FEMS Microbiology Reviews, 1–39
DOI: 10.1093/femsre/fuad008
Advance access publication date: 20 March 2023
Review Article

Four billion years of microbial terpenome evolution

Yosuke Hoshino [1,*] and Laura Villanueva [2,3,*]

[1]GFZ German Research Centre for Geosciences, Telegrafenberg, Potsdam 14473, Germany
[2]Department of Marine Microbiology and Biogeochemistry, Royal Netherlands Institute for Sea Research, Landsdiep 4, 't Horntje 1797, The Netherlands
[3]Department of Earth Sciences, Faculty of Geosciences, Utrecht University, Princetonlaan 8a, Utrecht 3584, The Netherlands
*Corresponding author. GFZ German Research Centre for Geosciences, Telegrafenberg, 14473 Potsdam, Germany. Tel: +49 331 6264 1781; E-mail: yhoshino@gfz-potsdam.de; Department of Marine Microbiology and Biogeochemistry, PO BOX 59, Royal Netherlands Institute for Sea Research, 1790 Den Burg, Netherlands. Tel: +31 222 369 300; E-mail: laura.villanueva@nioz.nl
Editor: [Sonja-Verena Albers]

Abstract

Terpenoids, also known as isoprenoids, are the largest and most diverse class of organic compounds in nature and are involved in many membrane-associated cellular processes, including membrane organization, electron transport chain, cell signaling, and phototrophy. Terpenoids are ancient compounds with their origin presumably before the last universal common ancestor. However, Bacteria and Archaea are known to possess two distinct terpenoid repertoires and utilize terpenoids differently. Most notably, archaea constitute their cellular membrane solely made of terpenoid-based phospholipids, contrary to the bacterial membrane that consists of fatty acid-based phospholipids. Thus, the composition of ancestral membranes at the beginning of cellular life and the diversification of terpenoids in early life remain enigmatic. This review addresses these key issues through comprehensive phylogenomic analyses of extant terpenoid biosynthesis enzymes in Bacteria and Archaea. We aim to infer the basal components of terpenoid biosynthesis machinery that have an ancient origin before the divergence of the two domains and shed light on the deep evolutionary connection between terpenoid biochemistry and early life.

Keywords: terpenoid, isoprenoid, paleobiochemistry, membrane evolution, archaea, origin of life

Introduction

The enormous diversity of biomolecules on Earth is a result of the endless expansion of biochemistry throughout Earth's history (Harms and Thornton 2013). Biochemistry serves as the molecular basis for biological evolution and it is also involved in shaping the Earth's environment through the interaction between the biosphere and the geosphere over the geological time scale (Planavsky et al. 2021). Life on Earth is composed of two major domains—Bacteria and Archaea—apart from eukaryotes that seems to be a chimera of the two domains (Eme et al. 2017). Bacteria and Archaea share the basal part of their biochemistry, but the product spectrum of biomolecules is distinct between the two domains as a result of the unique evolutionary history in the individual domains. Hence, comprehensive phylogenomic comparisons between Bacteria and Archaea can potentially give us clues about the biochemical diversity of molecules present in the last universal common ancestor (LUCA) (Markov and Laudet 2022).

Terpenoids, also known as isoprenoids, are a major class of biomolecules comprising more than 96 000 compounds ubiquitous in all life forms (Faylo et al. 2021). Terpenoid biosynthesis has an ancient history and is deeply intertwined with the establishment of biochemistry in its present form. Understanding the evolution of terpenoids, thus provides important information for the evolutionary history of biochemistry and host organisms. In this review, we describe major terpenoid biosynthesis pathways in Bacteria and Archaea and infer the evolutionary trajectory of terpenoid biosynthesis. We ultimately aim to shed light on the stepwise expansion of the terpenoid diversity—terpenome—from LUCA to individual bacterial and archaeal domains across 4 billion years of Earth's evolution.

This review consists of three major parts. The first part overviews the entire terpenoid biosynthesis and the product range (sections Introduction and Overview of microbial terpenoid biosynthesis). The second part depicts the individual steps of terpenoid biosynthesis pathway and associated enzymes in Bacteria and Archaea (sections Monomer formation, Chain elongation, Prenylation and Cyclization). The taxonomic distribution, physiological function, and the evolutionary relationship between individual enzyme families will be discussed. This part serves as the basis to obtain evolutionary implications for the following sections. The third part describes the profound relationship between terpenoids and fundamental cellular processes and the timeline of terpenome evolution since the emergence of life (sections Terpenoids at the origin of primary metabolism and Synthesis - 4 billion years of microbial terpenome evolution). While the second part mainly focuses on the enzymatic details of terpenoid biosynthesis, their evolutionary implications are more extensively discussed in the last part. Hence, readers who seek for the summary of the evolutionary implications about terpenoid biosynthesis may be referred directly to the last two sections and additionally relevant sections in the second part.

In this review, microbial terpenoids refer to terpenoids in Bacteria and Archaea, while terpenoids produced by eukaryotes are in most cases not discussed unless they are closely related to the terpenoid evolution in Bacteria and Archaea. Also, this review focuses on the evolutionary history of terpenoid biosynthesis enzymes. Hence, the catalytic mechanisms and the structures of

Received 9 November 2022; **revised** 13 February 2023; **accepted** 28 February 2023

individual enzymes are not discussed in details. Similarly, this review does not aim to exhaustively describe individual terpenoids in nature. Instead, it describes terpenoid biosynthesis as a unified biochemical system and places major enzymes in the evolutionary timeline.

Overview of microbial terpenoid biosynthesis

Terpenoid biosynthesis consists of three major stages: (I) prenyl monomer formation, (II) chain elongation, and (III) structural modifications (Stages I–III; Fig. 1A). All terpenoids are produced from C_5 prenyl monomer (isoprene) units. The conjugation of those monomers yields prenyl chains of various carbon lengths, depending on host organisms and physiological functions of final products (Fig. 1B). In microbial terpenoid biosynthesis, $C_{>60}$ chains are rare, while extremely long chains are known in some eukaryotes (e.g. rubbers; Cherian et al. 2019). Subsequent structural modifications include the transfer of prenyl groups to non-terpenoid substrates (prenylation) and the cyclization of linear prenyl chains. Individual terpenoid biosynthesis may require only prenylation or cyclization, or both. Additional modifications such as (de)saturation, (de)methylation, and hydroxylation are not discussed in this review since those modifications are generally more taxon-specific.

