The evolutionary history and modern diversity of triterpenoid cyclases

Hanon Solomon McShea[1,2], Robb A. Viens[1], Babatunde O. Olagunju[3], José-Luis Giner[3], and Paula V. Welander[1*]

[1]*Department of Earth System Science, Stanford University, Stanford, CA, USA*

[2]*Department of Microbiology and Immunology, University of California, San Francisco, San Francisco, CA, USA*

[3]*Department of Chemistry, SUNY-ESF, Syracuse, NY, USA*

*Correspondence: Paula Welander welander@stanford.edu

Abstract

Cyclic terpenoids are a class of lipid compounds containing immense structural and functional diversity, with many cyclic triterpenoids acting as regulators of the physical properties and spatial organization of lipid membranes. Cyclic terpenoids are also readily preserved as terpane fossils, such as steranes and hopanes, forming a rich record of the evolution of life on Earth. Formation of the multiple ring structure of all cyclic terpenoids is catalyzed by terpenoid cyclase enzymes, among which are whole clades of proteins – many from environmental metagenomes and uncultured organisms – whose substrates and products are completely unknown. We investigate the function of these divergent cyclases through biochemical assays, and the evolutionary processes that produced them by testing and applying a variety of evolutionary models. We find deep divergence between the diterpenoid cyclases and triterpenoid cyclases, with other clades branching between the two, rooting the triterpenoid cyclase subtree, between squalene-hopene cyclases and sterol cyclases. Through a simple test of evolutionary rate shifts, we find an elevated evolutionary rate in the enzyme active site on the squalene-hopene cyclase stem, potentially indicative of positive selection. Finally, by testing the activity of divergent cyclases for a variety of substrates, we find a group of early-branching sterol cyclases from bacteria that synthesize arborinols, two of which produce the molecular precursor to a Permian "orphan biomarker." Together, our data present an evolutionary framework for triterpenoid cyclases that can inform both the biochemical potential of these proteins and their products' occurrence in the geological record.

Introduction

In 1971, "geohopanes" were discovered in vast quantities in organic-rich sediments, ancient rocks, and fuel reservoirs (Ourisson & Albrecht, 1992). In the following years, hopanoid molecules from diverse bacteria were identified as the biological source of these molecular fossils (Förster et al., 1973; Ourisson & Rohmer, 1992). The finding precipitated a new era of research on the taxonomic distribution and physiological function of cyclic triterpenoids (hopanoids, sterols, and similar molecules shown in Figure 1A) and the triterpenoid cyclase enzymes that synthesize them, motivated by questions of what the molecular fossil record could reveal about the evolution of life on Earth, and of Earth itself (e.g. Fischer, 2008; Ourisson & Nakatani, 1994; Runnegar, 1991; Schaeffer et al., 1994). This body of research has uncovered diverse triterpane fossils, as well as an increasingly deep understanding of cyclic terpenoid structural diversity (Hayashi et al., 2007; Kontnik et al., 2008; Moosmann et al., 2020; Rudolf et al., 2021), taxonomic distribution (Mayer et al., 2021; Pearson et al., 2003; Ricci et al., 2014; Takishita et al., 2017; Villanueva et al., 2014, p. 201; Wei et al., 2016), biosynthesis (Banta et al., 2015; Brown et al., 2023; Lee et al., 2018, 2023; Pan et al., 2015; Pollier et al., 2019; Welander et al., 2010), and function (Brenac et al., 2019; Flesch & Rohmer, 1987; Gudde et al., 2019; Rivas-Marin et al., 2019; Schmerk et al., 2015; Welander et al., 2009, 2012; Welander &

Summons, 2012; Zhai et al., 2024). Unsurprisingly, the catalogues of triterpane fossils and modern triterpenoids are not isomorphic – there are many "orphan biomarkers" in the geological record with no known biological source (Brocks & Pearson, 2005), and conversely, there are many organisms and environments that have genes for triterpenoid cyclases, but no known biosynthetic products (Pearson et al., 2007). To reconcile such orphan biomarkers and mystery cyclases, we analyzed the modern diversity and evolutionary history of the terpenoid cyclase protein superfamily, of which triterpenoid cyclases are one protein family.

Terpenoid cyclases rearrange carbon-carbon bonds to form polycyclic molecules through an exquisitely precise carbocation cascade (Abe et al., 2014; Eschenmoser & Arigoni, 2005;Hoshino & Sato, 2002). We refer here specifically to the Type II terpenoid cyclase protein family, which protonates substrates with a general acid, and is distinct from the Type I terpenoid cyclase family, which adopts a different fold and employs a different, metal-dependent mechanism to cyclize C_{10}, C_{15}, and C_{20} substrates (Christianson, 2017). Type II terpenoid cyclases typically have a dumbbell shape composed of two homologous structural domains known as the β and γ domains. The enzyme active site is between the two domains, accessible by a substrate channel through the γ domain, while the β domain has a catalytic aspartic acid, and a water tunnel needed to replenish it. Some Type II cyclases have additional domains or lack the γ domain. All domains share the alpha-alpha toroid fold, which is a torus of 12 alpha helices arranged in two concentric rings of 6 (Figure 1). As described by the Structural Classification of Proteins – extended 2.08 (Chandonia et al., 2022; Fox et al., 2014), terpenoid cyclases and protein prenyltransferases make up one of the six superfamilies in the alpha-alpha toroid fold, the others being glycosidases and polysaccharide lyases (Syrén et al., 2016), an enzyme involved in the synthesis of lanthipeptide antibiotics (Zhang et al., 2012), and eukaryotic α2-macroglobulins and complement proteins (Cao et al., 2010).

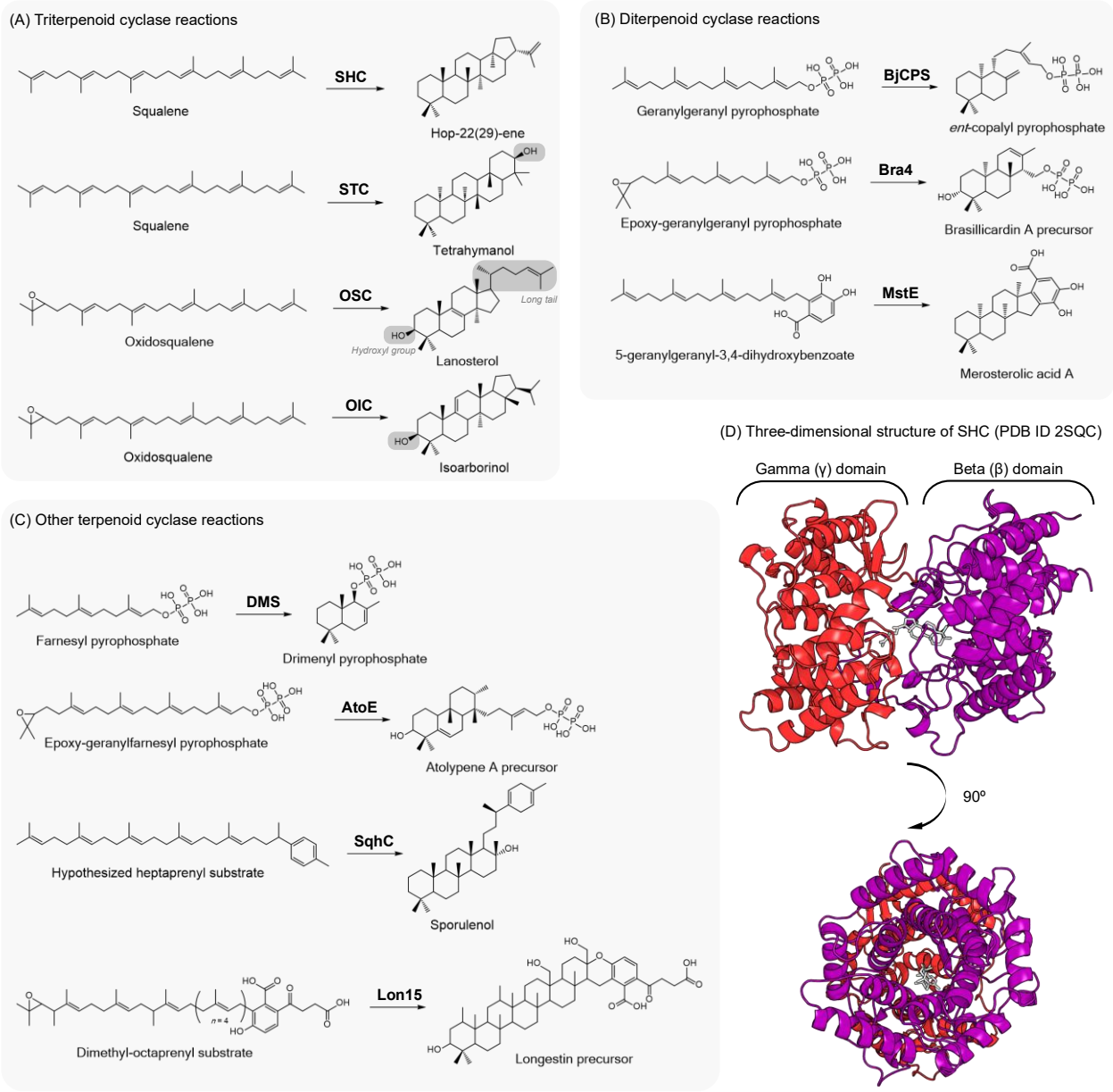

Figure 1: Reactions performed by Type II terpenoid cyclases, including on (A) C_{30} substrates in the case of triterpenoid cyclases (with annotation showing important chemical differences among products), (B) C_{20} substrates in the case of diterpenoid cyclases, and (C) C_{15}, C_{25}, C_{35}, and C_{40} substrates for other Type II terpenoid cyclases. Other reactions exist for some substrates; for example, in addition to lanosterol cyclases, there are oxidosqualene cyclases that synthesize parkeol, cycloartenol, or amyrins. (D) The crystal structure of squalene-hopene cyclase from *Alicyclobacillus acidocaldarius* (Wendt et al., 1999). SHC: squalene-hopene cyclase; STC: tetrahymanol cyclase; OSC: oxidosqualene cyclase; OIC: oxidosqualene-isoarborinol cyclase; BjCPS: *Bradyrhizobium japonicum* copalyl diphosphate synthase; Bra4: brasilicardin cyclase; MstE: merosterolic acid synthase; DMS: drimenol synthase; AtoE: atolypene synthase; SqhC: sporulenol synthase; Lon15: terpene cyclase (longestin).

Type II terpenoid cyclases include triterpenoid (C_{30} substrate) and diterpenoid (C_{20} substrate) cyclases, which share a catalytic mechanism and βγ domain architecture (Christianson, 2017). Triterpenoid cyclases include oxidosqualene cyclases, which cyclize an epoxidized

substrate to form tetracyclic sterols such as lanosterol, parkeol, cycloartenol, arborinols, and amyrins (R. Xu et al., 2004). These lipid alcohols are often modified to form functional membrane components such as cholesterol. Squalene-hopene cyclases and tetrahymanol cyclases cyclize squalene to form pentacyclic hopenes and tetrahymanol, respectively. Unmodified tetrahymanol functions as a membrane lipid (Conner et al., 1971), and hopanoids most often undergo polyfunctionalization at the "tail" alkene to become amphipathic membrane components as well (Belin et al., 2018; Bradley et al., 2010). Diterpenoid cyclases cyclize geranylgeranyl pyrophosphate (C_{20}) to form bicyclic copalyl (Ikeda et al., 2007; Morrone et al., 2009; Smanski et al., 2011), halimadienyl (Nakano et al., 2005), terpentedienyl (Dairi et al., 2001; Stowell et al., 2022), and kolavenyl (Nakano et al., 2015) diphosphates, among other molecules. Other diterpenoid cyclases work on epoxidized geranylgeranyl pyrophosphate, forming tricyclic molecules such as brasilicardin (Hayashi et al., 2008), phenalinolactone (Dürr et al., 2006), and tiancilactone (Dong et al., 2018) precursors. As shown in Figure 1, the protein superfamily includes the above and also C_{15} (Pan et al., 2022; Vo et al., 2022), C_{25} (Kim et al., 2019), C_{35} (Kontnik et al., 2008; Sato, Hoshino, et al., 2011; Sato, Yoshida, et al., 2011), and C_{40} (Hayashi et al., 2007; Ozaki et al., 2018) cyclases, which work on a variety of alkenyl, epoxidized, and otherwise functionalized isoprenoid substrates.

Products of triterpenoid cyclases often undergo modification by other enzymes and go on to regulate membrane homeostasis and dynamics, among other functions (Bloch, 1983). The best-studied example is cholesterol in the animal cell membrane, where it buffers membrane flexibility against temperature changes (Crockett, 1998; Mouritsen & Zuckermann, 2004) and organizes the membrane into liquid-ordered microdomains (Levental et al., 2020; Xu & London, 2000). There is evidence for similar function for hopanoids in bacteria (López & Kolter, 2010; Sáenz, 2010; Sáenz et al., 2012, 2015). Cyclic diterpenoids, on the other hand, are commonly modified with solubilizing functional groups after cyclization, and function in cell-cell and interspecific communication. Meanwhile, diterpenoids cyclized by Type I cyclasesare often insoluble and are involved in other functions such as plant wound healing in the form of resins. Other well-studied diterpenoids include the gibberellin hormones made by plants and their symbionts to regulate tissue growth (Nett et al., 2022). Cyclic diterpenoids are also involved in interspecies conflict – between microbes in the form of antibiotics such as terpentecin (Tamamura et al., 1985), platencin (Jayasuriya et al., 2007; J. Wang et al., 2007), platensimycin (J. Wang et al., 2006), and phenalinolactone (Gebhardt et al., 2011), antifungals such as viguiepinol and oxaloterpins (Bi & Yu, 2016), and between humans and their pathogens in the case of tuberculosinol from *Mycobacterium tuberculosis* (Mann et al., 2009) and brasilicardin A from *Nocardia brasiliensis* (Shigemori et al., 1998; Usui et al., 2006).

The evolutionary processes that generated the structural and functional diversity of terpenoids and of triterpenoids in particular have long been a subject of great scientific interest (Bloch, 1983; Ourisson, 1989; Ourisson & Nakatani, 1994). Fischer and Pearson (2007) presented several alternative models for relationships among triterpenoid cyclases under

124 parsimonious evolution of three functional traits, which they posit evolve independently and
125 slowly, if not irreversibly: substrate (squalene vs. oxidosqualene), reaction favorability (number
126 of anti-Markovnikov carbocations propagated during the reaction), and product stereochemistry
127 (whether the first three rings are chair-chair-chair vs. chair-boat-chair). In this study, we adopt
128 their rigorous evolutionary approach and update it with four new sources of information, in
129 pursuit of the same biological question: in what order, and through what processes, did the
130 diversity of polycyclic membrane regulators arise?

