, T.; Mast, Y.; Méndez, C.; Metsä-Ketelä, M.; Micklefield, J.; Mitchell, D. A.; Moore, B. S.; Moreira, L. M.; Müller, R.; Neilan, B. A.; Nett, M.; Nielsen, J.; O'Gara, F.; Oikawa, H.; Osbourn, A.; Osburne, M. S.; Ostash, B.; Payne, S. M.; Pernodet, J.-L.; Petricek, M.; Piel, J.; Ploux, O.; Raaijmakers, J. M.; Salas, J. A.; Schmitt, E. K.; Scott, B.; Seipke, R. F.; Shen, B.; Sherman, D. H.; Sivonen, K.; Smanski, M. J.; Sosio, M.; Stegmann, E.; Süßmuth, R. D.; Tahlan, K.; Thomas, C. M.; Tang, Y.; Truman, A. W.; Viaud, M.; Walton, J. D.; Walsh, C. T.; Weber, T.; van Wezel, G. P.; Wilkinson, B.; Willey, J. M.; Wohlleben, W.; Wright, G. D.; Ziemert, N.; Zhang, C.; Zotchev, S. B.; Breitling, R.; Takano, E.; Glöckner, F. O. *Nat. Chem. Biol.* **2015**, *11*, 625–631. doi:10.1038/nchembio.1890
- Chen, X.; Köllner, T. G.; Jia, Q.; Norris, A.; Santhanam, B.; Rabe, P.; Dickschat, J. S.; Shauly, G.; Gershenzon, J.; Chen, F. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 12132–12137. doi:10.1073/pnas.1610379113
- Song, C.; Mazzola, M.; Cheng, X.; Oetjen, J.; Alexandrov, T.; Dorrestein, P.; Watrous, J.; van der Voort, M.; Raaijmakers, J. M. *Sci. Rep.* **2015**, *5*, 12837. doi:10.1038/srep12837
- Yamada, Y.; Cane, D. E.; Ikeda, H. Diversity and Analysis of Bacterial Terpene Synthases. In *Natural Product Biosynthesis by Microorganisms and Plant, Part A*; Hopwood, D. A., Ed.; Elsevier: Amsterdam, Netherlands, 2012; Vol. 515, pp 123–162. doi:10.1016/b978-0-12-394290-6.00007-0
- Yamada, Y.; Kuzuyama, T.; Komatsu, M.; Shin-ya, K.; Omura, S.; Cane, D. E.; Ikeda, H. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 857–862. doi:10.1073/pnas.1422108112
- Takamatsu, S.; Lin, X.; Nara, A.; Komatsu, M.; Cane, D. E.; Ikeda, H. *Microb. Biotechnol.* **2011**, *4*, 184–191. doi:10.1111/j.1751-7915.2010.00209.x
- Holl, D. Biosynthesis and Biological Functions of Terpenoids in Plants. In *Biotechnology of Isoprenoids*; Schrader, J.; Bohlmann, J., Eds.; Advances in Biochemical Engineering/Biotechnology, Vol. 148; Springer: Cham, 2015; pp 63–106. doi:10.1007/10_2014_295
- Schulz-Bohm, K.; Martín-Sánchez, L.; Garbeva, P. *Front. Microbiol.* **2017**, *8*, No. 2484. doi:10.3389/fmicb.2017.02484
- Dickschat, J. S. *Nat. Prod. Rep.* **2016**, *33*, 87–110. doi:10.1039/c5np00102a
- Gerber, N. N.; Lechevalier, H. A. *Appl. Microbiol.* **1965**, *13*, 935–938.