The evolutionary trajectory of terpenoids seems to be retained in the reaction order of terpenoid biosynthesis. Stage III builds on the products of Stage II, which in turn builds on the products of Stage I. Hence, it is hypothesized that terpenoid biosynthesis evolved stepwise from Stage I to Stages II and III. However, this does not necessarily mean that Stage I was fully established prior to the evolution of Stages II and III. Stage I and the latter two stages are enzymatically distinct, while Stages II and III are more similar to each other and are partly catalyzed by homologous enzymes as described in details later. The majority of terpenoid biosynthesis reactions in Stages II and III are catalyzed by enzymes called prenyltransferases (PTs). PT encompasses enzymes that catalyze any type of prenyl transfer reactions and contains multiple evolutionarily independent enzyme families. PTs are also known for their high evolvability. PTs have developed novel terpenoid modification functions multiple times and these functions contributed to the modern enormous diversity of terpenoid structures and associated physiological functions, which are yet to be fully appreciated (Baunach et al. 2014, Rudolf et al. 2021). The functional plasticity of terpenoid biosynthesis enzymes has been exploited to engineer existing terpenoid biosynthesis pathways to yield further novel bioactive compounds for industrial and medical applications (Faylo et al. 2021).

Bacteria and Archaea share the chain elongation stage of terpenoid biosynthesis. This indicates the antiquity of terpenoids even before the divergence of Bacteria and Archaea. However, there are also important distinctions between the two domains for the monomer formation stage. After diverged from Archaea, Bacteria substantially expanded terpenoid biosynthesis, by recruiting new families of terpenoid-processing enzymes and also functionally diverting original enzymes. While terpenoids are involved in fundamental cellular processes such as respiration, phototrophy, cell signaling, cell wall biosynthesis, and glycosylation (Fig. 1A), bacterial terpenoids also take part in various forms of secondary metabolism. The diversity of modern bacterial terpenoids indeed resides on the large variety of secondary metabolites and also on well-known membrane regulators—hopanoids, steroids, and carotenoids (Fig. 1B). The majority of structural modifications, in particular cyclization, are observed only in Bacteria. The number of newly identified terpenoids in Bacteria is constantly increasing (Rudolf et al. 2021). In particular, advancements of genome mining have accelerated the identification of potential terpenoid biosynthesis genes (Yamada et al. 2015). The large diversity of the bacterial terpenome is in contrast to the archaeal terpenome that principally retains only basal parts of terpenoid biosynthesis and largely lacks secondary metabolite production. Instead, the archaeal terpenome contains some unique terpenoid-based phospholipids that constitute its cellular membrane (Fig. 1B). Following sections describe the details of the three stages.

Monomer formation

All terpenoids are constructed from C_5 prenyl monomers. Monomer formation can proceed via two distinct pathways—the mevalonate (MVA) pathway and the methylerythritol phosphate (MEP) pathway (Fig. 2). Starting compounds are all derived from glycolysis. Detailed enzymology for the two pathways has been previously described (Miziorko 2011, Zhao et al. 2013, Frank and Groll 2017). In general, Archaea possesses the MVA pathway, while Bacteria possesses the MEP pathway, although there are some exceptions and both pathways are in fact found in the two domains. The MVA pathway is known to have multiple variations, each of which has a different taxonomic distribution. In contrast, the MEP pathway is nearly identical among species. It has long been debated which pathway was present in LUCA (Boucher and Doolittle 2000, Lombard and Moreira 2011, Hoshino and Gaucher 2018), even though there is little doubt on the presence of terpenoids in LUCA. The complexity of both pathways, comprising seven or eight enzymatic steps, provoked a speculation that prenyl monomers may have even been abiotically provided in prebiotic Earth before the establishment of modern terpenoid biosynthesis (Nakatani et al. 2014). This section discusses the evolution of the two monomer formation pathways, independently of the ultimate origin of terpenoids.

MVA pathway

The MVA pathway was first described in eukaryotes and was originally thought to be the only pathway for prenyl monomer formation in all three domains of life (Frank and Groll 2017). However, the MEP pathway was subsequently discovered in Bacteria (Rohmer et al. 1993) and also several variations for the MVA pathway have since been described in Archaea (Dellas et al. 2013, Azami et al. 2014, Hayakawa et al. 2018) (Fig. 2). Due to this historical reason, the MVA pathway in eukaryotes is generally thought to be the canonical pathway, while the MVA pathways in Archaea are seen as modifications from the canonical pathway. However, detailed examinations of individual MVA variations provide a different picture for the evolutionary history of the MVA pathway. In fact, the eukaryotic MVA pathway is likely not an ancestral pathway, but instead archaeal MVA pathways seem to retain more ancestral traits, as described below.

Currently, four variations are known for the MVA pathway (Routes I–IV; Fig. 2). No species is found to have more than one variation simultaneously. All four variations are present in Archaea. In particular, Route I is nearly ubiquitously distributed, but is strictly limited to the archaeal domain (Hayakawa et al. 2018). Routes II–IV are found in a small subset of Archaea and Bacteria. Within Archaea, Route II is found in the class Haloarchaea, while Route III is found in the order Thermoplasmatales and Route IV is found in the order Sulfolobales (Lombard and