131 We first apply a probabilistic method (maximum likelihood) of estimating relationships
132 among cyclases, using an explicit model of protein sequence evolution. This allows us to relax
133 Fischer and Pearson's (2007) assumptions about the independence of the traits and the (slow)
134 speed at which they evolve. Therefore, we can observe the distribution of the selected cyclase
135 traits across a tree generated under unrelated assumptions (amino acid exchangeabilities and rate
136 distributions) rather than using these trait assumptions to build the tree. We also revise the list of
137 traits based on functional studies of cyclases over the past 16 years. Our use of probabilistic
138 phylogenetic methods follows that of previous workers (Desmond & Gribaldo, 2010; Frickey &
139 Kannenberg, 2009; Gold et al., 2017; Santana-Molina et al., 2020), and we find that our analyses,
140 where they overlap, recapitulate theirs. Second, we analyze the astounding cyclase diversity
141 since sequenced, with special attention to divergent sequences from environmental metagenomes
142 and metagenome-assembled genomes as a potential source of cyclases that synthesize orphan
143 biomarkers or novel natural products. Previous environmental sequencing efforts focused on
144 triterpenoid cyclases found entire clades comprised solely of metagenomic sequences (Pearson et
145 al., 2007; Pearson & Rusch, 2009), all of which remain uncharacterized. Third, we broaden the
146 analysis to include the entire terpenoid cyclase protein superfamily allowing us to root the
147 triterpenoid cyclase protein family tree, again independently of assumptions about the manner in
148 which functional traits evolve. Fourth and finally, using the evolutionary framework provided by
149 an explicit model of protein evolution, sequence diversity, and the alpha-alpha toroid fold, we
150 identify cyclases that branch at key transitions in the superfamily's evolution. We test the
151 function of these cyclases via heterologous expression in *Escherichia coli* and find two
152 surprising enzymes. One is a cyclase that synthesizes isoarborinol, an orphan biomarker from the
153 Paleozoic. The other is a group of cyclases within the squalene-hopene cyclase clade which
154 cyclize oxidosqualene (the native sterol cyclase substrate) but not squalene (the native squalene-
155 hopene cyclase substrate). In addition to other lines of evidence, the substrate specificity of these
156 cyclases suggests that alkenyl substrates may not be ancestral to triterpenoid cyclases, and were
157 gained relatively late in the squalene-hopene cyclase lineage.

158 Results and Discussion

159 Phylogeny

160 We began our analyses of cyclase evolution by estimating the phylogeny of all proteins known to
161 share the cyclase fold, including triterpenoid, diterpenoid, and meroterpenoid cyclases, and

cyclase homologs of unknown function from environmental metagenomes (Figure 2). To do so, we retrieved cyclase homologs from public databases using BLAST and Pfam-based methods (see Material and Methods). Combined, these searches resulted in 44,209 unique sequences after removing duplicates and filtering for length. Briefly, all sequences in this database were aligned, alignments were subset to 1,139 sequences, and phylogenetic trees were estimated under maximum likelihood.

We tested the robustness of tree topology to uncertainty in alignment, model choice, and history of domain duplication and loss. Figure 2 shows the cyclase phylogeny estimated under EX_EHO+R5, which applies separate evolutionary models to each site depending on its probability of being on the protein's surface or buried, and depending on whether it is in an alpha helix, a loop, or another structural element (Le & Gascuel, 2010). This model also calculates the probability of each site being in one of 5 evolutionary rate categories. While the distribution of these categories differed slightly between the β and γ domains, we found that the likelihood-maximizing partitioning scheme for the concatenation of domains was a single partition. The topology estimated under EX_EHO+R5 is robust to omission of γ domains, to alignment masking, to larger and smaller subsets of the sequence database, and to model misspecification, as the same deep splits were recovered under runner-up models (Supplementary Table 2). However, ultrafast bootstrap support improved dramatically under better-fitting models. Thus, although protein sequences contain limited evolutionary information, and can lose information as they evolve due to site saturation caused by repeated substitution (Xia et al., 2003), the phylogeny presented here represents a strong hypothesis given the sequence information available.

The domain structure of all cyclases in this tree is either β, βγ, βx, or αβγ. We noticed that β and γ domains crystal structures do not align well, despite clear secondary structural and topological homology (Supplementary Table 1). Similarly, diterpenoid γ domains and triterpenoid γ domains do not align well, due either to divergence or because they arose from independent duplications of the β domain. If so, the diterpenoid and triterpenoid cyclase γ domains would still be homologous but could have arisen from β domains with different sequences, in different organisms, resulting in different trajectories through fitness landscapes and resultingly independent evolutionary constraints, rates, and substitution processes. For this reason, we used structural alignments of protein crystals to create custom hidden Markov models (HMMs) for the β domain, the diterpenoid cyclase γ domain, and the triterpenoid cyclase γ domain. Our triterpenoid γ domain HMM did not pick up the diterpenoid γ domain, and vice versa. We estimated cyclase phylogeny with the two γ domains aligned separately and recovered the same topology as when γ domains were aligned together, or omitted.

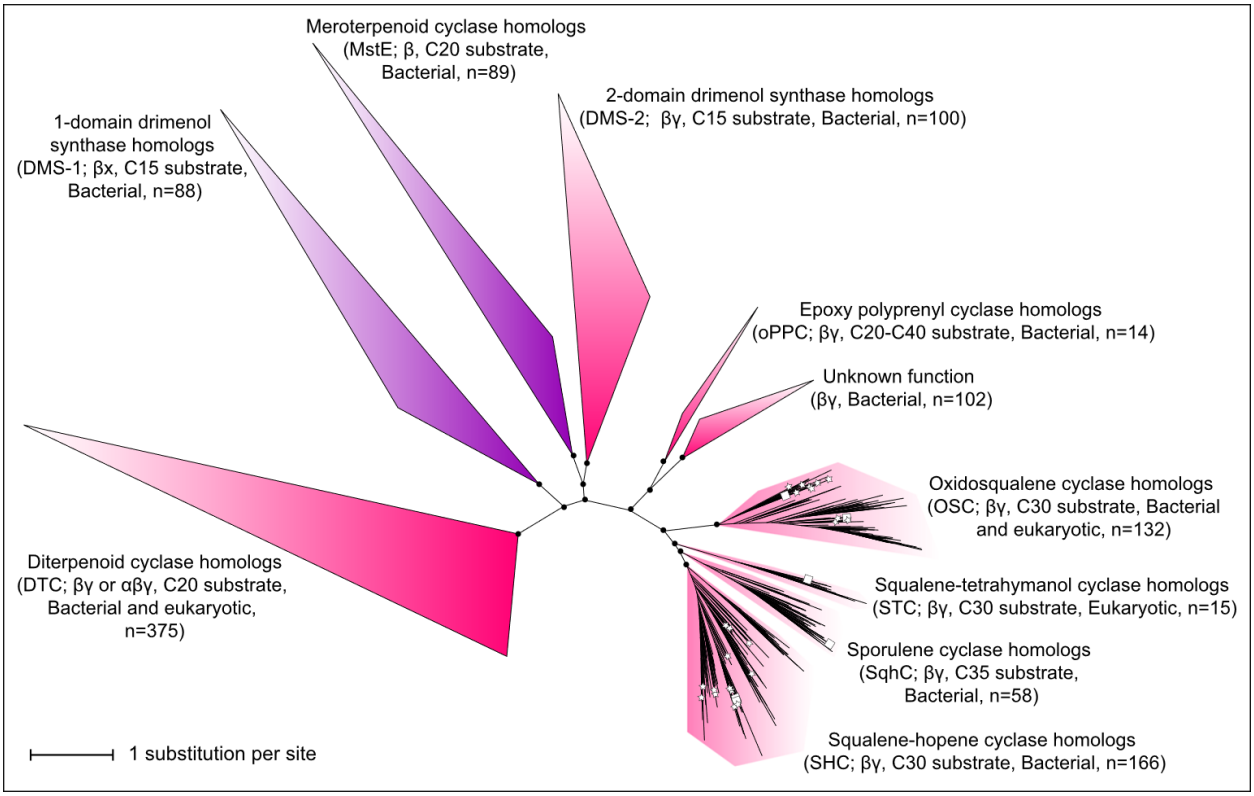

Figure 2: Maximum-likelihood phylogeny of terpenoid cyclases estimated under the EX_EHO model with 5 free rate categories. Stars indicate cyclases from environmental metagenomes; squares indicate cyclases of known function. Ultrafast bootstrap support is shown for deep splits, where black circles indicate bootstrap ≥ 97. Purple clades are composed of proteins without a γ domain (β-only or βx) as annotated; pink clades are composed of two-domain (βγ or αβγ) proteins.

Major clades

The overall tree topology is consistent with previously published triterpenoid and diterpenoid subtrees (Desmond & Gribaldo, 2010; Frickey & Kannenberg, 2009; Gold et al., 2017; Santana-Molina et al., 2020), and with a recent tree of all terpenoid cyclases (Hoshino & Villanueva, 2023). Squalene-hopene cyclase homologs (SHCs) and oxidosqualene cyclase homologs (OSCs) form large monophyletic clades, with squalene-tetrahymanol cyclase homologs (STCs) from microbial eukaryotes and sporulenol synthase homologs (SqhCs) from Bacillota forming small groups sister to SHC, and bacterial OSC groups branching earliest among the OSC subclades.

Sister to this mostly-triterpenoid group is another pair of clades, one of which is a small group of epoxy-polyprenyl cyclases. These enzymes take a substrate that is epoxidized like oxidosqualene but has all head-tail joined isoprene units rather than the central tail-tail linkage of squalene and oxidosqualene (Figure 1). The epoxy-polyprenyl cyclases include Lon15, a meroterpenoid octaprenyl (C_{40}) cyclase from Actinomycete *Streptomyces argenteolus* (Hayashi et al., 2007) and AtoE, a unique C_{25} cyclase from Actinomycete *Amycolatopsis tolypomycina* NRRL B-24205 (Kim et al., 2019). This clade is quite taxonomically restricted, with most

sequences coming from Actinomycetes and a handful from Chloroflexota and Clostridia. Surprisingly, those diterpenoid cyclases that utilize an epoxidized substrate – Bra4 and PlaT2, epoxy-diterpenoid cyclases from Actinomycetes *Nocardia brasiliensis* IFM 0406 (Hayashi et al., 2008) and *Streptomyces* sp. Tu¨ 6071 (Dürr et al., 2006), respectively – also fall in this clade rather than with other diterpenoid cyclases. It seems that the number of substrate isoprenoid chain carbons is a polyphyletic trait in the cyclase tree, as demonstrated by these diterpenoid cyclases and MstE (which takes a substrate with a dihydroxybenzoate-functionalized C20 isoprenoid chain) and by the triterpenoid cyclases, which are paraphyletic with respect to C35 SqhC. The position of SqhC also demonstrates the relative lability of isoprenoid chain linkage (all head-tail vs. tail-tail) as a trait.

Sister to the epoxy-polyprenyl cyclase group is a clade of unknown function referred to in the literature (e.g., Santana-Molina et al., 2020) as "SHC-like." This clade contains Gracilicutes sequences from the PVC group (Planctomycetes, Verrucomicrobia, Chlamydiae), the FCB group (Fibrobacterota, Chlorobiota, Bacteroidota), Spirochaetota, and Acidobacteriota. The function of these proteins, including those from cultured and genetically tractable organisms such as *Gemmata obscuriglobus,* has not been demonstrated, and their molecular structure has not been determined. However, we found that for both groups, the γ domains are detectable by the custom "triterpenoid" γ domain HMM we generated but not by our "diterpenoid" γ domain HMM. This could be due to functional divergence of the γ domains in these two halves of the tree (see "evolutionary implications" below).

On the tree's central long branch are three subgroups, each containing a single characterized enzyme. One is composed of monodomain (β-only) meroterpenoid cyclase homologs, including a single characterized enzyme from Cyanobacterium *Scytonema sp.* PCC 10023, which cyclizes a quinone precursor with a functionalized benzyl headgroup and a C_{20} isoprenoid chain (Moosman et al. 2020). Interestingly, this cyclase, called MstE, uses a DxD motif rather than the canonical DxDD motif to protonate an alkenyl substrate. Other sequences in this clade are from diverse bacterial taxa, including other Cyanobacteria as well as Alphaproteobacteria, Betaproteobacteria, Deltaproteobacteria, Gammaproteobacteria, Acidobacteriota, Myxococcota, Planctomycetes, Bacillota, Bacteroidota, Actinomycetota, and Chloroflexi, spanning the Gracilicutes and Terrabacteria divisions, although not recapitulating their branching order. There are also several archaeal sequences, including from cultured representatives such as *Methanosarcina acetivorans* C2A. The other subgroups are composed of drimenyl pyrophosphate synthases (DMS), recently-discovered enzymes that cyclize a C15 substrate. One ("DMS-2") contains the two-domain DMS characterized from Actinomycete *Streptomyces showdoensis* (Pan et al., 2022), along with homologs from other bacteria, while the other ("DMS-1") contains the DMS characterized from Bacteroidete *Aquimarina spongiae* (Vo et al., 2022), which has only one alpha-alpha toroid domain, fused to a haloacid dehalogenase domain. This clade also contains homologs from Actinomycetota, FCB group bacteria, and fungi.

The final major clade consists of diterpenoid cyclases. All characterized proteins in this group take geranylgeranyl pyrophosphate (GGPP) as a substrate. The two most basal clades consist of only bacterial proteins, including characterized bacterial cyclases that make *ent*-copalyl pyrophosphate (Jayasuriya et al., 2007; Morrone et al., 2009; Smanski et al., 2011), a gibberellin or antibiotic precursor, and the cyclase from *Kitasatospora griseola* that synthesizes the terpentecin precursor (Dairi et al., 2001). This is sister to a pair of clades containing fungal *ent*-copalyl pyrophosphate cyclases (Quin et al., 2014) and plant *ent*-copalyl pyrophosphate cyclases (Jia et al., 2022), respectively. The plant clade has many paraphyletic groups of prokaryotic homologs on its stem, including the *Mycobacterium tuberculosis* cyclase that synthesizes the tuberculosinyl adenosine precursor (Young et al., 2015) and homologs from the Asgardarchaeota that produce the same molecule (McShea et al., 2025).

Functional analysis of putative metagenomic cyclases

Cyclases from environmental metagenomes are distributed throughout the tree, especially in bacteria-dominated clades (Figure 3). They fall in the SHC and OSC crown groups, as well as on the stems of these clades, in small groups or as singleton branches. To determine if these metagenomic cyclases are capable of cyclizing squalene or oxidosqualene, we expressed a select set of cyclases in two strains of *E. coli,* engineered to synthesize squalene or oxidosqualene, respectively, and measured production of polycyclic lipids by gas chromatography-mass spectrometry (GC-MS). We identified 14 triterpenoid cyclase homologs from environmental metagenomes which either branched on the OSC or SHC stem lineage rather than within the crown group or branched within the crown in monophyletic subgroups with no cultured representatives. Many of these cyclases have substitutions at sites of known importance for cyclization of squalene or oxidosqualene.