15. Medsker, L. L.; Jenkins, D.; Thomas, J. F.; Koch, C. *Environ. Sci. Technol.* **1969**, *3*, 476–477. doi:10.1021/es60028a008
16. McDonald, B. R.; Currie, C. R. *mBio* **2017**, *8*, e00644-17. doi:10.1128/mbio.00644-17
17. Qin, S.; Feng, W.-W.; Wang, T.-T.; Ding, P.; Xing, K.; Jiang, J.-H. *Plant Soil* **2017**, *416*, 117–132. doi:10.1007/s11104-017-3192-2
18. Jiang, J.; He, X.; Cane, D. E. *Nat. Chem. Biol.* **2007**, *3*, 711–715. doi:10.1038/nchembio.2007.29
19. Baer, P.; Rabe, P.; Fischer, K.; Citron, C. A.; Klapschinski, T. A.; Groll, M.; Dickschat, J. S. *Angew. Chem., Int. Ed.* **2014**, *53*, 7652–7656. doi:10.1002/anie.201403648
20. Christianson, D. W. *Chem. Rev.* **2017**, *117*, 11570–11648. doi:10.1021/acs.chemrev.7b00287
21. Seemann, M.; Zhai, G.; de Kraker, J.-W.; Paschall, C. M.; Christianson, D. W.; Cane, D. E. *J. Am. Chem. Soc.* **2002**, *124*, 7681–7689. doi:10.1021/ja026058q
22. Dickschat, J. S.; Bode, H. B.; Mahmud, T.; Müller, R.; Schulz, S. *J. Org. Chem.* **2005**, *70*, 5174–5182. doi:10.1021/jo050449g
23. Jiang, J.; Cane, D. E. *J. Am. Chem. Soc.* **2008**, *130*, 428–429. doi:10.1021/ja077792i
24. Nawrath, T.; Dickschat, J. S.; Müller, R.; Jiang, J.; Cane, D. E.; Schulz, S. *J. Am. Chem. Soc.* **2008**, *130*, 430–431. doi:10.1021/ja077790y
25. Dickschat, J. S.; Nawrath, T.; Thiel, V.; Kunze, B.; Müller, R.; Schulz, S. *Angew. Chem., Int. Ed.* **2007**, *46*, 8287–8290. doi:10.1002/anie.200702496
26. Wang, C.-M.; Cane, D. E. *J. Am. Chem. Soc.* **2008**, *130*, 8908–8909. doi:10.1021/ja803639g
27. Komatsu, M.; Tsuda, M.; Omura, S.; Oikawa, H.; Ikeda, H. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 7422–7427. doi:10.1073/pnas.0802312105
28. Köksal, M.; Chou, W. K. W.; Cane, D. E.; Christianson, D. W. *Biochemistry* **2012**, *51*, 3011–3020. doi:10.1021/bi201827a
29. Zhao, B.; Lin, X.; Lei, L.; Lamb, D. C.; Kelly, S. L.; Waterman, M. R.; Cane, D. E. *J. Biol. Chem.* **2008**, *283*, 8183–8189. doi:10.1074/jbc.m710421200
30. Lin, X.; Hopson, R.; Cane, D. E. *J. Am. Chem. Soc.* **2006**, *128*, 6022–6023. doi:10.1021/ja061292s
31. Rabe, P.; Dickschat, J. S. *Angew. Chem., Int. Ed.* **2013**, *52*, 1810–1812. doi:10.1002/anie.201209103
32. Rabe, P.; Schmitz, T.; Dickschat, J. S. *Beilstein J. Org. Chem.* **2016**, *12*, 1839–1850. doi:10.3762/bjoc.12.173
33. Citron, C. A.; Gleitzmann, J.; Laurenzano, G.; Pukall, R.; Dickschat, J. S. *ChemBioChem* **2012**, *13*, 202–214. doi:10.1002/cbic.201100641
34. Citron, C. A.; Barra, L.; Wink, J.; Dickschat, J. S. *Org. Biomol. Chem.* **2015**, *13*, 2673–2683. doi:10.1039/c4ob02609h
35. Nakano, C.; Horinouchi, S.; Ohnishi, Y. *J. Biol. Chem.* **2011**, *286*, 27980–27987. doi:10.1074/jbc.m111.265652
36. Nakano, C.; Tezuka, T.; Horinouchi, S.; Ohnishi, Y. *J. Antibiot.* **2012**, *65*, 551–558. doi:10.1038/ja.2012.68
37. Cane, D. E.; Tandon, M. *Tetrahedron Lett.* **1994**, *35*, 5355–5358. doi:10.1016/s0040-4039(00)73498-8
38. Cane, D. E.; Tandon, M. *J. Am. Chem. Soc.* **1995**, *117*, 5602–5603. doi:10.1021/ja00125a029
39. Cane, D. E.; Tandon, M.; Prabhakaran, P. C. *J. Am. Chem. Soc.* **1993**, *115*, 8103–8106. doi:10.1021/ja00071a023
40. Kim, S.-Y.; Zhao, P.; Igarashi, M.; Sawa, R.; Tomita, T.; Nishiyama, M.; Kuzuyama, T. *Chem. Biol.* **2009**, *16*, 736–743. doi:10.1016/j.chembiol.2009.06.007
41. Janke, R.; Görner, C.; Hirte, M.; Brück, T.; Loll, B. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2014**, *70*, 1528–1537. doi:10.1107/s1399004714005513