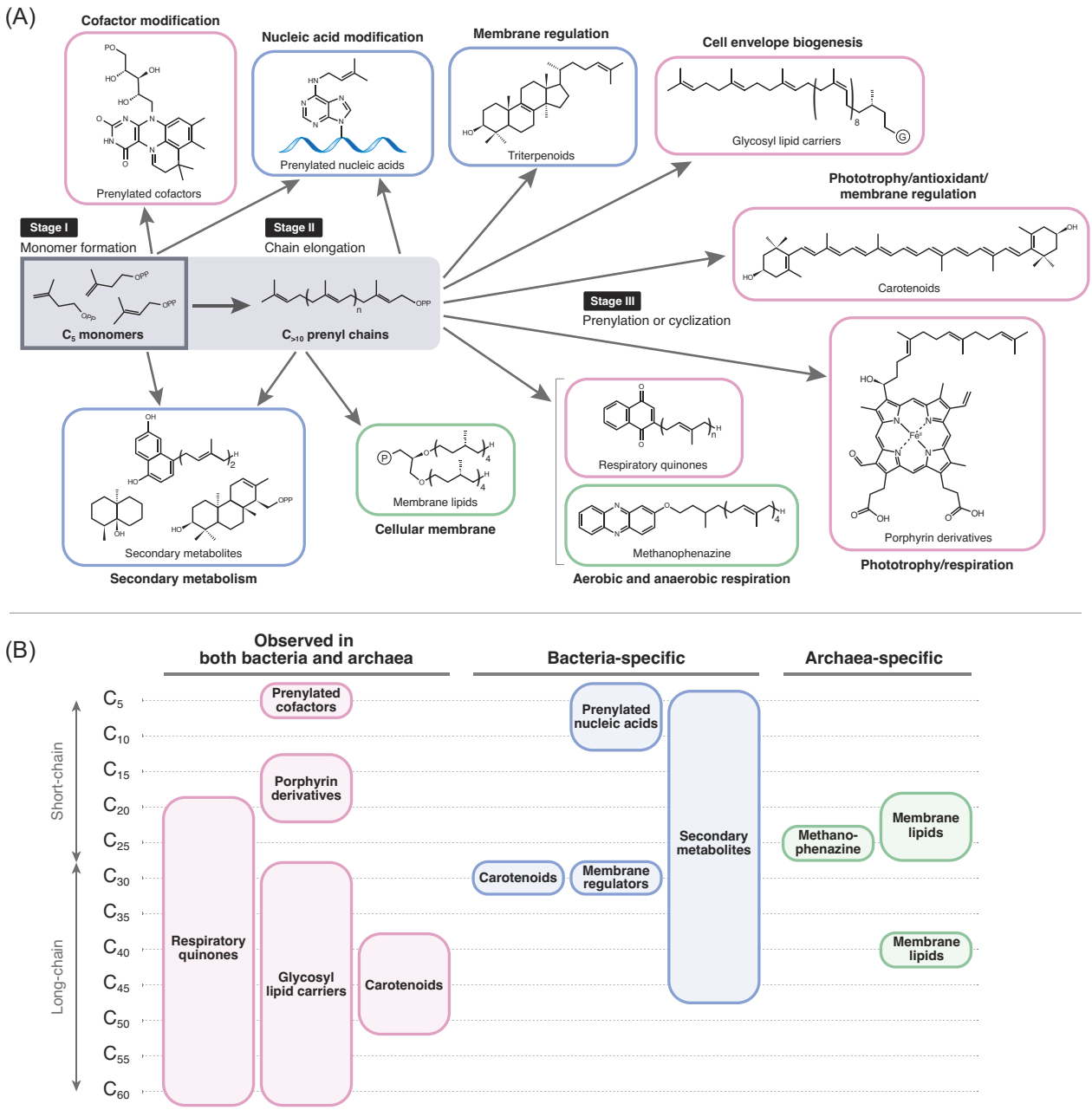

Figure 1. (A) Overview of microbial terpenoid biosynthesis and its functional diversity. Pink color indicates the presence in both Bacteria and Archaea. Blue color indicates the presence only in Bacteria, while green color indicates the presence only in Archaea. Abbreviations: G in a circle, glycosyl group; OP, phosphate; OPP, diphosphate; and P in a circle, phosphate head group. **(B)** Distribution of terpenoids in different carbon number classes (C_5–C_{60}).

Moreira 2011, Hoshino and Gaucher 2018). These observations suggest that Route I is the most ancestral route and was likely present in the last archaeal common ancestor, while the other three routes appeared later. Contrary to Archaea, MVA-containing bacteria possess mostly Route IV, except only for some chloroflexi that harbor Route II. Eukaryotes possess only Route IV. Therefore, Routes I–III are generally archaea-specific, while Route IV is generally tied to Bacteria and Eukarya. Isopentenyl phosphate (IP) is the key intermediate for Routes I–III, while only Route IV does not proceed by way of IP (Fig. 2).

It has been inferred that the eukaryotic MVA pathway originates in Bacteria (Boucher and Doolittle 2000, Lange et al. 2000, Hoshino and Gaucher 2018), although the MVA pathway is only sporadically distributed in Bacteria. Previously, the wide distribution of Route IV enzymes in the candidate phyla radiation (CPR) was utilized as the evidence for the presence of the MVA pathway in the bacterial common ancestor as well as in LUCA (Castelle and Banfield 2018, Hoshino and Gaucher 2018). This inference was based on the hypothesis that CPR is an ancestral bacterial clade that diverged earlier than the rest of all bacteria. However, recent studies suggest a later origin of CPR within Terrabacteria (Taib et al. 2020, Coleman et al. 2021, Moody et al. 2022). Hence, the presence of the MVA pathway in LUCA is not supported anymore by the presence of the MVA pathway coding genes in CPR.

In the MVA pathway, only the first three steps are shared by all four variations (Fig. 2). The latter steps from mevalonate to isopentenyl diphosphate (IPP) are mostly catalyzed by enzymes