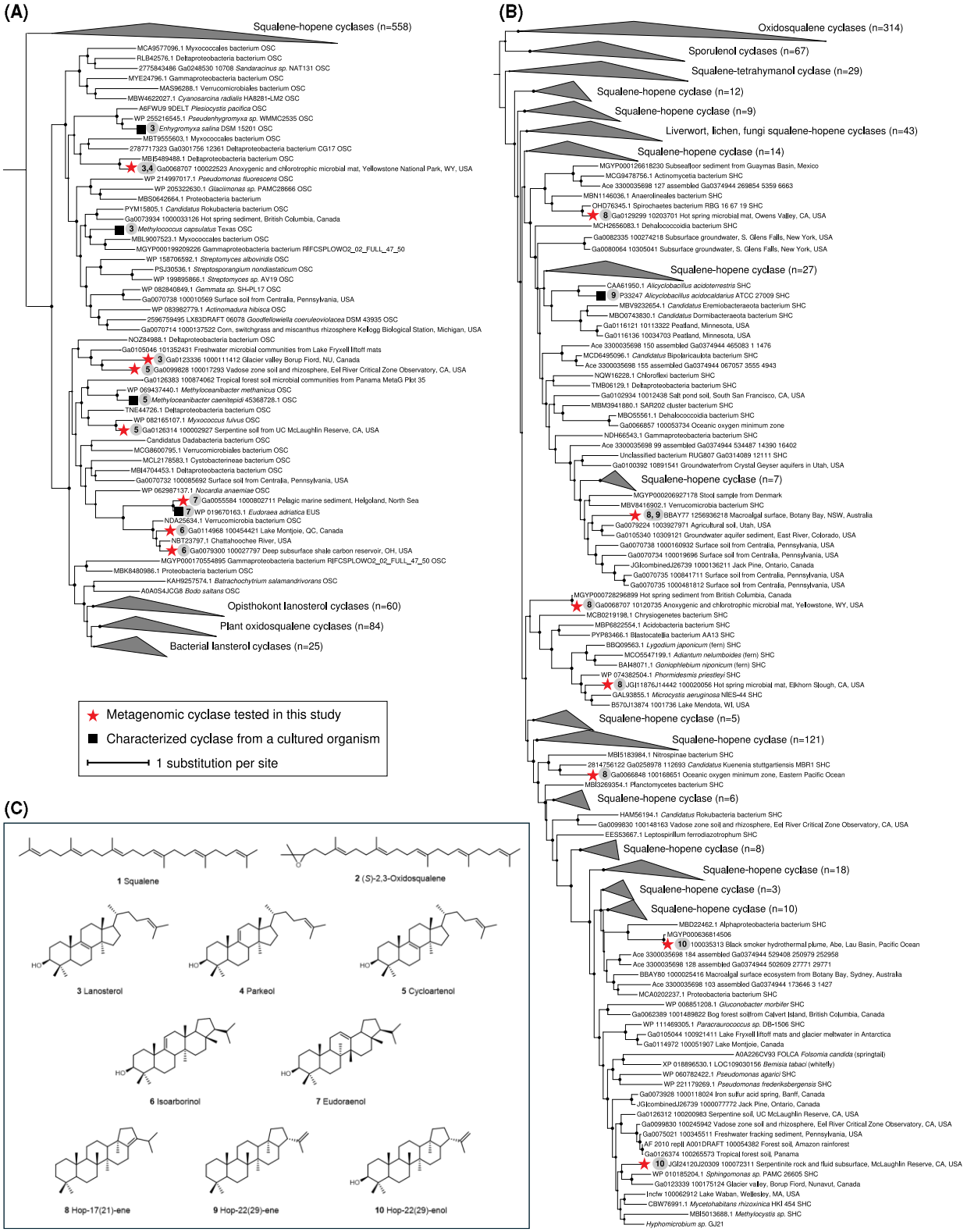

Figure 3: Phylogenetic positions of triterpenoid cyclases from environmental metagenomes heterologously expressed in *E. coli*. Metagenomic cyclases are marked with red stars, while previously characterized enzymes from

cultured organisms are marked with black squares, and both are annotated with product profiles. (A) Oxidosqualene cyclases. (B) Squalene-hopene cyclases. (C) Substrates and products of cyclases from environmental metagenomes.

Many triterpenoid cyclases were active in the heterologous expression system. Most enzymes from early-branching OSC subclades produce lanosterol or cycloartenol (Table 1; structures shown in Figure 3C), consistent with their phylogenetic proximity to cyclases from cultured bacteria known to make those products (Figure 3A). None of the cyclases from early-branching bacterial OSC groups had activity for squalene, which requires the greater active site acidity of an SHC (Xu et al., 2004).

One cyclase from a marine metagenome produces eudoraenol, a recently discovered arborane triterpenoid. It is the second eudoraenol-synthesizing enzyme to be reported after the discovery of a eudoraenol cyclase in *Eudoraea adriatica* (Banta et al., 2017), a marine heterotroph from phylum Bacteroidetes (Alain et al., 2008). The new eudoraenol cyclase has the same substitutions shown by Banta et al. (2017) to be sufficient for eudoraenol production – tryptophan to serine at position 230 (*Homo sapiens* OSC numbering), histidine to tyrosine at position 232, tyrosine to valine at position 521, and asparagine to tyrosine at position 697. In the enzyme active site, these residues are clustered near the substrate tail where the fifth ring forms in arborinols.

Closely related to the eudoraenol cyclases are two cyclases from deep subsurface and inland lake ecosystems which we found synthesize isoarborinol (Figure 4). These cyclases have the eudoraenol cyclase substitutions at positions 230, 232, and 697 and a unique cysteine at position 503. The subsurface cyclase contig is predicted by the Whokaryote classifier (Pronk & Medema, 2022) to be prokaryotic, making it the first reported bacterial source of isoarborinol. While the lake cyclase contig is too short for classification, its phylogenetic proximity to the subsurface and other bacterial cyclases make it likely to be prokaryotic as well. Arborinol is named for the tree from which it was first isolated (Kennard et al., 1965; Vorbrüggen et al., 1963), but the arborane fossil record stretches back to the Permian (Hauke et al., 1995), well before the evolution of angiosperms, leading to the hypothesis (Ourisson et al., 1982) that microbial sources of arborinols, and specifically isoarborinol, must exist. The formation of isoarborinol by these two cyclases identifies a modern and most likely bacterial source for this "orphan" biomarker. Bacterial arborinol cyclases form a monophyletic clade within one of the bacterial OSC clades. Thus, it is unclear whether bacterial arborinol cylcases arose before or after the radiation of eukaryotic OSCs, which show a pattern of vertical inheritance (i.e., major clades of plant, fungal, and animal cyclases largely recapitulate organismal phylogeny). The bacterial arborinol cyclases are thus a reasonable, but not definitive, source for Paleozoic and Proterozoic arborane fossils.

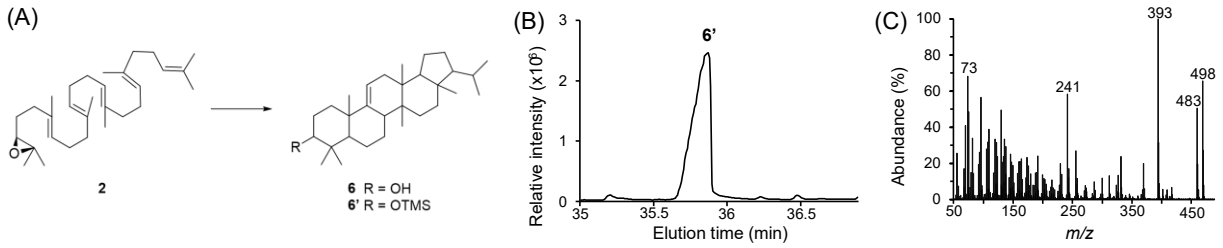

Figure 4: Biosynthesis of isoarborinol by a triterpenoid cyclase from a deep subsurface shale metagenome. (A) The reaction performed by isoarborinol cyclase. (B) GC-MS total ion chromatogram of total lipid extract from *E. coli* strain HM1 expressing the oxidosqualene biosynthesis pathway and Ga0079300_100027797, derivatized to trimethylsilyl (TMS) ethers. (C) Mass spectrum of peak at 36.0 min, identified by comparison to Vorbrüggen et al. (1963) and by nuclear magnetic resonance (NMR), see Supplementary Figure 3 and Supplementary Table 3.

Most enzymes from early branching SHC subclades (Figure 3B) that we tested produce typical hopene isomers hop-17(21)-ene and/or hop-22(29)-ene from a squalene substrate (Table 1). Hopene-producing enzymes also all had activity for oxidosqualene, typically producing a single hopanol isomer (hopan-22(29)-ol), consistent with previous work which has shown squalene cyclases to be nonspecific (Abe & Rohmer, 1994; Hammer et al., 2013; Rohmer et al., 1980; Seitz et al., 2012). However, a small group of cyclases within the SHC crown had activity for only oxidosqualene and no activity for squalene, producing only hopanols. This is surprising given that all other cyclases in the SHC clade have some activity for squalene, including sporulenol cyclases, which are functional for both β-curcumene and squalene, and tetrahymanol cyclases. These oxidosqualene cyclases in the SHC clade do not lack any residues needed to activate squalene. Indeed, they have the electrophilic DxD[D] motif thought to be necessary to protonate an alkenyl substrate. It is possible that these enzymes are in fact true SHCs but cannot cyclize squalene in the *E. coli* cell under the conditions tested, perhaps due to minor protein misfolding that weakens the electrophilicity of the active site, retaining activity for the more reactive substrate but abolishing it for the more recalcitrant one. It is also possible that these hopanol-producing enzymes are true OSCs in the functional sense, despite being SHCs in the phylogenetic sense, and that substrate functionalization is a relatively labile trait in this group.

Table 1: Substrates and products of triterpenoid cyclases from environmental metagenomes heterologously expressed in *E. coli*. Products were determined by comparison to published GC-MS spectra (see Supplementary Figures 1 and 2), and in the case of isoarborinol, confirmed by NMR (Supplementary Figure 3 and Supplementary Table 3).

Cyclase gene ID	Environmental Metagenome	Clade	Substrate	Major products	Strain
Ga0079300_100027797	Deep subsurface shale carbon reservoir, OH, USA	OSC	Oxidosqualene	Isoarborinol	HM1
Ga0123336_1000111412	Glacier valley Borup Fiord, NU, Canada	OSC	Oxidosqualene	Lanosterol	HM3
Ga0114968_100454421	Lake Montjoie, QC, Canada	OSC	Oxidosqualene	Isoarborinol	HM4
Ga0055584_1000802711	Pelagic marine sediment, Helgoland, North Sea	OSC	Oxidosqualene	Eudoraenol	HM6
Ga0099828_100017293	Vadose zone soil and rhizosphere, Eel River Critical Zone Observatory, CA, USA	OSC	Oxidosqualene	Cycloartenol	HM23
Ga0126314_100002927	Serpentine soil from UC McLaughlin Reserve, CA, USA	OSC	Oxidosqualene	Cycloartenol	HM31
Ga0068707_100022523	Anoxygenic and chlorotrophic microbial mat, Yellowstone National Park, WY, USA	OSC	Oxidosqualene	Lanosterol, parkeol	HM33
Ga0129299_10203701	Hot spring microbial mat, Owens Valley, CA, USA	SHC	Squalene	Hop-17(21)-ene	RAV2
Ga0068707_10120735	Anoxygenic and chlorotrophic microbial mat, Yellowstone National Park, WY, USA	SHC	Squalene	Hop-17(21)-ene	RAV3
Ga0066848_100168651	Oceanic oxygen minimum zone, Eastern Pacific Ocean	SHC	Squalene	Hop-17(21)-ene	RAV7
JGI11876J14442_100020056	Hot spring microbial mat, Elkhorn Slough, CA, USA	SHC	Squalene	Hop-17(21)-ene	RAV10
BBAY77_1256936218	Macroalgal surface, Botany Bay, NSW, Australia	SHC	Squalene	Hop-17(21)-ene, hop-22(29)-ene	RAV13
Abe_100035313	Black smoker hydrothermal plume, Abe, Lau Basin, Pacific Ocean	SHC	Oxidosqualene / squalene	Hopanol / no product	HM163
JGI24120J20309_100072311	Serpentinite rock and fluid subsurface biosphere, McLaughlin Reserve, CA, USA	SHC	Oxidosqualene / squalene	Hopanol / no product	HM164

Adaptation in the terpenoid cyclase superfamily

What evolutionary dynamics drove the diversification of terpenoid cyclases? Did the diversity of substrates and products in this superfamily arise from neutral divergence and drift, or did natural selection play a role? To assess the relative influences of drift and selection, we employed a quantitative test that interprets positive selection as an elevated rate of amino acid mutations in functional regions of the protein structure, along a given branch in a phylogenetic tree (similar to Maddamsetti & Grant, 2022; Ritchie et al., 2021). This approach applies the theory developed for protein-coding nucleotide sequences (e.g. the branch-site test, Gharib & Robinson-Rechavi, 2013; McDonald & Kreitman, 1991; Messer & Petrov, 2012; Yang & Dos Reis, 2011) to protein

structures, which allows the test to work on deeply divergent groups where nucleotide substitutions have saturated but structure has evolved more slowly.

Along all the major internal branches of the terpenoid cyclase phylogeny, there are two where a functional domain had a significantly elevated proportion of amino acid substitutions over background (Figure 5). One is the stem of the triterpenoid cyclases which take an alkenyl substrate (that is, squalene-hopene, sporulenol, and tetrahymanol cyclases), where the active site of the protein has an unusually high substitution density. The second is the stem of the group of unknown function, where the SHC-type dimer contact surface had unusually high substitution density. We interpret clustering of substitutions in a functional region as a potential signal of protein adaptation, in which case these may be two instances of positive selective pressure and/or functional divergence in the evolutionary history of triterpenoid cyclase homologs. Adaptation could be to new environmental conditions (e.g. to a new host organism after horizontal gene transfer), to new function, or both. That only two instances of rate shifts in functional regions were detectable means either that adaptation on other branches occurred at a molecular length scale undetectable when focusing on functional regions (e.g. single-residue substitutions), at a temporal scale undetectable when focusing on long branches (e.g. rapid evolution within crown groups), or that triterpenoid cyclase divergence was otherwise dominated by drift.

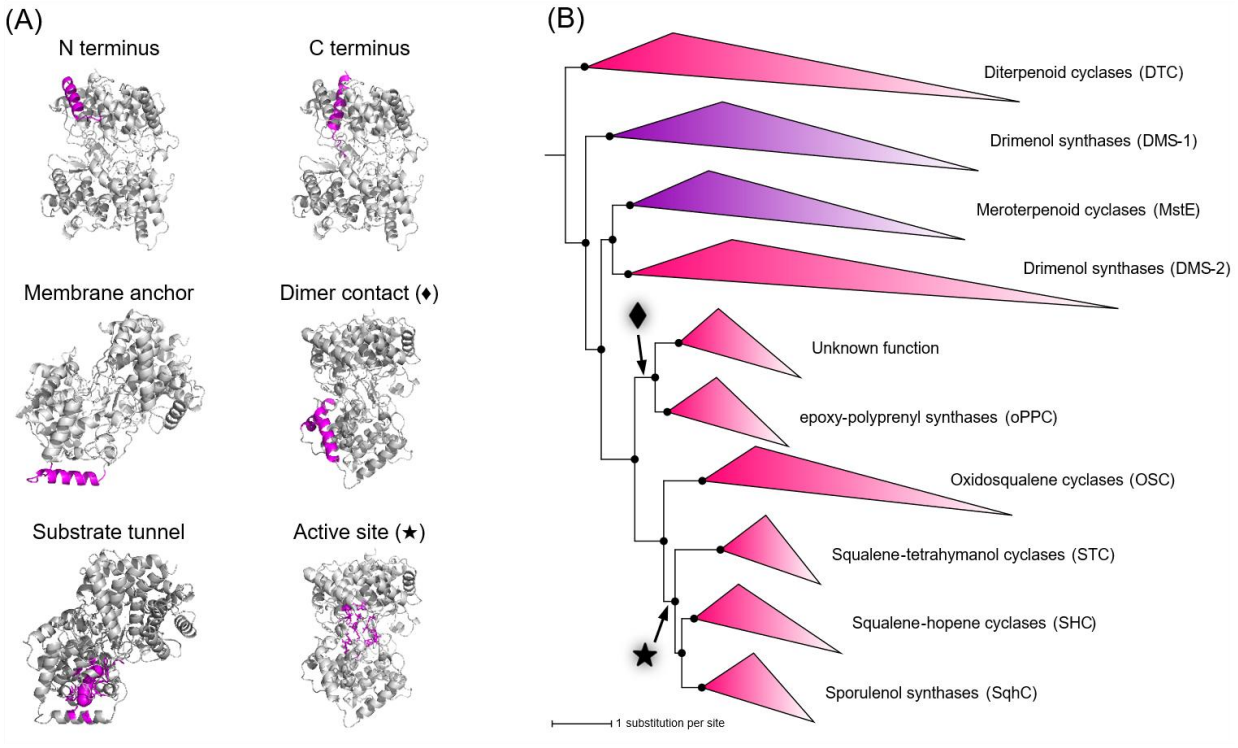

Figure 5: Windows of rapid evolution in functionally important domains of triterpenoid cyclases. (A) Functional regions shown on the squalene-hopene cyclase structure (PDB ID 1SQC). Regions found to have an evolutionary rate elevated above background are marked with a rhombus (♦) and a star (★). (B) Branches with a rapidly-evolving functional region, marked with symbols corresponding to panel A. As in Figure 2, purple clades are composed of proteins without a G domain (β-only or βX); pink clades are composed of two-domain (βγ or βγX) proteins.