42. Riclea, R.; Citron, C. A.; Rinkel, J.; Dickschat, J. S. *Chem. Commun.* **2014**, *50*, 4228–4230. doi:10.1039/c4cc00177j
43. Rabe, P.; Samborskyy, M.; Leadlay, P. F.; Dickschat, J. S. *Org. Biomol. Chem.* **2017**, *15*, 2353–2358. doi:10.1039/c7ob00234c
44. Aoyagi, T.; Aoyama, T.; Kojima, F.; Hattori, S.; Honma, Y.; Hamada, M.; Takeuchi, T. *J. Antibiot.* **1992**, *45*, 1587–1591. doi:10.7164/antibiotics.45.1587
45. Cane, D. E.; Abell, C.; Tillman, A. M. *Bioorg. Chem.* **1984**, *12*, 312–328. doi:10.1016/0045-2068(84)90013-0
46. Cane, D. E.; Tillman, A. M. *J. Am. Chem. Soc.* **1983**, *105*, 122–124. doi:10.1021/ja00339a026
47. Lesburg, C. A.; Zhai, G.; Cane, D. E.; Christianson, D. W. *Science* **1997**, *277*, 1820–1824. doi:10.1126/science.277.5333.1820
48. Cane, D. E.; Oliver, J. S.; Harrison, P. H. M.; Abell, C.; Hubbard, B. R.; Kane, C. T.; Lattman, R. *J. Am. Chem. Soc.* **1990**, *112*, 4513–4524. doi:10.1021/ja00167a059
49. Cane, D. E.; Sohng, J.-K.; Lamberson, C. R.; Rudnicki, S. M.; Wu, Z.; Lloyd, M. D.; Oliver, J. S.; Hubbard, B. R. *Biochemistry* **1994**, *33*, 5846–5857. doi:10.1021/bi00185a024
50. Gutta, P.; Tantillo, D. J. *J. Am. Chem. Soc.* **2006**, *128*, 6172–6179. doi:10.1021/ja058031n
51. Zu, L.; Xu, M.; Lodewyk, M. W.; Cane, D. E.; Peters, R. J.; Tantillo, D. J. *J. Am. Chem. Soc.* **2012**, *134*, 11369–11371. doi:10.1021/ja3043245
52. Darby, C. A.; Stolzer, M.; Ropp, P. J.; Barker, D.; Durand, D. *Bioinformatics* **2017**, *33*, 640–649. doi:10.1093/bioinformatics/btw686
53. Emms, D. M.; Kelly, S. *Genome Biol.* **2015**, *16*, 157. doi:10.1186/s13059-015-0721-2
54. Emms, D. M.; Kelly, S. *Mol. Biol. Evol.* **2017**, *34*, 3267–3278. doi:10.1093/molbev/msx259
55. Katoh, K.; Standley, D. M. *Mol. Biol. Evol.* **2013**, *30*, 772–780. doi:10.1093/molbev/mst010
56. Criscuolo, A.; Gribaldo, S. *BMC Evol. Biol.* **2010**, *10*, 210. doi:10.1186/1471-2148-10-210
57. Darriba, D.; Taboada, G. L.; Doallo, R.; Posada, D. *Bioinformatics* **2011**, *27*, 1164–1165. doi:10.1093/bioinformatics/btr088
58. Stamatakis, A. *Bioinformatics* **2014**, *30*, 1312–1313. doi:10.1093/bioinformatics/btu033
59. Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C.; Thierer, T.; Ashton, B.; Meintjes, P.; Drummond, A. *Bioinformatics* **2012**, *28*, 1647–1649. doi:10.1093/bioinformatics/bts199
60. Sydykova, D. K.; Jack, B. R.; Spielman, S. J.; Wilke, C. O. *F1000Research* **2018**, *6*, 1845. doi:10.12688/f1000research.12874.2
61. Pond, S. L. K.; Frost, S. D. W.; Muse, S. V. *Bioinformatics* **2005**, *21*, 676–679. doi:10.1093/bioinformatics/bti079
62. Smith, M. D.; Wertheim, J. O.; Weaver, S.; Murrell, B.; Scheffler, K.; Kosakovsky Pond, S. L. *Mol. Biol. Evol.* **2015**, *32*, 1342–1353. doi:10.1093/molbev/msv022
63. Murrell, B.; Moola, S.; Mabona, A.; Weighill, T.; Sheward, D.; Kosakovsky Pond, S. L.; Scheffler, K. *Mol. Biol. Evol.* **2013**, *30*, 1196–1205. doi:10.1093/molbev/mst030
64. Bansal, M. S.; Wu, Y.-C.; Alm, E. J.; Kellis, M. *Bioinformatics* **2015**, *31*, 1211–1218. doi:10.1093/bioinformatics/btu806

License and Terms

This is an Open Access article under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>). Please note that the reuse, redistribution and reproduction in particular requires that the authors and source are credited.

The license is subject to the *Beilstein Journal of Organic Chemistry* terms and conditions: (<https://www.beilstein-journals.org/bjoc>)

The definitive version of this article is the electronic one which can be found at:
[doi:10.3762/bjoc.15.115](https://doi.org/10.3762/bjoc.15.115)