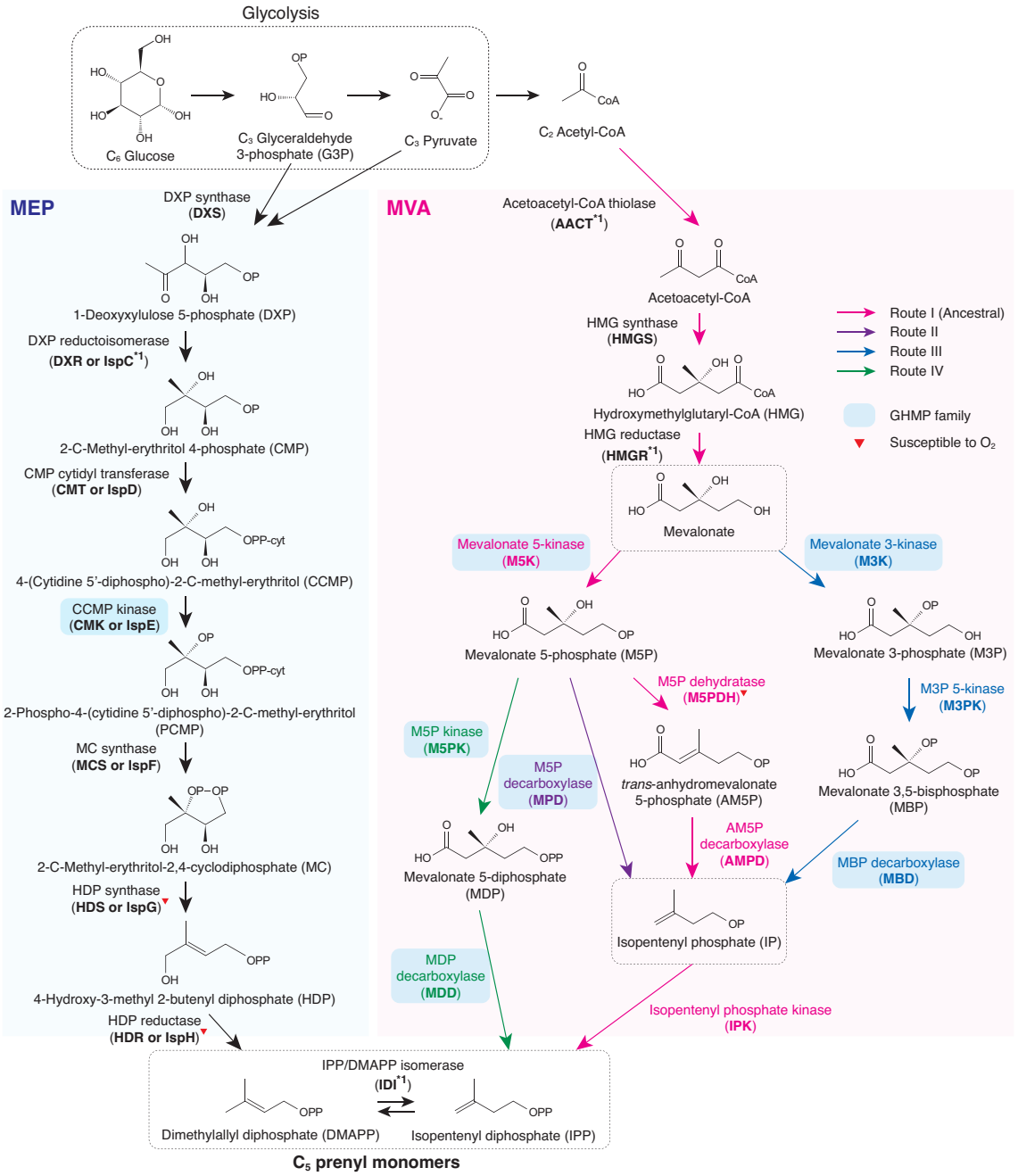

Figure. 2. MVA and MEP pathways for the terpenoid monomer formation. IPP/DMAPP isomerase is shared by both the MVA and MEP pathways. Among the four routes for the MVA pathway, Route I is inferred to be the most ancestral. IPK and MDD produce only IPP, while HDR produces both DMAPP and IPP. ∗1: there are two distinct enzymes, but they are functionally equivalent for each step. For AACT and HMGR, those two enzymes are distantly related to each other, but for DXR and IDI those two enzymes are nonhomologous to each other. See Supplementary Data 1 for the complete list of enzymes and the accession numbers for representative proteins. Abbreviations: cyt, cytidyl group; GHMP, galacto-/homoserine/mevalonate/phosphomevalonate kinase; Isp, isoprenoid biosynthesis enzyme; MEP, methylerythritol phosphate; and MVA, mevalonate.

that belong to the GHMP protein family, which is named after four representative enzymes—galacto-, homoserine, mevalonate, and phosphomevalonate kinases (Fig. 2). Each MVA route utilizes different GHMP proteins and hence the evolutionary history of the MVA pathway is associated to that of GHMP proteins. The most ancestral Route I contains only one GHMP enzyme (mevalonate 5-kinase; M5K), while the other three routes contain multiple GHMP enzymes. An up-to-date phylogeny of GHMP homologs in the MVA pathway was drawn for this review (Fig. 3). The phylogeny suggests that GHMP homologs form two major clades that are closely related to each other: the M5K clade and the mevalonate 5-diphosphate decarboxylase (MDD) clade. The M5K clade contains M5K and mevalonate 5-phosphate kinase (M5PK), while the MDD clade contains MDD, mevalonate 5-phosphate decarboxylase (MPD), mevalonate 3,5-bisphosphate decarboxylase (MBD), and mevalonate 3-kinase (M3K). A recent structural study of GHMP enzymes proposed the evolution of the MVA pathway from Routes I to II (emergence of the MDD clade) and then

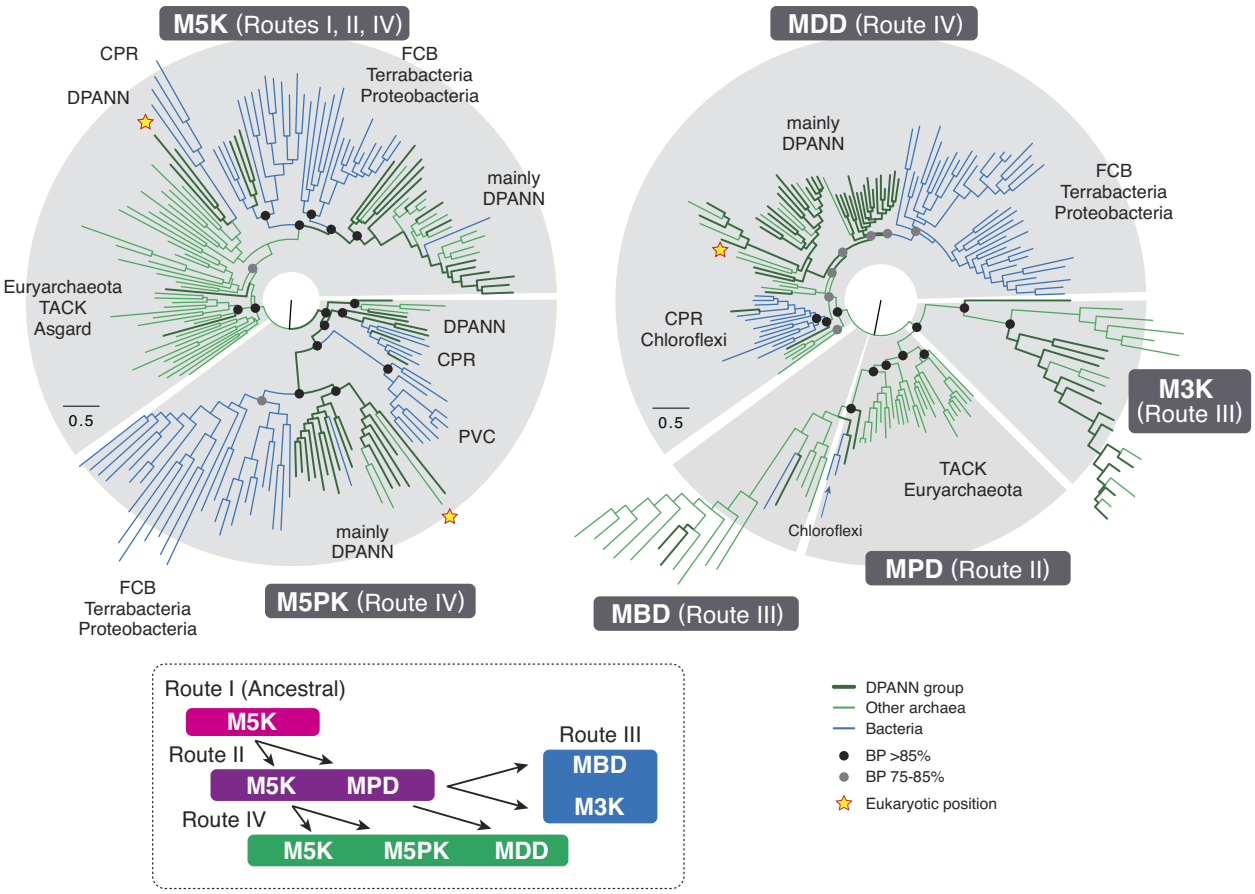

Figure 3. Maximum-likelihood phylogenies for the GHMP proteins in the MVA pathway. The M5K clade and the MDD clade are drawn separately for clarity. Labeling of clades is based solely on the tree topology and does not indicate the functionality of constituent proteins. Eukaryotic proteins are not included in the dataset, but the approximate phylogenetic positions are indicated for reference purposes, based on Hoshino and Gaucher (2018). Protein sequences were retrieved from the NCBI database (https://www.ncbi.nlm.nih.gov/protein/) by Blastp search, using representative proteins (see below) as the query sequences with the cutoff threshold of e-value 10^{-5} and >60% query identity coverage. Taxonomically redundant sequences were excluded. The sequences were aligned using MUSCLE v. 3.8.31. Phylogenetic analyses were performed using IQ-TREE v. 2.0.6. Bootstrap values (BP) were based on 1000 Ultrafast bootstraps. The experimental settings and source data are in the supplementary information. Taxonomy abbreviations: CPR, candidate phyla radiation; DPANN, Diapherotrites, Parvarchaeota, Aenigmarchaeota, Nanoarchaeota, and Nanohaloarchaeota superphylum; FCB, Fibrobacterota, Chlorobiota, and Bacteroidota superphylum; and TACK, Thaumarchaeota, Aigarchaeota, Crenarchaeota, and Korarchaeota superphylum. Enzyme abbreviations and accession numbers for representative proteins in the trees: M5K, mevalonate 5-kinase (EPX62344); M5PK, mevalonate 5-phosphate kinase (EPX62342); MDD, mevalonate 5-diphosphate decarboxylase (EPX62343); MPD, mevalonate 5-phosphate decarboxylase (ABQ92626); MBD, mevalonate 3,5-bisphosphate decarboxylase (BAB59938); and M3K, mevalonate 3-kinase (BAB59465).

independently to III and IV (Aoki et al. 2022). Based on this recent study and the GHMP phylogeny produced in this review, a possible evolutionary order of GHMP proteins for the MVA pathway is shown in Fig. 3. Novel GHMP enzymes likely emerged via the combination of gene duplication and neofunctionalization.