Evolutionary Implications

sTo interrogate domain evolution in the cyclase family, we turned to protein structure, which evolves more slowly than sequence composition (Caetano-Anollés et al., 2009; Ingles-Prieto et al., 2013). We applied structural data to estimate relationships in three ways: we used structures to better align sequences before estimating phylogeny under a model of sequence evolution (Figure 6A), we measured relationships as simple distances between structural coordinates (Figure 6B), and we measured relationships using the local structural alphabet implemented in Foldtree (Moi et al., 2023) (Figure 6C).

The structure-aligned sequence tree (Figure 6A) and the RMSD dendrogram (Figure 6B) show a deep split between β domains and γ domains. The β and γ clades have the same internal branching order in the structure-aligned sequence tree, consistent with a single duplication event at the root followed by parallel divergence, with the only other domain rearrangements being the loss of the γ domain in monodomain cyclase MstE (PDB ID 6SBB), gain of the α domain in the ancestor of plant cyclases, and loss of the γ domain later within that clade. The RMSD dendrogram also supports this scenario, and differs only in the relationships among β domains, placing the β domains of bacterial diterpenoid cyclases sister to those of the triterpenoid cyclases rather than those of the plant cyclases. The Foldtree estimate (Figure 6C) predicts a much more complex scenario, with a deep split between plant β domains and the rest of the structures. After the split with plant βs, the first branches are triterpenoid and sesquiterpenoid γ domains, then tri- and sesqui-terpenoid βs, then diterpenoid βs and γs. This is consistent with our hypothesis of triterpenoid and diterpenoid β domains arising from independent duplication events but suggests a surprising γ domain ancestor for all non-plant β domains. Given that Foldtree does not yet measure statistical support, and given the agreement between the structure-aligned sequence tree, the RMSD dendrogram, and our large sequence-based phylogeny (Figure 2), we tentatively reject the Foldtree estimate and the independent duplications hypothesis. Most evidence supports a single evolution of the terpenoid cyclase γ domain. However, this scenario should be revisited as structural phylogenetics (Garg & Hochberg, 2025; Malik et al., 2020; Moi et al., 2023) becomes more sophisticated (Mutti et al., 2025).

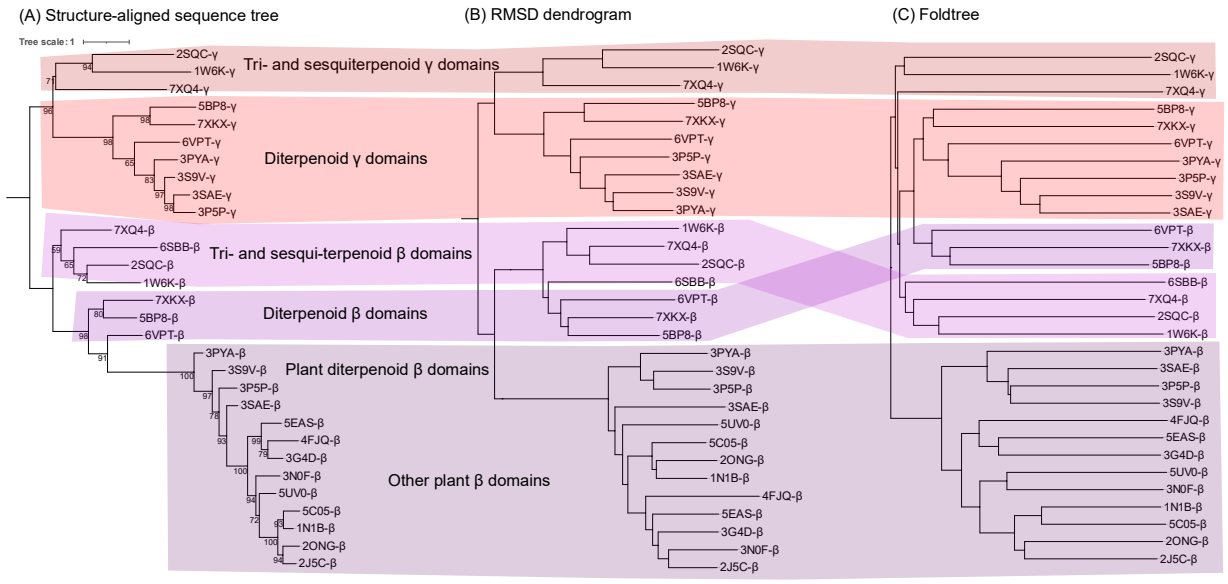

Figure 6: Relationships between cyclase domains estimated using three structure-informed methods. (A) Phylogeny estimated using standard sequence-based method after aligning sequences using structure. (B) Neighbor-joined dendrogram of domains using root mean squared deviation (RMSD) as distances. (C) Structure relationships estimated using the Foldtree algorithm. Red and purple stripes are for visual aid, to show the position of each group across trees.

Given the possible selection on oligomerization and the active site revealed through our structural analyses of cyclases, we can posit and assess evolutionary scenarios focused on related traits. The longest internal branch of the cyclase phylogeny is between the diterpenoid cyclases and all other proteins (Figure 7A). Rooting the tree on this branch, in agreement with the root position in the β and γ domain subtrees (Figure 6), places all-bacterial clades basal to the eukaryotic clades. This is consistent with separate transfer events of diterpenoid cyclases from bacteria to plants and to fungi, likely at different times and involving different donor organisms and proteins. For this and most possible superfamily roots, the triterpenoid cyclase root is between the squalene-hopene (and STC and SqhC) and oxidosqualene cyclases. This OSC-SHC root position does not further elucidate the biosynthetic substrate and product of the triterpenoid cyclase ancestor (i.e. whether it produced hopanoids, sterols, or a different molecule). Given that squalene-hopene cyclase was likely present in the common ancestor of Gracilicutes, one of the major divisions of bacteria (Santana-Molina et al., 2020), the OSC-SHC ancestor is likely quite ancient, although hopanes and steranes do not appear in the fossil record until 1.6 billion years ago (Brocks et al., 2005; French et al., 2015).

However, rooting on the diterpenoid cyclases implies either three independent evolutions of the γ domain (Figure 7B), or two losses of the γ domain on the MstE and DMS-1 stems (Figure 7C). The first scenario is not supported by the analysis of domain evolution discussed above (Figure 6). The second scenario is consistent with the domain trees but seems biochemically unlikely given that both domains are necessary for function in modern two-

434 domain enzymes for substrate positioning and water exclusion (Abe, 2007; Fischer & Pearson,
435 2007; Wendt, 2005). The unusual cyclase DMS-1 may have had an easier path, with the novel
436 haloacid dehalogenase-like domain (Vo et al., 2022) replacing the γ domain, or inserted between
437 β and γ, in one mutational event, never requiring an evolutionary intermediate with a water-
438 exposed active site. On the other hand, the monodomain cyclase merosterolic acid synthase
439 (MstE) excludes water with a mobile loop that caps the active site when substrate is bound
440 (Moosmann et al., 2020). This loop is significantly shorter in the closest related two-domain
441 cyclase, drimenyl diphosphate synthase (Pan et al., 2022), suggesting that unless loop extension
442 and domain loss occurred simultaneously, the active site of the MstE ancestor would have been
443 exposed to water, making catalysis impossible. However, domain loss is not unknown in
444 terpenoid cyclases (Oldfield & Lin, 2012). Biochemical investigation of basal or ancestral mono-
445 and di-domain cyclases, as well as of, e.g., synthetic constructs of monodomain cyclases with
446 shortened mobile loops, are necessary to establish the relative probabilities of domain loss and
447 duplication in this superfamily.

448 If we then take domain loss in an MstE ancestor to be less likely than domain loss in a
449 DMS-1 ancestor, a monodomain cyclase root (Figure 7D) becomes worth investigating. Rooting
450 the structure-informed domain tree and the RMSD domain dendrogram on the monodomain
451 cyclase (PDB ID 6SBB) places sesqui- and tri-terpenoid β domains as the next most basal,
452 followed by remaining β domains, from which a monophyletic γ domain clade emerges.
453 Although this position is in all cases far from the tree midpoint, it remains plausible given the
454 evolutionary rarity of domain loss (Bridgham et al., 2009), and balances that rarity with the
455 evidence for a single origin of the γ domain.

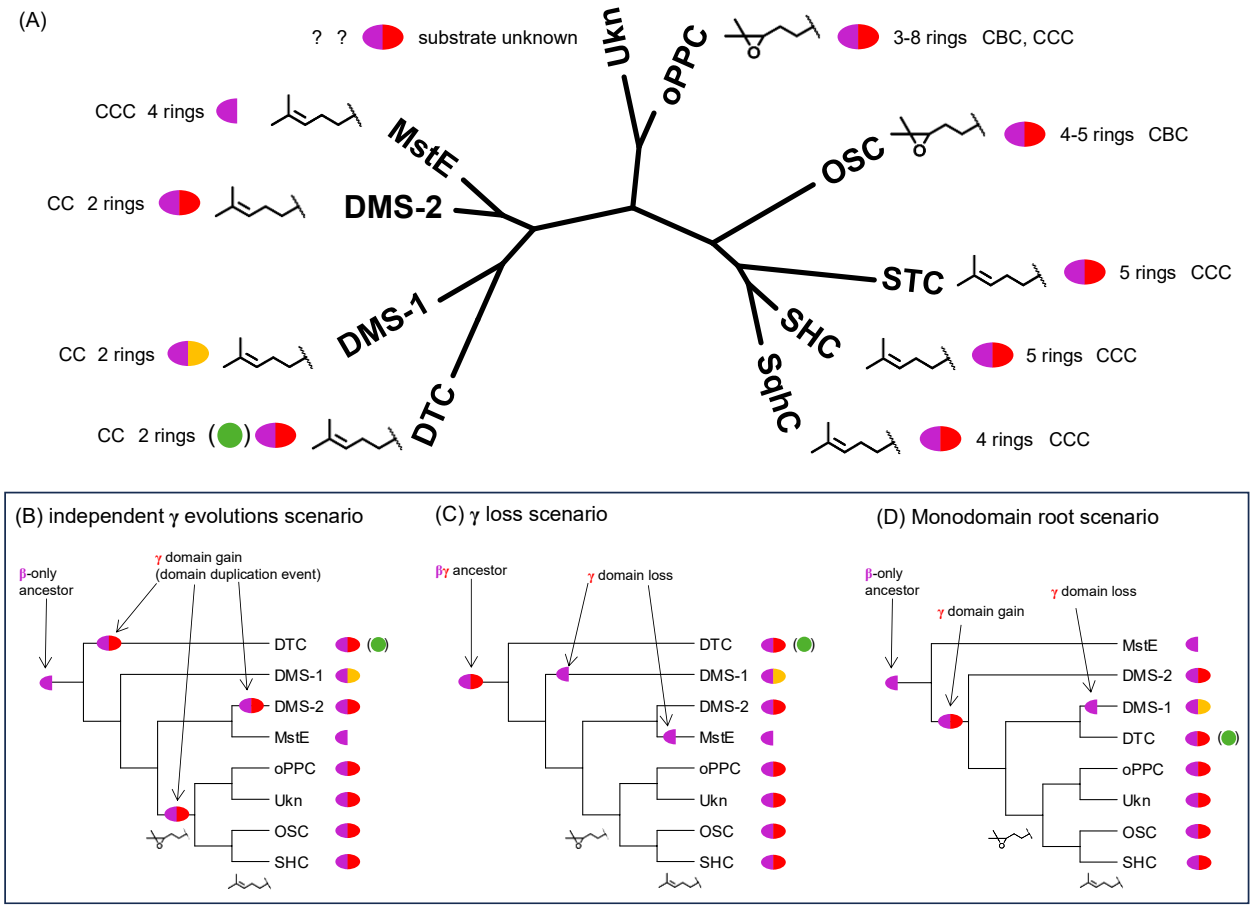

Figure 7: Distribution of terpenoid cyclase traits and hypotheses for superfamily root position. (A) Cyclase traits mapped across the unrooted tree, which has the same topology and internal branch lengths as the tree shown in Figure 2. Indicated traits include the number of rings in the final product, the structural conformation of the substrate (CBC: chair-boat-chair or CCC: chair-chair-chair); whether the substrate is epoxidized, and the number of domains (β = purple half-oval, γ = red half-oval, α (present in the plant subclade of diterpenoid cyclases) = green circle, haloacid dehalogenase domain = orange half-oval). (B-D) Alternate hypotheses for the root position of the cyclase tree with substrate and domain structure mapped. For simplicity, STC and SqhC are not shown, as their domain architecture and substrate functionalization are identical to SHC. OSC: oxidosqualene cyclases; SHC: squalene-hopene cyclases; MstE: meroterpenoid cyclases; DTC: diterpenoid cyclases; DMS: drimenyl synthases; Unk: homologs of unknown function; oPPC: epoxy-polyprenyl cyclases.

Within the triterpenoid cyclases, the oxidosqualene cyclase clade contains enzymes that form different numbers of rings (4 for sterol cyclases, 5 for arborinol cyclases). The squalene-hopene cyclases, which form 5 rings, have a stem group comprised of enzymes that form 4 (sporulenol cyclases) or 5 (squalene-tetrahymanol cyclase) rings respectively. Although we have found an outgroup with fewer rings (3 for Bra4 and PlaT2) as predicted by Fischer and Pearson (2007), this group also contains an enzyme that forms many more (8 for Lon15). A gradualist model of evolution, whereby additional rings evolve irreversibly and one-at-a-time, does not fit the data for this trait; rather, the number of rings is relatively labile. This is not surprising since anti-Markovnikov (secondary) carbocations are no longer considered to be cyclization

476 intermediates (Hess, 2002; Hess & Smentek, 2013; Tantillo, 2010), removing the physical
477 argument for a teleology of ring number.