Interestingly, the revised GHMP phylogeny also suggests a possible horizontal transfer of GHMP genes from Archaea to Bacteria and Eukarya (Fig. 3). Previously, Route IV-specific GHMP enzymes (M5PK and MDD) were nearly unknown in Archaea (Nishimura et al. 2013). However, in this review, homologs of those enzymes are found to be widely distributed in many archaea, in particular DPANN archaea (Fig. 3). Further, those archaeal homologs cluster close to bacterial (and eukaryotic) homologs. The same tree topology is observed for M5K as well. These observations suggest that M5K, M5PK, and MDD in Bacteria and Eukarya are derived from DPANN archaea via horizontal gene transfer (HGT). Although the possibility of gene transfer from symbiotic bacterial/eukaryotic hosts to DPANN archaea is not excluded, the broad absence of GHMP genes in Bacteria and contrarily the wide distribution of GHMP genes in DPANN archaea (and in Archaea in general) suggests that the gene transfer from Archaea to Bacteria and Eukarya is more likely. Hence, there is a possibility that all four variations of the MVA pathway evolved within Archaea and transferred only later to Bacteria and Eukarya, possibly independently. Additionally, CPR and other bacteria do not form a monophyletic clade for those three enzymes and thus the MVA genes seem to have similarly transferred to CPR and other bacteria independently.

Lastly, the presence of *trans*-anhydromevalonate 5-phosphate decarboxylase (AMPD) in Route I provides an intriguing and important implication about the origin of the MVA pathway (Fig. 2). The decarboxylation activity of AMPD is dependent on a unique cofactor—prenylated flavin mononucleotide (FMN). FMN is prenylated by flavin prenyltransferase (UbiX), using DMAPP or dimethylallyl monophosphate (DMAP) that is derived from DMAPP (Marshall et al. 2019). Hence, the biosynthesis of prenyl monomers via Route I is dependent on the monomers themselves (Leys 2018). This implies the presence of even an earlier version of the prenyl monomer biosynthesis pathway before the establishment of the DMAPP-feedback loop (see also section UbiX family).

MEP pathway

Unlike the MVA pathway, no variation is known for the MEP pathway thus far (Fig. 2). It is not known what selective pressure caused the diversification of only the MVA pathway. There is little similarity between the MEP and the MVA pathways, except that the fourth step of the MEP pathway is catalyzed also by a GHMP enzyme. The MEP pathway is widely distributed in Bacteria and photosynthetic eukaryotes, while only a small fraction of Archaea are known to possess this pathway (Lohr et al. 2012, Castelle and Banfield 2018, Castelle et al. 2021). The eukaryotic MEP pathway is likely derived from ancestral symbiotic cyanobacteria that transformed into plastids, although other bacterial contributions are additionally suggested (Zeng and Dehesh 2021). Also, the presence of genes encoding some of the enzymes in the MEP pathway is exclusive to the DPANN superphylum in Archaea and is also sparse within the group, unlike the wide distribution of the MVA pathway in DPANN archaea. Therefore, archaeal MEP genes may be derived from symbiotic bacteria via HGT. The MEP pathway was most likely present in the bacterial common ancestor, while the MVA pathway was likely present in the archaeal common ancestor. Hence, although the ubiquity of terpenoid biosynthesis as a whole suggests the presence of terpenoids in LUCA, the ancestral monomer biosynthesis pathway in LUCA remains unresolved.

Chain elongation

Enzymatic repertoire

The majority of reactions for the chain elongation and subsequent structural modifications are catalyzed by PTs. In particular, two PT superfamilies—*trans*-PT and *cis*-PT—are the largest and also the oldest among all PT families(Fig. 4). *Trans*- and *cis*-PT enzymes are metalloenzymes that utilize Mg^{2+} ions, but are evolutionarily independent from each other (Li 2016, Christianson 2017, Chen et al. 2020a). In modern terpenoid biosynthesis, *trans*-PT and *cis*-PT proteins engage in chain elongation, prenylation, and cyclization reactions, but their ancestral functions are most likely confined to chain elongation. Divergent enzymes that involve in prenylation and cyclization are likely evolved later within the *trans*-PT superfamily. In addition to *trans*- and *cis*-PTs, there are many smaller PT families (Fig. 4). These families do not share sequence or structural similarity and thus are evolutionarily independent from one another.

The chain elongation proceeds through the condensation of prenyl monomers (DMAPP and IPP), forming prenyl chains of various carbon lengths in the form of isoprenyl diphosphates (Fig. 5). The chain elongation consists of two consecutive enzymatic steps: (1) $C_{<25}$ short chain formation and (2) $C_{>30}$ long chain formation. The second step utilizes the products of the first step as primers. The chain elongation has several variations, depending on the product configuration and the monomer condensation mode. The first step is performed by isoprenyl diphosphate synthase (IPPS), while the second step is performed by either IPPS or squalene synthase (SQS) family enzymes (Fig. 5). SQS enzymes catalyze a different type of condensation from IPPS). Depending on the stereochemistry of the product, there are two groups of IPPS—*trans (E)*-IPPS and *cis (Z)*-IPPS (Liang et al. 2002). *Trans*-IPPS transfers a prenyl group to a substrate in the *trans*-configuration, while *cis*-IPPS transfers in the *cis*-configuration. *Trans*- and *cis*-IPPSs are evolutionarily independent and the catalytic mechanism is also distinct between the two groups. In contrast, *trans*-IPPS and SQS families are distantly related to each other (*trans*-PT superfamily). The chain elongation stage, in particular the short chain formation, is universally conserved in Bacteria and Archaea and thus provides strong evidence for the presence of terpenoids in and before LUCA. However, there are interesting distinctions between the two domains in the details of the reactions. The chain elongation products serve as important branching points for subsequent modifications. Lastly, several unusual chain elongation mechanisms that likely evolved in specific lineages of Bacteria and Archaea are also briefly described.

Trans-IPPS family

Trans-IPPS produces isoprenyl diphosphates in the *trans*-configuration (Fig. 5). *Trans*-IPPSs are soluble proteins and universally possess a unique α-helical fold that contains two highly conserved Asp-rich motifs, binding to Mg^{2+} ions. The fold is called the terpene synthase fold or the α-fold. The α-fold is conserved in the entire *trans*-PT superfamily (Oldfield and Lin 2012, Li 2016, Christianson 2017). *Trans*-IPPS is divided into two homologous, but distinct enzyme groups: short-chain (C_{10}–C_{25}) and long-chain ($C_{>30}$) groups. Short-chain group catalyzes the first step of the chain elongation, while the long-chain group catalyzes the second step (Fig. 5). The two most prevalent short-chain *trans*-IPPSs are farnesyl diphosphate synthase (FPPS) and geranylgeranyl diphosphate synthase (GGPPS). FPPS produces C_{15} farnesyl diphosphate (FPP), using DMAPP and two molecules of IPP. GGPPS similarly produces C_{20} geranylgeranyl diphosphate (GGPP). In most cases, FPP and GGPP serve as primers for downstream reactions, including $C_{>30}$ chain elongation, prenylation, and cyclization. The distinction between short-chain and long-chain IPPSs for the product chain length has exceptions. For instance, C_{25} prenyl chains in methanogenic archaea are synthesized not by short-chain *trans*-IPPS, but by long-chain *trans*-IPPS (Ogawa et al. 2010). IPPS product chain length is dependent on not only IPPS structure but also primer concentration and theoretical modeling has been developed to predict the chain length (Wallrapp et al. 2013).

The majority of Bacteria and Archaea have a single copy of short-chain *trans*-IPPS, unlike eukaryotes that often have multiple copies (Feng et al. 2020, Satta et al. 2022). The majority of Bacteria, which do not produce carotenoids, have FPPS. In contrast, carotenoid-producing bacteria (e.g. cyanobacteria) have GGPPS to utilize GGPP for C_{40} carotenoid biosynthesis, although the substrate specificity is not necessarily strict (Satta et al. 2022). It is not known if GGPPS in carotenoid-producing bacteria functionally diverged from ancestral FPPS, or alternatively FPPS was replaced with GGPPS that was horizontally acquired. Contrary to the majority of Bacteria, Archaea generally have only GGPPS (Matsumi et al. 2011), although some archaea have a bifunctional enzyme that produces both FPP and GGPP (Chen and Poulter 1993, Fujiwara et al. 2004). The ubiquity of GGPPS in Archaea is consistent with the composition of the archaeal lipid membrane that is composed primarily of C_{20} prenyl chains (Jain et al. 2014). Hence, the major short-chain product is different between Bacteria and Archaea. A similar dichotomy is observed for long-chain *trans*-IPPS that utilizes the products of short-chain *trans*-IPPS. FPP is the substrate for the long-chain *trans*-IPPS in the majority of Bacteria (Liang et al. 2002). In contrast, GGPP is the likely substrate for the archaeal long-chain *trans*-IPPS, although this is not experimentally confirmed.

Previous phylogenetic analyses suggest that both short-chain and long-chain *trans*-IPPSs were present in LUCA (Lombard and Moreira 2011). However, there is a notable exception for the

	Enzyme family	Reaction	Domain	Example	Accession #	Enzyme family	Reaction	Domain	Example	Accession #
Trans-PT										
Trans-IPPS	**Short-chain *trans*-IPPS**	E	AB	FPPS	P22939	IPPS-like	P	B	ComQ	P33690
	Long-chain *trans*-IPPS	E	AB	OctPPS	P0AD57					
SQS	HpnCDE-route SQS	E	AB	HpnD	Q6N3F2	SQS-like	P	B	NzsG	BBF24926
	Sqs-route SQS	E	AB	Sqs	Q7NE22					
	Carotenoid synthase	E	AB	CrtB	P37294					
UbiA	**Aromatic PT** *1	P	AB	MenA	P32166	UbiA-like	C	B	PtmT1	D8L2U4
	Non-aromatic PT *1	P	AB	DGGGPS	Q9UWY6					
						Class I TC	C	B	CIS	B5GMG2
						Class IB TC	C	B	BsuTS	O34707
Cis-PT	***Cis*-IPPS**	E	AB	UPPS	P60472	IPPS-like	E	B	Mcl22	M4T4U9
							P	B	Rv3378c	P9WJ61
							C	B	CLDS	X5IYJ5
GGGPS	**GGGPS**	P	AB		Q9HJH3	GGGPS-like	P	B	PcrB	O34790
	UbiX	P	AB		P0AG03	Class II TC	C	B	SHC	P33247
	LC	C	AB	CrtY	P21687	Protein PT *5	P	-	FTase	Q9LX33
	TES	E	A(B) *2		Q58036	**MiaA**	P	B		P16384
	GrsAB	C	A(B) *2	GrsA	Q4J8I0	SelU	P	B		P33667
						ABBA	P	B	NphB	Q4R2T2
	PGT (polytopic)	(P)	AB	AglH	P39465	F	P	B	AgcF	BCR82584
	PGT (monotopic)	(P)	AB	WbaP	P26406	IDT	C	B	XiaE	I7KIT8
	GT2	(P)	AB	AglD	D4GUA0					
	Xrt/Art *3	(P)	AB	ArtA	D4GUZ4					
	Lgt-like *3	(P)	(A)B *4		P60955					
	WS2	(P)	B		A3RE51					