478 Finally, substrate functionalization could have followed two evolutionary paths in the
479 updated phylogeny. Fischer and Pearson (2007) hypothesized that the use of an epoxidized
480 substrate evolved from an alkenyl ancestor once, on the oxidosqualene cyclase stem after the
481 divergence of OSC and SHC. In the updated phylogeny, because epoxy-polyprenyl cyclases also
482 used an epoxidized substrate, epoxide use could have evolved twice independently, once in the
483 ancestor of each group. Because alkenyl-substrate enzymes can also cyclize epoxidized
484 substrates, but epoxy-substrate enzymes cannot cyclize alkenyl substrates, evolutionary
485 transitions from alkenyl to epoxy substrates seem in general more likely than the reverse. This is
486 consistent with our discovery of enzymes within the squalene-hopene cyclase crown that only
487 work on an epoxidized substrate (Table 1). Alternately, use of an epoxidized substrate may have
488 evolved only once, in the common ancestor of oxidosqualene cyclases and epoxy-polyprenyl
489 cyclases, and subsequently been lost in the squalene-hopene cyclase/squalene-tetrahymanol
490 cyclase/sporulenol cyclase ancestor. Because epoxidized substrate biosynthesis requires
491 molecular oxygen (Jahnke, 1986; Jahnke & Nichols, 1986), this scenario is less likely if the
492 OSC-SHC-oPPC ancestor predates the radiation of Gracilicutes (Santana-Molina et al., 2020)
493 and if Gracilicutes predate the rise of molecular oxygen around 2.4 billion years ago (Gumsley et
494 al., 2017; Luo et al., 2016; Martinez-Gutierrez et al., 2023; Ostrander et al., 2024; Wang & Luo,
495 2025). However, it bears further investigation, e.g. via molecular clock and duplication-transfer-
496 loss analysis, given the uncertainties in both antecedents. The relative probabilities of the two
497 scenarios for substrate evolution will also be better-informed when the substrate of the
498 "unknown" clade is discovered.

499 Conclusions

500 Here we present an updated phylogenetic and functional analysis of triterpenoid cyclases. One
501 result is that the triterpenoid cyclase active site has an elevated evolutionary rate on the SHC-
502 STC-SqhC stem lineage. We hypothesize that this may be due to adaptation to an alkenyl
503 substrate from an epoxy ancestor of all triterpenoid and epoxy-polyprenyl cyclases, *if* this
504 ancestor evolved after the great oxidation event. Further, we found possible reversions to an
505 epoxidized substrate in the SHC clade. This experimental result, as well as the discovery of a
506 novel bacterial oxidosqualene-isoarborinol cyclase, shows that metagenomic cyclases from
507 uncultured sources are a significant source of diversity and can help inform our understanding of
508 the biochemical potential and evolutionary history of these fascinating enzymes. Future work
509 focused on the functional and structural characterization of cyclases, particularly those from the
510 uncharacterized clade sister to the epoxy-polyprenyl cyclases, will provide invaluable insight that
511 will help constrain the evolutionary scenarios proposed in this study.

512 # Materials and Methods

513 **Gene synthesis and molecular cloning**

514 Cyclase DNA sequences were codon-optimized for expression in *Escherichia coli* and
515 synthesized through the Department of Energy Joint Genome Institute (DOE JGI) DNA
516 Synthesis Science Program. These genes were obtained in the IPTG-inducible plasmid
517 pSRKGm-*lac*UV5-rbs5 (Banta et al., 2017) in *E. coli* TOP10. Plasmid DNA was isolated using
518 the GeneJET Plasmid Miniprep Kit (ThermoFisher Scientific) and sequenced to confirm
519 promoter and gene sequences by ELIM Biopharm (Hayward, CA) using the following primers:
520 5′-AATGCAGCTGGCACGACAGG-3′ (forward) and 5′-CCAGGGTTTTCCCAGTCAC-3′
521 (reverse), purchased from Integrated DNA technologies (Coralville, IA).

522 **Bacterial culture and heterologous expression**

523 *E. coli* DH10B expression strains were transformed by electroporation with the cyclase-bearing
524 pSRKGm plasmid (Khan et al., 2008) and two other plasmids: pJBEI2997 (Addgene plasmid
525 #351515) (Peralta-Yahya et al., 2011) encoding the MEV pathway and a pTrc99a derivative
526 (Amann et al., 1988) encoding squalene synthase (and, in some cases, squalene
527 monooxygenase). Gentamicin (15 µg/mL), carbenicillin (100 µg/mL), and/or chloramphenicol
528 (20 µg/mL) were added to media as appropriate to select for maintenance of each plasmid.
529 Plasmids are described in detail in Supplementary Table 4.

530 Strains were cultured in 50 mL TYGPN broth (RECIPE) at 37 °C while shaking at 225 rpm.
531 Cultures were induced with 500 µM isopropyl *β*-D-1-thiogalactopyranoside (IPTG) at an OD_{600}
532 of ~0.6, then incubated an additional 48 h at 30 °C while shaking at 225 rpm. Cells were
533 harvested by centrifugation at 4500 × g for 10 min at 4 °C. Cell pellets were stored at −20 °C
534 until lipid extraction. In every experiment, a negative control strain with an empty pSRKgm
535 plasmid and relevant positive controls (strains with pSRKGm containing a triterpenoid cyclase of
536 known function) were grown and extracted alongside experimental strains.

537 **Lipid extraction**

538 Cell pellets were extracted using a modified Bligh–Dyer method (Bligh & Dyer, 1959; Welander
539 et al., 2012): cells were resuspended in 2 mL of deionized water and transferred to a solvent-
540 washed Teflon centrifuge tube containing 5 mL of methanol and 2.5 mL of dichloromethane
541 (DCM), vortexed for 30 seconds, and sonicated for 1 hour in a water bath sonicator. 10 mL of
542 deionized water and 10 mL of DCM were then added, and samples were vortexed and incubated
543 overnight at −20 °C. Samples were centrifuged for 10 min at 2,800 × g, and the organic layer
544 was transferred to a baked glass vial and evaporated at 40 °C under a gentle stream of N_2 to yield
545 the total lipid extract (TLE). TLE was stored at −20 °C, and derivatized to trimethylsilyl ethers
546 with 1:1 (v:v) bis(trimethylsilyl)trifluoroacetamide (BSTFA):pyridine for 1 h at 70 °C before

547 analysis by gas chromatography–mass spectrometry (GC-MS). The alcohol-soluble fractions of
548 some TLEs were further purified by silicon column chromatography before derivatization, using
549 a solvent series of hexanes, 8:2 hexanes:DCM, DCM, 1:1 DCM:ethyl acetate, and ethyl acetate
550 (Summons et al., 2013).

551 **GC-MS analysis**

552 Lipids were separated with an Agilent 7890B Series GC equipped with two Agilent DB-17HT
553 columns (30 m × 0.25 mm i.d. × 0.15 µm film thickness) in tandem with helium as the carrier gas
554 at a constant flow of 1.1 ml/min and programmed as follows: 100 °C for 2 min, then 12 °C/min to
555 250 °C and held for 10 min, then 10 °C/min to 330 °C and held for 17.5 min. 2 µL of each sample
556 was injected in splitless mode at 250 °C. The GC was coupled to an Agilent 5977 A Series MSD
557 with the ion source at 230 °C and operated at 70 eV in electron ionization (EI) mode scanning
558 from 50 to 850 Da in 0.5 s. All lipids except isoarborinol were identified based on their retention
559 time and comparison with published spectra. Isoarborinol was identified as a likely sterol by GC-
560 MS, and its structure was confirmed by nuclear magnetic resonance spectroscopy (NMR).

561 **NMR Analysis**

562 NMR experiments were performed following Banta et al. (2017). TLE saponification was
563 accomplished by heating with 10% (vol/vol) sodium hydroxide (NaOH)/MeOH at reflux for 16
564 h. The reaction mixture was partitioned between water and hexane/ethyl acetate (EtOAc) 2:1; the
565 organic layers were filtered through neutral alumina and concentrated to dryness with a stream of
566 nitrogen. The saponified lipids were fractionated by preparative TLC on glass-backed plates (10
567 cm in length) coated with a 0.25-mm layer of silica gel 60 F254 using hexane/EtOAc 4:1 as the
568 developing solvent. The triterpenol fraction was further fractionated by reversed-phase HPLC
569 with a system consisting of a Waters 6000A pump, Waters 410 differential refractometer, and
570 two Altex Ultrasphere ODS 5-µm 10 × 250 mm columns in series using a flow rate of 3 mL/min
571 MeOH. After evaporation of the HPLC solvent, the triterpenols were characterized by NMR
572 using a Bruker Avance III HD with an Ascend 800 MHz magnet and a 5-mm TCI cryoprobe at
573 30 °C using deuterated chloroform (CDCl3) as the solvent. Calibration was by the residual
574 solvent signal (^{1}H: 7.26 ppm; ^{13}C 77.0 ppm).

575 **Sequence search and curation**

576 Cyclase homologs were retrieved from the Joint Genome Institute Integrated Microbial Genomes
577 & Microbiomes database (JGI IMG; https://img.jgi.doe.gov) and the National Center for
578 Biological Informatics nonredundant sequence database (https://www.ncbi.nlm.nih.gov/genbank)
579 using the set of query sequences listed in Supplementary Table 5. Protein-protein BLAST
580 algorithms were used in both cases, with an expect threshold of 0.05 and a word size of 5 for
581 NCBI, and an e-value cutoff of 1e-5 for JGI. Additionally, we downloaded all hits for Pfams
582 PF13243 (squalene-hopene cyclase C-terminal domain) and PF13249 (squalene-hopene cyclase
583 N-terminal domain) from the European Bioinformatics Institute (EBI;

584 https://www.ebi.ac.uk/interpro/entry/pfam/#table). Combined, these searches resulted in an
585 initial set of 44,209 unique sequences. BLAST results were filtered by length dependent on
586 query sequence as follows: 600-800 amino acids (aa) for triterpenoid cyclases, 550-700 aa for
587 epoxy-polyprenyl cyclases and members of the "unknown function" clade, 700-1,000 aa for
588 plant and fungal diterpenoid cyclases, 450-600 aa for bacterial diterpenoid cyclases, 550-650 aa
589 for DMS1, 450-550 aa for DMS2, and 300-400 aa for MstE. These length ranges were
590 determined based on length distribution of results and length range of functionally characterized
591 representatives of these groups. Sequences from EBI were filtered to a length range of 350 (the
592 length of the smallest known Type II cyclase; Moosmann et al., 2020) to 900 (the length of the
593 largest known Type II domain-bearing cyclase; Mitsuhashi et al., 2017) amino acids. Remaining
594 sequences were subsetted with CD-hit (command cd-hit -i in.fasta -o out.fasta -c 0.90 -n 5 -M
595 6000 -d 0 -T 8) (Fu et al., 2012) and aligned with the MAFFT v7.402 fftns2 algorithm
596 (Rozewicki et al., 2019), after which sequences lacking an aligned DxD motif were removed
597 (except for oPPC homologs with an [E/D]SA[E/N] motif; Dürr et al., 2006; Hayashi et al., 2008)
598 to yield a database of 13,680 sequences.

599 **Phylogenetic estimation**

600 All sequences in the database described above were aligned using the MAFFT linsi algorithm
601 (Rozewicki et al., 2019) implemented in Magus (Smirnov & Warnow, 2021). The alignment was
602 trimmed using trimAl with a gap threshold of 0.1 (command trimal -in in.aln -out out.aln -gt
603 0.1), and subsetted with a maximum sequence identity of 50% (command trimal -in in.aln -out
604 out.aln -maxidentity 0.5) (Capella-Gutiérrez et al., 2009). Model selection and maximum
605 likelihood phylogeny estimation were performed with IQ-TREE 2.2.0 (Bui et al., 2020), with the
606 model EX_EHO+R5 (Le & Gascuel, 2010) found to be the best model by Bayesian Information
607 Criterion and Akaike Information Criterion (Kalyaanamoorthy et al., 2017). For all phylogeny
608 estimation, we performed ≤ 3,000 ultrafast bootstrap replicates (Hoang et al., 2018). All input
609 commands are reported in Supplementary Table 2. Resulting phylogenetic trees were visualized
610 in Dendroscope (Huson & Scornavacca, 2012), Figtree (Rambaut, 2009), or iTOL (Letunic &
611 Bork, 2021).

612 **Structural phylogenetics**

613 We used the rigid-body superposition algorithm mTM-align (R. Dong et al., 2018) to align
614 representative cyclase crystal structures (Supplementary Table 6). To generate the character state
615 matrix, we used this multiple structure alignment to measure and encode every chemical and
616 structural character of the proteins that was both unlikely to be subject to artifacts of
617 crystallization and likely to evolve as slowly or slower than the protein's amino acid sequence:
618 reaction class, active site motif, subunits, loops and helices, and N-terminal tail. The resulting
619 character state matrix became input for Bayesian phylogeny estimation using MrBayes 3.2.6
620 (Ronquist et al., 2012). We compared these phylogenies to LDDT and TM score trees calculated

using a neighbor-joining algorithm implemented in Foldtree (Moi et al., 2023), as well as structural trees calculated using the FoldTree method and sequence trees.

Analysis of adaptation

Using the phylogeny estimated under the model EX_EHO+R5 in IQ-TREE, ancestral protein sequences were reconstructed using IQ-TREE. For each of the branches between major clades (highlighted in Figure 5), most-likely ancestral sequences at the branches' terminal nodes were modeled using ColabFold (Mirdita et al., 2022). We then used Fisher's exact test (2-tail, multiple-test corrected) to measure whether the amino acid substitutions that occurred along that branch (differences between the two ancestral proteins) are statistically clustered in functional regions of the protein. 6 functional regions were tested: the protein's N terminus, the C terminus, the helix that acts as a membrane anchor, the dimer contact surface, the tunnel from the surface to the active site, and the active site itself (precise region definitions can be found in Supplementary Table 7). Substrate tunnel residues were calculated with Caver (Pavelka et al., 2016) and active site was defined as residues within 5 Å of substrate molecule.

Data availability

The data supporting the findings in this study are available in the main text and supplementary information. The terpenoid cyclase alignment and phylogeny (shown in Figures 2, 3, and 5) are available as supplementary files.

Acknowledgments

The authors would like to thank the students of ESS 143: Molecular Geomicrobiology Lab for hard work and curiosity in initial experiments characterizing the lipid products of several metagenomic triterpenoid cyclases. The authors would like to thank members of the Welander lab, M. Horst, B. Kapili, C.D. Stark, and B. Barros-McShea for critically examining, and thereby improving, this work. We would also like to thank Stanford Research Computing for access to and assistance with the Sherlock High-Performance Computing Environment, where we performed phylogeny estimation and protein structure prediction, and the Stanford Geomicrobiology Shared Laboratories Core Facility (RRID:SCR_025000), where we performed portions of the experimental work. H.S.M. and P.V.W. were supported by NSF Grant EAR-1752564. H.S.M. was supported by the National Science Foundation (NSF) Graduate Research Fellowship, the Stanford Enhancing Diversity in Graduate Education (EDGE) Fellowship, and the Stanford School of Sustainability McGee and Levorsen Graduate Research Grant. R.A.V. was supported by the Stanford Earth Summer Undergraduate Research (SESUR) Program. B.O. and J.L.G. were supported (in part) by National Institutes of Health (NIH) Grant R15GM143714; the 800 MHz NMR was acquired through NIH S10 OD012254.

References

Abe, I. (2007). Enzymatic synthesis of cyclic triterpenes. *Natural Product Reports*, *24*(6), 1311–1331. https://doi.org/10.1039/b616857b

Abe, I., Olano, C., & Carmen, M. (2014). *Evolving Enzymes , Evolving Pathways: The Oxidosqualene Cyclases: One Substrate , Diverse Products Part 1*.

Abe, I., & Rohmer, M. (1994). Enzymic cyclization of 2,3-dihydrosqualene and squalene 2,3-epoxide by squalene cyclases: From pentacyclic to tetracyclic triterpenes. *Journal of the Chemical Society, Perkin Transactions 1*, *0*(7), 783–791. https://doi.org/10.1039/P19940000783

Alain, K., Intertaglia, L., Catala, P., & Lebaron, P. (2008). Eudoraea adriatica gen. Nov., sp. Nov., a novel marine bacterium of the family Flavobacteriaceae. *International Journal of Systematic and Evolutionary Microbiology*, *58*(10), 2275–2281. https://doi.org/10.1099/ijs.0.65446-0

Amann, E., Ochs, B., & Abel, K.-J. (1988). Tightly regulated tac promoter vectors useful for the expression of unfused and fused proteins in Escherichia coli. *Gene*, *69*(2), 301–315. https://doi.org/10.1016/0378-1119(88)90440-4

Banta, A. B., Wei, J. H., Gill, C. C. C., Giner, J.-L., & Welander, P. V. (2017). Synthesis of arborane triterpenols by a bacterial oxidosqualene cyclase. *Proceedings of the National Academy of Sciences*, *114*(2), 245–250. https://doi.org/10.1073/pnas.1617231114

Banta, A. B., Wei, J. H., & Welander, P. V. (2015). A distinct pathway for tetrahymanol synthesis in bacteria. *Proceedings of the National Academy of Sciences*, *112*(44), 13478–13483. https://doi.org/10.1073/pnas.1511482112

Belin, B. J., Busset, N., Giraud, E., Molinaro, A., Silipo, A., & Newman, D. K. (2018). Hopanoid lipids: From membranes to plant-bacteria interactions. *Nature Reviews Microbiology*, *16*(5), 304–315. https://doi.org/10.1038/nrmicro.2017.173

Bi, Y., & Yu, Z. (2016). Diterpenoids from Streptomyces sp. SN194 and Their Antifungal Activity against Botrytis cinerea. *Journal of Agricultural and Food Chemistry*, *64*(45), 8525–8529. https://doi.org/10.1021/acs.jafc.6b03645

Bligh, E. G., & Dyer, W. J. (1959). A rapid method of total lipid extraction and purification. *Canadian Journal of Biochemistry and Physiology*, *37*(8), 911–917. https://doi.org/10.1139/o59-099

Bloch, K. E. (1983). Sterol structure and membrane function. *Critical Reviews in Biochemistry and Molecular Biology*, *14*(1), 47–92. https://doi.org/10.3109/10409238309102790

Bradley, A. S., Pearson, A., Sáenz, J. P., & Marx, C. J. (2010). Adenosylhopane: The first intermediate in hopanoid side chain biosynthesis. *Organic Geochemistry*, *41*(10), 1075–1081. https://doi.org/10.1016/j.orggeochem.2010.07.003

Brenac, L., Baidoo, E. E. K., Keasling, J. D., & Budin, I. (2019). Distinct functional roles for hopanoid composition in the chemical tolerance of Zymomonas mobilis. *Molecular Microbiology*, *112*(5), 1564–1575. https://doi.org/10.1111/mmi.14380

Bridgham, J. T., Ortlund, E. A., & Thornton, J. W. (2009). An epistatic ratchet constrains the direction of glucocorticoid receptor evolution. *Nature*, *461*(7263), 515–519. https://doi.org/10.1038/nature08249

Brocks, J. J., Love, G. D., Summons, R. E., Knoll, A. H., Logan, G. A., & Bowden, S. A. (2005). Biomarker evidence for green and purple sulphur bacteria in a stratified Palaeoproterozoic sea. *Nature*, *437*(7060), 866–870. https://doi.org/10.1038/nature04068

Brocks, J. J., & Pearson, A. (2005). Building the Biomarker Tree of Life. *Reviews in Mineralogy and Geochemistry*, *59*(1), 233–258. https://doi.org/10.2138/rmg.2005.59.10

Brown, M. O., Olagunju, B. O., Giner, J.-L., & Welander, P. V. (2023). Sterol methyltransferases in uncultured bacteria complicate eukaryotic biomarker interpretations. *Nature Communications*, *14*(1), Article 1. https://doi.org/10.1038/s41467-023-37552-3

Bui, M. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference

in the Genomic Era. *Molecular Biology and Evolution*, *37*(5), 1530–1534. https://doi.org/10.1093/molbev/msaa015

Caetano-Anollés, G., Wang, M., Caetano-Anollés, D., & Mittenthal, J. E. (2009). The origin, evolution and structure of the protein world. *Biochemical Journal*, *417*(3), 621–637. https://doi.org/10.1042/BJ20082063

Cao, R., Zhang, Y., Mann, F. M., Huang, C., Mukkamala, D., Hudock, M. P., Mead, M. E., Prisic, S., Wang, K., Lin, F.-Y., Chang, T.-K., Peters, R. J., & Oldfield, E. (2010). Diterpene cyclases and the nature of the isoprene fold. *Proteins: Structure, Function, and Bioinformatics*, *78*(11), 2417–2432. https://doi.org/10.1002/prot.22751

Capella-Gutiérrez, S., Silla-Martínez, J. M., & Gabaldón, T. (2009). trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, *25*(15), 1972–1973. https://doi.org/10.1093/bioinformatics/btp348

Chandonia, J.-M., Guan, L., Lin, S., Yu, C., Fox, N. K., & Brenner, S. E. (2022). SCOPe: Improvements to the structural classification of proteins – extended database to facilitate variant interpretation and machine learning. *Nucleic Acids Research*, *50*(D1), D553–D559. https://doi.org/10.1093/nar/gkab1054

Christianson, D. W. (2017). Structural and Chemical Biology of Terpenoid Cyclases. *Chemical Reviews*, *117*(17), 11570–11648. https://doi.org/10.1021/acs.chemrev.7b00287

Conner, R. L., Mellory, F. B., Landrey, J. R., Ferguson, K. A., Kaneshiro, E. S., & Ray, E. (1971). Ergosterol replacement of tetrahymanol in Tetrahymena<math><mtext>Tetrahymena</mtext></math> membranes. *Biochemical and Biophysical Research Communications*, *44*(4), 995–1000. https://doi.org/10.1016/0006-291X(71)90810-2

Crockett, E. L. (1998). Cholesterol Function in Plasma Membranes from Ectotherms: Membrane-Specific Roles in Adaptation to Temperature. *American Zoologist*, *38*(2), 291–304. https://doi.org/10.1093/icb/38.2.291

Dairi, T., Hamano, Y., Kuzuyama, T., Itoh, N., Furihata, K., & Seto, H. (2001). Eubacterial Diterpene Cyclase Genes Essential for Production of the Isoprenoid Antibiotic Terpentecin. *Journal of Bacteriology*, *183*(20), 6085–6094. https://doi.org/10.1128/JB.183.20.6085-6094.2001

Desmond, E., & Gribaldo, S. (2010). Phylogenomics of Sterol Synthesis: Insights into the Origin, Evolution, and Diversity of a Key Eukaryotic Feature. *Genome Biology and Evolution*, *1*(0), 364–381. https://doi.org/10.1093/gbe/evp036

Dong, L.-B., Rudolf, J. D., Deng, M.-R., Yan, X., & Shen, B. (2018). Discovery of the Tiancilactone Antibiotics by Genome Mining of Atypical Bacterial Type II Diterpene Synthases. *ChemBioChem*, *19*(16), 1727–1733. https://doi.org/10.1002/cbic.201800285

Dong, R., Peng, Z., Zhang, Y., & Yang, J. (2018). mTM-align: An algorithm for fast and accurate multiple protein structure alignment. *Bioinformatics*, *34*(10), 1719–1725. https://doi.org/10.1093/bioinformatics/btx828

Dürr, C., Schnell, H.-J., Luzhetskyy, A., Murillo, R., Weber, M., Welzel, K., Vente, A., & Bechthold, A. (2006). Biosynthesis of the terpene phenalinolactone in Streptomyces sp. Tü6071: Analysis of the gene cluster and generation of derivatives. *Chemistry & Biology*, *13*(4), 365–377. https://doi.org/10.1016/j.chembiol.2006.01.011

Eschenmoser, A., & Arigoni, D. (2005). Revisited after 50 Years: The 'Stereochemical Interpretation of the Biogenetic Isoprene Rule for the Triterpenes.' *Helvetica Chimica Acta*, *88*(12), 3011–3050. https://doi.org/10.1002/hlca.200590245

Fischer, W. W. (2008). Life before the rise of oxygen. *Nature*, *455*(7216), Article 7216. https://doi.org/10.1038/4551051a

Fischer, W. W., & Pearson, A. (2007). Hypotheses for the origin and early evolution of triterpenoid cyclases. *Geobiology*, *5*(1), 19–34. https://doi.org/10.1111/j.1472-4669.2007.00096.x

Flesch, G., & Rohmer, M. (1987). Growth inhibition of hopanoid synthesizing bacteria by squalene cyclase inhibitors. *Archives of Microbiology*, *147*(1), 100–104. https://doi.org/10.1007/BF00492912

Förster, H. J., Biemann, K., Haigh, W. G., Tattrie, N. H., & Colvin, J. R. (1973). The structure of novel C35 pentacyclic terpenes from Acetobacter xylinum. *Biochemical Journal*, *135*(1), 133–143. https://doi.org/10.1042/bj1350133

Fox, N. K., Brenner, S. E., & Chandonia, J.-M. (2014). SCOPe: Structural Classification of Proteins—extended, integrating SCOP and ASTRAL data and classification of new structures. *Nucleic Acids Research*, *42*(D1), D304–D309. https://doi.org/10.1093/nar/gkt1240

French, K. L., Hallmann, C., Hope, J. M., Schoon, P. L., Zumberge, J. A., Hoshino, Y., Peters, C. A., George, S. C., Love, G. D., Brocks, J. J., Buick, R., & Summons, R. E. (2015). Reappraisal of hydrocarbon biomarkers in Archean rocks. *Proceedings of the National Academy of Sciences*, *112*(19), 5915–5920. https://doi.org/10.1073/pnas.1419563112

Frickey, T., & Kannenberg, E. (2009). Phylogenetic analysis of the triterpene cyclase protein family in prokaryotes and eukaryotes suggests bidirectional lateral gene transfer. *Environmental Microbiology*, *11*(5), 1224–1241. https://doi.org/10.1111/j.1462-2920.2008.01851.x

Fu, L., Niu, B., Zhu, Z., Wu, S., & Li, W. (2012). CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics*, *28*(23), 3150–3152. https://doi.org/10.1093/bioinformatics/bts565

Garg, S. G., & Hochberg, G. K. A. (2025). A General Substitution Matrix for Structural Phylogenetics. *Molecular Biology and Evolution*, *42*(6), msaf124. https://doi.org/10.1093/molbev/msaf124

Gebhardt, K., Meyer, S. W., Schinko, J., Bringmann, G., Zeeck, A., & Fiedler, H.-P. (2011). Phenalinolactones A–D, terpenoglycoside antibiotics from Streptomyces sp. Tü 6071. *The Journal of Antibiotics*, *64*(3), Article 3. https://doi.org/10.1038/ja.2010.165

Gharib, W. H., & Robinson-Rechavi, M. (2013). The Branch-Site Test of Positive Selection Is Surprisingly Robust but Lacks Power under Synonymous Substitution Saturation and Variation in GC. *Molecular Biology and Evolution*, *30*(7), 1675–1686. https://doi.org/10.1093/molbev/mst062

Gold, D. A., Caron, A., Fournier, G. P., & Summons, R. E. (2017). Paleoproterozoic sterol biosynthesis and the rise of oxygen. *Nature*, *543*(7645), 420–423. https://doi.org/10.1038/nature21412

Gudde, L. R., Hulce, M., Largen, A. H., & Franke, J. D. (2019). Sterol synthesis is essential for viability in the planctomycete bacterium Gemmata obscuriglobus. *FEMS Microbiology Letters*, *366*(3). https://doi.org/10.1093/femsle/fnz019

Gumsley, A. P., Chamberlain, K. R., Bleeker, W., Söderlund, U., de Kock, M. O., Larsson, E. R., & Bekker, A. (2017). Timing and tempo of the Great Oxidation Event. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(8), 1811–1816. https://doi.org/10.1073/pnas.1608824114

Hammer, S. C., Syrén, P.-O., Seitz, M., Nestl, B. M., & Hauer, B. (2013). Squalene hopene cyclases: Highly promiscuous and evolvable catalysts for stereoselective CC and CX bond formation. *Current Opinion in Chemical Biology*, *17*(2), 293–300. https://doi.org/10.1016/j.cbpa.2013.01.016

Hauke, V., Adam, P., Trendel, J. M., Albrecht, P., Schwark, L., Vliex, M., Hagemann, H., & Püttmann, W. (1995). Isoarborinol through geological times: Evidence for its presence in the Permian and Triassic. *Organic Geochemistry*, *23*(1), 91–93. https://doi.org/10.1016/0146-6380(95)00002-V

Hayashi, Y., Matsuura, N., Toshima, H., Itoh, N., Ishikawa, J., Mikami, Y., & Dairi, T. (2008). Cloning of the Gene Cluster Responsible for the Biosynthesis of Brasilicardin A, a Unique Diterpenoid. *The Journal of Antibiotics*, *61*(3), Article 3. https://doi.org/10.1038/ja.2008.126

Hayashi, Y., Onaka, H., Itoh, N., Seto, H., & Dairi, T. (2007). Cloning of the Gene Cluster Responsible for Biosynthesis of KS-505a (Longestin), a Unique Tetraterpenoid. *Bioscience, Biotechnology, and Biochemistry*, *71*(12), 3072–3081. https://doi.org/10.1271/bbb.70477

Hess, B. A. (2002). Concomitant C-Ring Expansion and D-Ring Formation in Lanosterol Biosynthesis from Squalene without Violation of Markovnikov's Rule. *Journal of the American Chemical Society*, *124*(35), 10286–10287. https://doi.org/10.1021/ja026850r

Hess, B. A., & Smentek, L. (2013). The Concerted Nature of the Cyclization of Squalene Oxide to the Protosterol Cation. *Angewandte Chemie International Edition*, *52*(42), 11029–11033. https://doi.org/10.1002/anie.201302886

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution*, *35*(2), 518–522. https://doi.org/10.1093/molbev/msx281

Hoshino, T., & Sato, T. (2002). Squalene – hopene cyclase: Catalytic mechanism and substrate recognition. *Chemical Communications*, *0*(4), 291–301. https://doi.org/10.1039/B108995C

Hoshino, Y., & Villanueva, L. (2023). Four billion years of microbial terpenome evolution. *FEMS Microbiology Reviews*, *47*(2), fuad008. https://doi.org/10.1093/femsre/fuad008

Huson, D. H., & Scornavacca, C. (2012). Dendroscope 3: An Interactive Tool for Rooted Phylogenetic Trees and Networks. *Systematic Biology*, *61*(6), 1061–1067. https://doi.org/10.1093/sysbio/sys062

Ikeda, C., Hayashi, Y., Itoh, N., Seto, H., & Dairi, T. (2007). Functional Analysis of Eubacterial ent - Copalyl Diphosphate Synthase and Pimara-9(11),15-Diene Synthase with Unique Primary Sequences. *The Journal of Biochemistry*, *141*(1), 37–45. https://doi.org/10.1093/jb/mvm004