Figure 4. Summary of enzymes that engage in chain elongation (E), prenylation (P), and cyclization (C). The enzyme families in boldface indicate that their origin possibly dates back to the common ancestor of Bacteria and/or Archaea. Domain labels A and B represent Archaea and Bacteria. The pink background indicates the presence of enzymes in both Bacteria and Archaea. The blue background indicates the presence only in Bacteria, while the green background indicates only in Archaea. The gray background indicates that enzymes are non-PT enzymes. Not all characterized cyclase families are shown in the list. See also Supplementary Data 1 for the complete list of enzymes that are described in this review. Accession numbers for representative proteins are retrieved from UniProt (https://www.uniprot.org) or the NCBI database (https://www.ncbi.nlm.nih.gov/). ∗1: the separation of aromatic and nonaromatic PTs is only for reference purposes and is not based on the sequence similarity (see Fig. 8). ∗2: the function of bacterial homologs is unknown. ∗3: the enzymatic activity can be seen as prenylation only in the archaeal membrane, while the same activity is equal to fatty acylation in the bacterial membrane (see section Xrt/Art and Lgt-like families). ∗4: Lgt homologs are not observed in Archaea, but the presence of a functionally analogous enzyme is predicted in the domain. ∗5: protein PTs are nearly absent in Bacteria and Archaea, but the family possibly has an evolutionary link to the bacterial class II terpene cyclase and thus is included. Abbreviations (ordered as in the list): PT, prenyltransferase; IPPS, isoprenyl diphosphate synthase; FPPS, farnesyl diphosphate synthase; OctPPS, octaprenyl diphosphate synthase; SQS, squalene synthase; Hpn, hopanoid biosynthesis enzyme; HpnD, presqualene diphosphate synthase; Sqs, squalene synthase; CrtB, phytoene synthase (carotenoid biosynthesis enzyme B); UbiA, 4-hydroxybenzoate PT; MenA, 1,4-dihydroxy-2-naphthoate PT; DGGGPS, digeranylgeranylglyceryl phosphate synthase; UPPS, undecaprenyl diphosphate synthase; GGGPS, geranylgeranylglyceryl phosphate synthase; UbiX, flavin PT; LC, lycopene cyclase; CrtY, lycopene cyclase β-cyclase; TES, tetraether synthase; GrsAB, glycerol dibiphytanyl glycerol tetraether ring synthases AB; PGT, phosphoglycosyl transferase; AglH, archaeal glycosylation enzyme H; WbaP, galactose-1-phosphate transferase; GT2, family 2 glycosyltransferase; AglD, β-mannosyltransferase; Xrt/Art: Exosortase/Archaeosortase; ArtA, archaeosortase A; Lgt-like, prolipoprotein diacylglyceryl transferase-like; WS2, wax ester synthase 2; ComQ, tryptophan PT; NzsG, carbazole biosynthesis enzyme G; PtmT1, *ent*-atiserene synthase; TC, terpene cyclase; CIS, 1,8-cineole synthase; BsuTS (YtpB), tetraprenyl-β-curcumene synthase; Mcl22, isosesquilavandulyl diphosphate synthase; Rv3378c, tuberculosinyladenosine synthase; CLDS, cyclolavandulyl diphosphate synthase; PcrB, heptaprenylglyceryl phosphate synthase; SHC, squalene–hopene cyclase; FTase, farnesyl transferase; MiaA, tRNA dimethylallyltransferase; SelU, tRNA 2-selenouridine synthase; ABBA, $\alpha\beta\beta\alpha$-fold; NphB, naphterpin biosynthesis enzyme B; F, peptide PT; AgcF, argicyclamide biosynthesis enzyme F; IDT, indole diterpene synthase; and XiaE, xiamycin cyclase.

presence of *trans*-IPPS in Archaea. While short-chain *trans*-IPPS is universally distributed in the domain, long-chain *trans*-IPPS seems to be absent in some early-branching taxa (e.g. DPANN superphylum, methanogens) as discussed in more details later (section Quinones and methanophenazine). Long-chain *trans*-IPPS is closely associated with respiration and thus the evolution of long-chain *trans*-IPPS has an important implication for constraining the origin of respiration. Long-chain *trans*-IPPS relies on the products of short-chain *trans*-IPPS (FPP or GGPP). This implies that short-chain *trans*-IPPS has an older origin than long-chain *trans*-IPPS. Hence, long-chain *trans*-IPPS likely evolved from within short-chain *trans*-IPPS by gene duplication. Among C_{10}–C_{25} *trans*-IPPS,