Ingles-Prieto, A., Ibarra-Molero, B., Delgado-Delgado, A., Perez-Jimenez, R., Fernandez, J. M., Gaucher, E. A., Sanchez-Ruiz, J. M., & Gavira, J. A. (2013). Conservation of Protein Structure over Four Billion Years. *Structure*, *21*(9), 1690–1697. https://doi.org/10.1016/j.str.2013.06.020

Jahnke, L. L. (1986). The effects of low oxygen on the synthesis of unsaturated fatty acids and sterols: Implications for the evolution of eukaryotes. *Origins of Life and Evolution of the Biosphere*, *16*(3), 317–318. https://doi.org/10.1007/BF02422046

Jahnke, L. L., & Nichols, P. D. (1986). Methyl sterol and cyclopropane fatty acid composition of Methylococcus capsulatus grown at low oxygen tensions. *Journal of Bacteriology*, *167*(1), 238–242. https://doi.org/10.1128/jb.167.1.238-242.1986

Jayasuriya, H., Herath, K. B., Zhang, C., Zink, D. L., Basilio, A., Genilloud, O., Diez, M. T., Vicente, F., Gonzalez, I., Salazar, O., Pelaez, F., Cummings, R., Ha, S., Wang, J., & Singh, S. B. (2007). Isolation and Structure of Platencin: A FabH and FabF Dual Inhibitor with Potent Broad-Spectrum Antibiotic Activity. *Angewandte Chemie International Edition*, *46*(25), 4684–4688. https://doi.org/10.1002/anie.200701058

Jia, Q., Brown, R., Köllner, T. G., Fu, J., Chen, X., Wong, G. K.-S., Gershenzon, J., Peters, R. J., & Chen, F. (2022). Origin and early evolution of the plant terpene synthase family. *Proceedings of the National Academy of Sciences*, *119*(15), e2100361119. https://doi.org/10.1073/pnas.2100361119

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*(6), Article 6. https://doi.org/10.1038/nmeth.4285

Kennard, O., Riva di Sanseverino, L., Vorbrűggen, H., & Djerassi, C. (1965). The complete structure of the triterpene arborinol. *Tetrahedron Letters*, *6*(39), 3433–3438. https://doi.org/10.1016/S0040-4039(01)89324-2

Khan, S. R., Gaines, J., Roop, R. M., & Farrand, S. K. (2008). Broad-Host-Range Expression Vectors with Tightly Regulated Promoters and Their Use To Examine the Influence of TraR and TraM Expression on Ti Plasmid Quorum Sensing. *Applied and Environmental Microbiology*, *74*(16), 5053–5062. https://doi.org/10.1128/AEM.01098-08

Kim, S.-H., Lu, W., Ahmadi, M. K., Montiel, D., Ternei, M. A., & Brady, S. F. (2019). Atolypenes, Tricyclic Bacterial Sesterterpenes Discovered Using a Multiplexed In Vitro Cas9-TAR Gene Cluster Refactoring Approach. *ACS Synthetic Biology*, *8*(1), 109–118. https://doi.org/10.1021/acssynbio.8b00361

Kontnik, R., Bosak, T., Butcher, R. A., Brocks, J. J., Losick, R., Clardy, J., & Pearson, A. (2008). Sporulenes, heptaprenyl metabolites from Bacillus subtilis spores. *Organic Letters*, *10*(16), 3551–3554. https://doi.org/10.1021/ol801314k

Le, S. Q., & Gascuel, O. (2010). Accounting for Solvent Accessibility and Secondary Structure in Protein Phylogenetics Is Clearly Beneficial. *Systematic Biology*, *59*(3), 277–287. https://doi.org/10.1093/sysbio/syq002

Lee, A. K., Banta, A. B., Wei, J. H., Kiemle, D. J., Feng, J., Giner, J.-L., & Welander, P. V. (2018). C-4 sterol demethylation enzymes distinguish bacterial and eukaryotic sterol synthesis. *Proceedings of the National Academy of Sciences*, *115*(23), 5884–5889. https://doi.org/10.1073/pnas.1802930115

Lee, A. K., Wei, J. H., & Welander, P. V. (2023). De novo cholesterol biosynthesis in bacteria. *Nature Communications*, *14*(1), Article 1. https://doi.org/10.1038/s41467-023-38638-8

Letunic, I., & Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Research*, *49*(W1), W293–W296. https://doi.org/10.1093/nar/gkab301

Levental, I., Levental, K. R., & Heberle, F. A. (2020). Lipid Rafts: Controversies Resolved, Mysteries Remain. *Trends in Cell Biology*, *30*(5), 341–353. https://doi.org/10.1016/j.tcb.2020.01.009

López, D., & Kolter, R. (2010). Functional microdomains in bacterial membranes. *Genes & Development*, *24*(17), 893–902. https://doi.org/10.1101/gad.1945010

Luo, G., Ono, S., Beukes, N. J., Wang, D. T., Xie, S., & Summons, R. E. (2016). Rapid oxygenation of Earth's atmosphere 2.33 billion years ago. *Science Advances*, *2*(5), e1600134. https://doi.org/10.1126/sciadv.1600134

Maddamsetti, R., & Grant, N. A. (2022). Discovery of positive and purifying selection in metagenomic time series of hypermutator microbial populations. *PLOS Genetics*, *18*(8), e1010324. https://doi.org/10.1371/journal.pgen.1010324

Malik, A. J., Poole, A. M., & Allison, J. R. (2020). Structural Phylogenetics with Confidence. *Molecular Biology and Evolution*, *37*(9), 2711–2726. https://doi.org/10.1093/molbev/msaa100

Mann, F. M., Xu, M., Chen, X., Fulton, D. B., Russell, D. G., & Peters, R. J. (2009). Edaxadiene: A New Bioactive Diterpene from Mycobacterium tuberculosis. *Journal of the American Chemical Society*, *131*(48), 17526–17527. https://doi.org/10.1021/ja9019287

Martinez-Gutierrez, C. A., Uyeda, J. C., & Aylward, F. O. (2023). A timeline of bacterial and archaeal diversification in the ocean. *eLife*, *12*, RP88268. https://doi.org/10.7554/eLife.88268

Mayer, M. H., Parenteau, M. N., Kempher, M. L., Madigan, M. T., Jahnke, L. L., & Welander, P. V. (2021). Anaerobic 3-methylhopanoid production by an acidophilic photosynthetic purple bacterium. *Archives of Microbiology*, *203*(10), 6041–6052. https://doi.org/10.1007/s00203-021-02561-7

McDonald, J. H., & Kreitman, M. (1991). Adaptive protein evolution at the Adh locus in Drosophila. *Nature*, *351*(6328), Article 6328. https://doi.org/10.1038/351652a0

McShea, H., Anda, V. D., Brocks, J. J., Welander, P. V., & Baker, B. J. (2025). *Archaeal lineages related to eukaryotes encode functional diterpenoid cyclases* (p. 2025.02.07.637177). bioRxiv. https://doi.org/10.1101/2025.02.07.637177

Messer, P. W., & Petrov, D. A. (2012). *The McDonald-Kreitman Test and its Extensions under Frequent Adaptation: Problems and Solutions*. *2013*. https://doi.org/10.1073/pnas.1220835110

Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., & Steinegger, M. (2022). ColabFold: Making protein folding accessible to all. *Nature Methods*, *19*(6), Article 6. https://doi.org/10.1038/s41592-022-01488-1

Mitsuhashi, T., Okada, M., & Abe, I. (2017). Identification of Chimeric αβγ Diterpene Synthases Possessing both Type II Terpene Cyclase and Prenyltransferase Activities. *ChemBioChem*, *18*(21), 2104–2109. https://doi.org/10.1002/cbic.201700445

Moi, D., Bernard, C., Steinegger, M., Nevers, Y., Langleib, M., & Dessimoz, C. (2023). *Structural phylogenetics unravels the evolutionary diversification of communication systems in gram-*

positive bacteria and their viruses (p. 2023.09.19.558401). bioRxiv. https://doi.org/10.1101/2023.09.19.558401

Moosmann, P., Ecker, F., Leopold-Messer, S., Cahn, J. K. B., Dieterich, C. L., Groll, M., & Piel, J. (2020). A monodomain class II terpene cyclase assembles complex isoprenoid scaffolds. *Nature Chemistry*, *12*(10), Article 10. https://doi.org/10.1038/s41557-020-0515-3

Morrone, D., Chambers, J., Lowry, L., Kim, G., Anterola, A., Bender, K., & Peters, R. J. (2009). Gibberellin biosynthesis in bacteria: Separate ent-copalyl diphosphate and ent-kaurene synthases in Bradyrhizobium japonicum. *FEBS Letters*, *583*(2), 475–480. https://doi.org/10.1016/j.febslet.2008.12.052

Mouritsen, O. G., & Zuckermann, M. J. (2004). What's so special about cholesterol? *Lipids*, *39*(11), 1101–1113. https://doi.org/10.1007/s11745-004-1336-x

Mutti, G., Ocaña-Pallarès, E., & Gabaldón, T. (2025). Newly Developed Structure-Based Methods Do Not Outperform Standard Sequence-Based Methods for Large-Scale Phylogenomics. *Molecular Biology and Evolution*, *42*(7), msaf149. https://doi.org/10.1093/molbev/msaf149

Nakano, C., Okamura, T., Sato, T., Dairi, T., & Hoshino, T. (2005). Mycobacterium tuberculosis H37Rv3377c encodes the diterpene cyclase for producing the halimane skeleton. *Chemical Communications*, *8*, 1016–1018. https://doi.org/10.1039/B415346D

Nakano, C., Oshima, M., Kurashima, N., & Hoshino, T. (2015). Identification of a New Diterpene Biosynthetic Gene Cluster that Produces O-Methylkolavelool in Herpetosiphon aurantiacus. *ChemBioChem*, *16*(5), 772–781. https://doi.org/10.1002/cbic.201402652

Nett, R. S., Bender, K. S., & Peters, R. J. (2022). Production of the plant hormone gibberellin by rhizobia increases host legume nodule size. *The ISME Journal*, *16*(7), Article 7. https://doi.org/10.1038/s41396-022-01236-5

Oldfield, E., & Lin, F.-Y. (2012). Terpene Biosynthesis: Modularity Rules. *Angewandte Chemie International Edition*, *51*(5), 1124–1137. https://doi.org/10.1002/anie.201103110

Ostrander, C. M., Heard, A. W., Shu, Y., Bekker, A., Poulton, S. W., Olesen, K. P., & Nielsen, S. G. (2024). Onset of coupled atmosphere–ocean oxygenation 2.3 billion years ago. *Nature*, *631*(8020), 335–339. https://doi.org/10.1038/s41586-024-07551-5

Ourisson, G. (1989). The evolution of terpenes to sterols. *Pure and Applied Chemistry*, *61*(3), 345–348. https://doi.org/10.1351/pac198961030345

Ourisson, G., & Albrecht, P. (1992). Hopanoids. 1. Geohopanoids: The most abundant natural products on Earth? *Accounts of Chemical Research*, *25*(9), 398–402. https://doi.org/10.1021/ar00021a003

Ourisson, G., Albrecht, P., & Rohmer, M. (1982). Predictive microbial biochemistry—From molecular fossils to procaryotic membranes. *Trends in Biochemical Sciences*, *7*(7), 236–239. https://doi.org/10.1016/0968-0004(82)90028-7

Ourisson, G., & Nakatani, Y. (1994). The terpenoid theory of the origin of cellular life: The evolution of terpenoids to cholesterol. *Chemistry and Biology*, *1*(1), 11–23. https://doi.org/10.1016/1074-5521(94)90036-1

Ourisson, G., & Rohmer, M. (1992). Hopanoids. 2. Biohopanoids: A novel class of bacterial lipids. *Accounts of Chemical Research*, *25*(9), 403–408. https://doi.org/10.1021/ar00021a004

Ozaki, T., Shinde, S. S., Gao, L., Okuizumi, R., Liu, C., Ogasawara, Y., Lei, X., Dairi, T., Minami, A., & Oikawa, H. (2018). Enzymatic Formation of a Skipped Methyl-Substituted Octaprenyl Side Chain of Longestin (KS-505a): Involvement of Homo-IPP as a Common Extender Unit. *Angewandte Chemie International Edition*, *57*(22), 6629–6632. https://doi.org/10.1002/anie.201802116

Pan, J.-J., Solbiati, J. O., Ramamoorthy, G., Hillerich, B. S., Seidel, R. D., Cronan, J. E., Almo, S. C., & Poulter, C. D. (2015). Biosynthesis of Squalene from Farnesyl Diphosphate in Bacteria: Three Steps Catalyzed by Three Enzymes. *ACS Central Science*, *1*(2), 77–82. https://doi.org/10.1021/acscentsci.5b00115

Pan, X., Du, W., Zhang, X., Lin, X., Li, F.-R., Yang, Q., Wang, H., Rudolf, J. D., Zhang, B., & Dong, L.-B. (2022). Discovery, Structure, and Mechanism of a Class II Sesquiterpene Cyclase. *Journal of the American Chemical Society*, *144*(48), 22067–22074. https://doi.org/10.1021/jacs.2c09412

Pavelka, A., Sebestova, E., Kozlikova, B., Brezovsky, J., Sochor, J., & Damborsky, J. (2016). CAVER: Algorithms for Analyzing Dynamics of Tunnels in Macromolecules. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *13*(3), 505–517. IEEE/ACM Transactions on Computational Biology and Bioinformatics. https://doi.org/10.1109/TCBB.2015.2459680

Pearson, A., Budin, M., & Brocks, J. J. (2003). Phylogenetic and biochemical evidence for sterol synthesis in the bacterium Gemmata obscuriglobus. *Proceedings of the National Academy of Sciences*, *100*(26), 15352–15357. https://doi.org/10.1073/pnas.2536559100

Pearson, A., Flood Page, S. R., Jorgenson, T. L., Fischer, W. W., & Higgins, M. B. (2007). Novel hopanoid cyclases from the environment. *Environmental Microbiology*, *9*(9), 2175–2188. https://doi.org/10.1111/j.1462-2920.2007.01331.x

Pearson, A., & Rusch, D. B. (2009). Distribution of microbial terpenoid lipid cyclases in the global ocean metagenome. *ISME Journal*, *3*(3), 352–363. https://doi.org/10.1038/ismej.2008.116

Peralta-Yahya, P. P., Ouellet, M., Chan, R., Mukhopadhyay, A., Keasling, J. D., & Lee, T. S. (2011). Identification and microbial production of a terpene-based advanced biofuel. *Nature Communications*, *2*(1), Article 1. https://doi.org/10.1038/ncomms1494

Pollier, J., Vancaester, E., Kuzhiumparambil, U., Vickers, C. E., Vandepoele, K., Goossens, A., & Fabris, M. (2019). A widespread alternative squalene epoxidase participates in eukaryote steroid biosynthesis. *Nature Microbiology*, *4*(2), 226–233. https://doi.org/10.1038/s41564-018-0305-5

Pronk, L. J. U., & Medema, M. H. (2022). Whokaryote: Distinguishing eukaryotic and prokaryotic contigs in metagenomes based on gene structure. *Microbial Genomics*, *8*(5), 000823. https://doi.org/10.1099/mgen.0.000823

Quin, M. B., Flynn, C. M., & Schmidt-Dannert, C. (2014). Traversing the fungal terpenome. *Natural Product Reports*, *31*(10), 1449–1473. https://doi.org/10.1039/C4NP00075G

Rambaut, A. (2009). *FigTree, version 1.4.3*. http://tree.bio.ed.ac.uk/software/Figtree/

Ricci, J. N., Coleman, M. L., Welander, P. V., Sessions, A. L., Summons, R. E., Spear, J. R., & Newman, D. K. (2014). Diverse capacity for 2-methylhopanoid production correlates with a specific ecological niche. *The ISME Journal*, *8*(3), Article 3. https://doi.org/10.1038/ismej.2013.191

Ritchie, A. M., Stark, T. L., & Liberles, D. A. (2021). Inferring the number and position of changes in selective regime in a non-equilibrium mutation-selection framework. *BMC Ecology and Evolution*, *21*(1), 39. https://doi.org/10.1186/s12862-021-01770-4

Rivas-Marin, E., Stettner, S., Gottshall, E. Y., Santana-Molina, C., Helling, M., Basile, F., Ward, N. L., & Devos, D. P. (2019). Essentiality of sterol synthesis genes in the planctomycete bacterium Gemmata obscuriglobus. *Nature Communications*, *10*(1), 2916. https://doi.org/10.1038/s41467-019-10983-7

Rohmer, M., Anding, C., & Ourisson, G. (1980). Non-specific biosynthesis of hopane triterpenes by a cell-free system from Acetobacter pasteurianum. *European Journal of Biochemistry*, *112*(3), 541–547. https://doi.org/10.1111/j.1432-1033.1980.tb06117.x

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M. A., & Huelsenbeck, J. P. (2012). MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Systematic Biology*, *61*(3), 539–542. https://doi.org/10.1093/sysbio/sys029

Rozewicki, J., Li, S., Amada, K. M., Standley, D. M., & Katoh, K. (2019). MAFFT-DASH: Integrated protein sequence and structural alignment. *Nucleic Acids Research*, *47*(W1), W5–W10. https://doi.org/10.1093/nar/gkz342

Rudolf, J. D., Alsup, T. A., Xu, B., & Li, Z. (2021). Bacterial terpenome. *Natural Product Reports*, *38*(5), 905–980. https://doi.org/10.1039/D0NP00066C

Runnegar, B. (1991). Precambrian oxygen levels estimated from the biochemistry and physiology of early eukaryotes. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *97*(1), 97–111. https://doi.org/10.1016/0031-0182(91)90186-U

Sáenz, J. P. (2010). Hopanoid enrichment in a detergent resistant membrane fraction of Crocosphaera watsonii: Implications for bacterial lipid raft formation. *Organic Geochemistry*, *41*(8), 853–856. https://doi.org/10.1016/j.orggeochem.2010.05.005

Sáenz, J. P., Grosser, D., Bradley, A. S., Lagny, T. J., Lavrynenko, O., Broda, M., & Simons, K. (2015). Hopanoids as functional analogues of cholesterol in bacterial membranes. *Proceedings of the National Academy of Sciences*, *112*(38), 11971–11976. https://doi.org/10.1073/pnas.1515607112

Sáenz, J. P., Sezgin, E., Schwille, P., & Simons, K. (2012). Functional convergence of hopanoids and sterols in membrane ordering. *Proceedings of the National Academy of Sciences*, *109*(35), 14236–14240. https://doi.org/10.1073/pnas.1212141109

Santana-Molina, C., Rivas-Marin, E., Rojas, A. M., & Devos, D. P. (2020). Origin and Evolution of Polycyclic Triterpene Synthesis. *Molecular Biology and Evolution*, *37*(7), 1925–1941. https://doi.org/10.1093/molbev/msaa054

Sato, T., Hoshino, H., Yoshida, S., Nakajima, M., & Hoshino, T. (2011). Bifunctional triterpene/sesquarterpene cyclase: Tetraprenyl-β- curcumene cyclase is also squalene cyclase in bacillus megaterium. *Journal of the American Chemical Society*, *133*(44), 17540–17543. https://doi.org/10.1021/ja2060319

Sato, T., Yoshida, S., Hoshino, H., Tanno, M., Nakajima, M., & Hoshino, T. (2011). Sesquarterpenes (C35 Terpenes) Biosynthesized via the Cyclization of a Linear C35 Isoprenoid by a Tetraprenyl-β-curcumene Synthase and a Tetraprenyl-β-curcumene Cyclase: Identification of a New Terpene Cyclase. *Journal of the American Chemical Society*, *133*(25), 9734–9737. https://doi.org/10.1021/ja203779h

Schaeffer, P., Poinsot, J., Hauke, V., Adam, P., Wehrung, P., Trendel, J.-M., Albrecht, P., Dessort, D., & Connan, J. (1994). Novel Optically Active Hydrocarbons in Sediments: Evidence for an Extensive Biological Cyclization of Higher Regular Polyprenols. *Angewandte Chemie International Edition in English*, *33*(11), 1166–1169. https://doi.org/10.1002/anie.199411661

Schmerk, C. L., Welander, P. V., Hamad, M. A., Bain, K. L., Bernards, M. A., Summons, R. E., & Valvano, M. A. (2015). Elucidation of the Burkholderia cenocepacia hopanoid biosynthesis pathway uncovers functions for conserved proteins in hopanoid-producing bacteria. *Environmental Microbiology*, *17*(3), 735–750. https://doi.org/10.1111/1462-2920.12509

Seitz, M., Klebensberger, J., Siebenhaller, S., Breuer, M., Siedenburg, G., Jendrossek, D., & Hauer, B. (2012). Substrate specificity of a novel squalene-hopene cyclase from Zymomonas mobilis. *Journal of Molecular Catalysis B: Enzymatic*, *84*, 72–77. https://doi.org/10.1016/j.molcatb.2012.02.007

Shigemori, H., Komaki, H., Yazawa, K., Mikami, Y., Nemoto, A., Tanaka, Y., Sasaki, T., In, Y., Ishida, T., & Kobayashi, J. (1998). Brasilicardin A. A Novel Tricyclic Metabolite with Potent Immunosuppressive Activity from Actinomycete Nocardia brasiliensis. *The Journal of Organic Chemistry*, *63*(20), 6900–6904. https://doi.org/10.1021/jo9807114

Smanski, M. J., Yu, Z., Casper, J., Lin, S., Peterson, R. M., Chen, Y., Wendt-Pienkowski, E., Rajski, S. R., & Shen, B. (2011). Dedicated ent-kaurene and ent-atiserene synthases for platensimycin and platencin biosynthesis. *Proceedings of the National Academy of Sciences*, *108*(33), 13498–13503. https://doi.org/10.1073/pnas.1106919108

Smirnov, V., & Warnow, T. (2021). MAGUS: Multiple sequence Alignment using Graph clUStering. *Bioinformatics*, *37*(12), 1666–1672. https://doi.org/10.1093/bioinformatics/btaa992

Stowell, E. A., Ehrenberger, M. A., Lin, Y.-L., Chang, C.-Y., & Rudolf, J. D. (2022). Structure-guided product determination of the bacterial type II diterpene synthase Tpn2. *Communications Chemistry*, *5*(1), Article 1. https://doi.org/10.1038/s42004-022-00765-6

Summons, R. E., Bird, L. R., Gillespie, A. L., Pruss, S. B., Roberts, M., & Sessions, A. L. (2013). Lipid biomarkers in ooids from different locations and ages: Evidence for a common bacterial flora. *Geobiology*, *11*(5), 420–436. https://doi.org/10.1111/gbi.12047

Syrén, P.-O., Henche, S., Eichler, A., Nestl, B. M., & Hauer, B. (2016). Squalene-hopene cyclases—Evolution, dynamics and catalytic scope. *Current Opinion in Structural Biology*, *41*, 73–82. https://doi.org/10.1016/j.sbi.2016.05.019

Takishita, K., Chikaraishi, Y., Tanifuji, G., Ohkouchi, N., Hashimoto, T., Fujikura, K., & Roger, A. J. (2017). Microbial Eukaryotes that Lack Sterols. *Journal of Eukaryotic Microbiology*, *16*(1), 1–8. https://doi.org/10.1111/jeu.12426

Tamamura, T., Sawa, T., Isshiki, K., Masuda, T., Homma, Y., Iinuma, H., Naganawa, H., Hamada, M., Takeuchi, T., & Umezawa, H. (1985). ISOLATION AND CHARACTERIZATION OF TERPENTECIN, A NEW ANTITUMOR ANTIBIOTIC. *The Journal of Antibiotics*, *38*(12), 1664–1669. https://doi.org/10.7164/antibiotics.38.1664

Tantillo, D. J. (2010). The carbocation continuum in terpene biosynthesis—Where are the secondary cations? *Chemical Society Reviews*, *39*(8), 2847–2854. https://doi.org/10.1039/B917107J

Usui, T., Nagumo, Y., Watanabe, A., Kubota, T., Komatsu, K., Kobayashi, J., & Osada, H. (2006). Brasilicardin A, a Natural Immunosuppressant, Targets Amino Acid Transport System L. *Chemistry & Biology*, *13*(11), 1153–1160. https://doi.org/10.1016/j.chembiol.2006.09.006

Villanueva, L., Rijpstra, W. I. C., Schouten, S., & Damsté, J. S. S. (2014). Genetic biomarkers of the sterol-biosynthetic pathway in microalgae. *Environmental Microbiology Reports*, *6*(1), 35–44. https://doi.org/10.1111/1758-2229.12106

Vo, N. N. Q., Nomura, Y., Kinugasa, K., Takagi, H., & Takahashi, S. (2022). Identification and Characterization of Bifunctional Drimenol Synthases of Marine Bacterial Origin. *ACS Chemical Biology*, *17*(5), 1226–1238. https://doi.org/10.1021/acschembio.2c00163

Vorbrüggen, H., Pakrashi, S. C., & Djerassi, C. (1963). Terpenoide, LIV. Arborinol, ein neuer Triterpen-Typus. *Justus Liebigs Annalen der Chemie*, *668*(1), 57–76. https://doi.org/10.1002/jlac.19636680107

Wang, J., Kodali, S., Lee, S. H., Galgoci, A., Painter, R., Dorso, K., Racine, F., Motyl, M., Hernandez, L., Tinney, E., Colletti, S. L., Herath, K., Cummings, R., Salazar, O., González, I., Basilio, A., Vicente, F., Genilloud, O., Pelaez, F., … Singh, S. B. (2007). Discovery of platencin, a dual FabF and FabH inhibitor with in vivo antibiotic properties. *Proceedings of the National Academy of Sciences*, *104*(18), 7612–7616. https://doi.org/10.1073/pnas.0700746104

Wang, J., Soisson, S. M., Young, K., Shoop, W., Kodali, S., Galgoci, A., Painter, R., Parthasarathy, G., Tang, Y. S., Cummings, R., Ha, S., Dorso, K., Motyl, M., Jayasuriya, H., Ondeyka, J., Herath, K., Zhang, C., Hernandez, L., Allocco, J., … Singh, S. B. (2006). Platensimycin is a selective FabF inhibitor with potent antibiotic properties. *Nature*, *441*(7091), Article 7091. https://doi.org/10.1038/nature04784

Wang, S., & Luo, H. (2025). Dating the Bacterial Tree of Life Based on Ancient Symbiosis. *Systematic Biology*, syae071. https://doi.org/10.1093/sysbio/syae071

Wei, J. H., Yin, X., & Welander, P. V. (2016). Sterol synthesis in diverse bacteria. *Frontiers in Microbiology*, *7*(JUN), 1–19. https://doi.org/10.3389/fmicb.2016.00990

Welander, P. V., Coleman, M. L., Sessions, A. L., Summons, R. E., & Newman, D. K. (2010). Identification of a methylase required for 2-methylhopanoid production and implications for the interpretation of sedimentary hopanes. *Proceedings of the National Academy of Sciences*, *107*(19), 8537–8542. https://doi.org/10.1073/pnas.0912949107

Welander, P. V., Doughty, D. M., Wu, C.-H., Mehay, S., Summons, R. E., & Newman, D. K. (2012). Identification and characterization of Rhodopseudomonas palustris TIE-1 hopanoid biosynthesis mutants: Hopanoid biosynthesis in R. palustris TIE-1. *Geobiology*, *10*(2), 163–177. https://doi.org/10.1111/j.1472-4669.2011.00314.x

Welander, P. V., Hunter, R. C., Zhang, L., Sessions, A. L., Summons, R. E., & Newman, D. K. (2009). *Hopanoids Play a Role in Membrane Integrity and pH Homeostasis in Rhodopseudomonas palustris TIE-1 ♋†*. *191*(19), 6145–6156. https://doi.org/10.1128/JB.00460-09

Welander, P. V., & Summons, R. E. (2012). Discovery, taxonomic distribution, and phenotypic characterization of a gene required for 3-methylhopanoid production. *Proceedings of the National Academy of Sciences*, *109*(32), 12905–12910. https://doi.org/10.1073/pnas.1208255109

Wendt, K. U. (2005). Enzyme Mechanisms for Triterpene Cyclization: New Pieces of the Puzzle. *Angewandte Chemie International Edition*, *44*(26), 3966–3971. https://doi.org/10.1002/anie.200500804

Wendt, K. U., Lenhart, A., & Schulz, G. E. (1999). The structure of the membrane protein squalene-hopene cyclase at 2.0 Å resolution. *Journal of Molecular Biology*, *286*(1), 175–187. https://doi.org/10.1006/jmbi.1998.2470

Xia, X., Xie, Z., Salemi, M., Chen, L., & Wang, Y. (2003). An index of substitution saturation and its application. *Molecular Phylogenetics and Evolution*, *26*(1), 1–7. https://doi.org/10.1016/S1055-7903(02)00326-3

Xu, R., Fazio, G. C., & Matsuda, S. P. T. (2004). On the origins of triterpenoid skeletal diversity. *Phytochemistry*, *65*(3), 261–291. https://doi.org/10.1016/j.phytochem.2003.11.014

Xu, X., & London, E. (2000). The Effect of Sterol Structure on Membrane Lipid Domains Reveals How Cholesterol Can Induce Lipid Domain Formation. *Biochemistry*, *39*(5), 843–849. https://doi.org/10.1021/bi992543v

Yang, Z., & Dos Reis, M. (2011). Statistical properties of the branch-site test of positive selection. *Molecular Biology and Evolution*, *28*(3), 1217–1228. https://doi.org/10.1093/molbev/msq303

Young, D. C., Layre, E., Pan, S.-J., Tapley, A., Adamson, J., Seshadri, C., Wu, Z., Buter, J., Minnaard, A. J., Coscolla, M., Gagneux, S., Copin, R., Ernst, J. D., Bishai, W. R., Snider, B. B., & Moody, D. B. (2015). In Vivo Biosynthesis of Terpene Nucleosides Provides Unique Chemical Markers of Mycobacterium tuberculosis Infection. *Chemistry & Biology*, *22*(4), 516–526. https://doi.org/10.1016/j.chembiol.2015.03.015

Zhai, L., Bonds, A. C., Smith, C. A., Oo, H., Chou, J. C.-C., Welander, P. V., & Dassama, L. M. (2024). Novel sterol binding domains in bacteria. *eLife*, *12*, RP90696. https://doi.org/10.7554/eLife.90696

Zhang, Q., Yu, Y., Vélasquez, J. E., & van der Donk, W. A. (2012). Evolution of lanthipeptide synthetases. *Proceedings of the National Academy of Sciences*, *109*(45), 18361–18366. https://doi.org/10.1073/pnas.1210393109