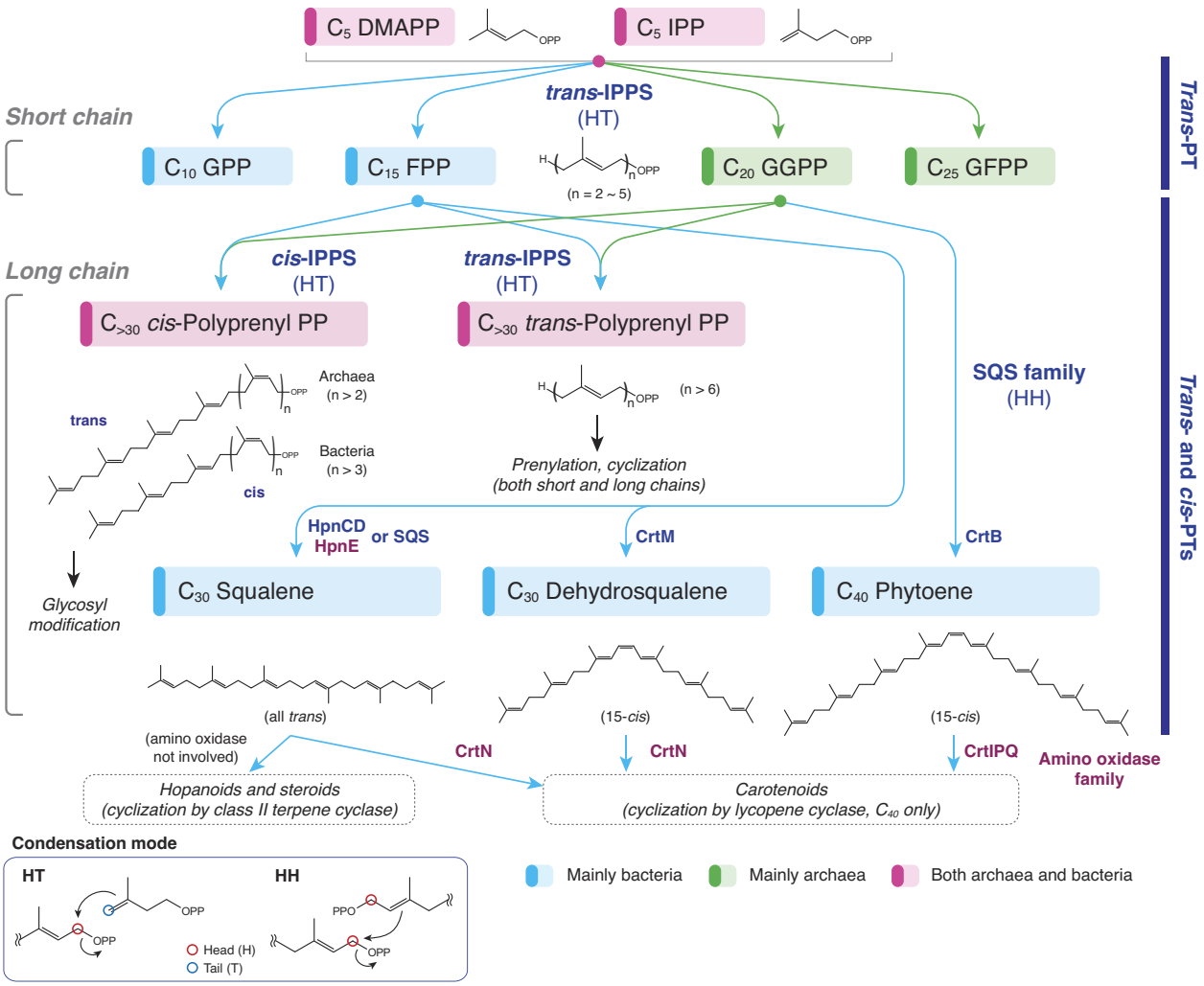

Figure 5. Schematic diagram of prenyl chain elongation. The head-to-tail (HT) condensation is the canonical (ancestral) reaction, while the head-to-head (HH) condensation represents a noncanonical (and more recent) condensation. $C_{>30}$ indicates prenyl chains of C_{30} or longer. Exceptions from the color code for Bacteria and Archaea (green and blue) include carotenoid producers: (1) bacterial species produce C_{20} GGPP instead of or in addition to C_{15} FPP and (2) archaeal species perform the HH condensation to produce C_{40} phytoene. ∗1: the chemical structure of archaeal *cis*-polyprenyl PP is not experimentally confirmed and remains theoretical. Abbreviations and accession numbers for representative proteins: DMAPP, dimethylallyl diphosphate; IPP, isopentenyl diphosphate; IPPS, isoprenyl diphosphate synthase (P22939, short-chain; P0AD57, long-chain; P60472, *cis*-chain); GPP, geranyl diphosphate; FPP, farnesyl diphosphate; GGPP, geranylgeranyl diphosphate; GFPP, geranylfarnesyl diphosphate; PP/OPP, diphosphate; SQS, squalene synthase; HpnD, presqualene diphosphate synthase (ABE49737); HpnC, hydroxysqualene synthase (ABE49736); HpnE, hydroxysqualene dehydroxylase (ABE49738); Sqs, squalene synthase (BAC91999); CrtM, dehydrosqualene synthase (ACA82621); CrtB, phytoene synthase (BAC89685); CrtN, dehydrosqualene desaturase (O07855); CrtI/P, phytoene desaturase (P17054/Q53589); and CrtQ, ζ-carotene desaturase (Q7A3E0).

C_{10} geranyl diphosphate synthase (GPPS) and C_{25} geranylfarnesyl diphosphate synthase (GFPPS) are rare (Tachibana et al. 2000, Mann et al. 2011, Hou and Dickschat 2020). Therefore, a late origin of GPPS and GFPPS is suggested, while FPPS and/or GGPPS are inferred to be the two oldest short-chain *trans*-IPPS enzymes. It is unknown if the earliest *trans*-IPPS produced only FPP or GGPP, or was bifunctional, producing both FPP and GGPP (Chen et al. 1994, Szkopińska and Płochocka 2005).

SQS family

Terpenoid chain elongation by *trans*- and *cis*-IPPSs proceeds through the "canonical" head-to-tail (HT) condensation, by which the head (diphosphate end) of IPP is added to the tail (aliphatic end) of pre-existing isoprenyl diphosphate molecules (Fig. 5). However, there is an additional mechanism for chain elongation: the head-to-head (HH) condensation. In this mechanism, the head of an isoprenyl diphosphate molecule is attached to the head of another isoprenyl diphosphate molecule. Unlike the HT condensation, the HH condensation is a one-time reaction and is not repeated. The HH condensation is catalyzed by the SQS family enzymes that are characterized by the presence of the so-called ε-fold that displays a high structural similarity to the α-fold in *trans*-IPPS, sharing a common ancestry (Cao et al. 2010, Oldfield and Lin 2012). Therefore, the HH condensation is a functional off-shoot of the HT condensation. The important feature of the HH condensation is that the condensation removes phosphate groups from substrates, and thus the products are entirely hydrophobic, unlike HT condensation products. Therefore, HH condensation products are embedded within lipid membranes and subsequent modifications also take place in membranes. Consistently, SQS proteins are membrane-bound, while *trans*-IPPS proteins are soluble.

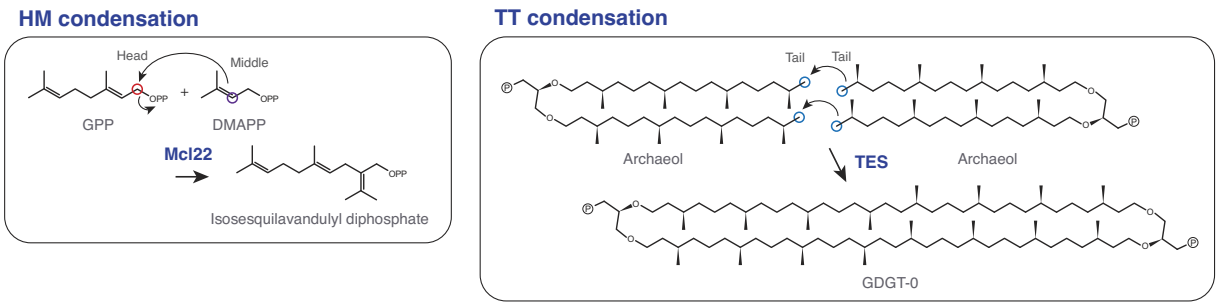

Figure 6. Unconventional modes of chain elongation—the head-to-middle (HM) and the tail-to-tail (TT) condensations. Abbreviations and accession numbers for representative proteins: Mcl22, isosesquilavandulyl diphosphate synthase (merochlorin biosynthesis enzyme 22) (M4T4U9); GPP, geranyl diphosphate; DMAPP, dimethylallyl diphosphate; OPP, diphosphate; TES, tetraether synthase (Q58036); GDGT-0, glycerol dibiphytanyl glycerol tetraether-0; and P in a circle, phosphate head group.

The HH condensation yields all-*trans* products (C_{30} squalene) or 15-*cis* products (C_{30} dehydrosqualene and C_{40} phytoene) (Fig. 5). Dehydrosqualene and phytoene are the simplest members of C_{30} and C_{40} carotenoids, while squalene is the precursor for all hopanoids, steroids, and carotenoids. Squalene is produced by two routes: HpnCDE and Sqs routes (Pan et al. 2015). The first route comprises two SQS family proteins (HpnC and HpnD) and additionally a FAD-dependent amino oxidase HpnE. In contrast, the second route comprises only Sqs. Both HpnCDE and Sqs are found in Bacteria and Archaea, but their distribution is largely exclusive to each other at the species level. Dehydrosqualene and phytoene are produced by different SQS enzymes CrtM and CrtB, respectively (Sandmann 2021). Interestingly, the combination of SQS and amino oxidase proteins is also observed in the carotenoid biosynthesis pathway and amino oxidase acts on dehydrosqualene and phytoene as desaturases (CrtN and CrtIPQ) (Fig. 5). Hence, the early steps of carotenoid biosynthesis and hopanoid/steroid biosynthesis are similar to each other, although the function of amino oxidase is different (desaturase and dehydroxylase).

Phylogenetic analyses of SQS and amino oxidase proteins suggest that squalene biosynthesis enzymes HpnCD evolved from carotenoid biosynthesis enzymes CrtBM (Santana-Molina et al. 2020), while the associated amino oxidases evolved independently in each pathway from an ancestor of unknown function (Santana-Molina et al. 2022). Within squalene biosynthesis, the first route (HpnCDE) seems to be more ancestral than the second route (Sqs). All amino oxidases in carotenoid/squalene biosynthesis use flavin adenine dinucleotide ($FADH_2$) as the cofactor. For carotenoid biosynthesis, the cofactor is not required during the HH condensation, but is utilized in the next desaturation step. In contrast, for squalene biosynthesis the cofactor is required to complete the HH condensation, due to an additional reductive step (Pan et al. 2015). This step is catalyzed by a separate enzyme HpnE in the first route, while Sqs in the second route performs the whole HH condensation process, utilizing a different cofactor nicotinamide adenine dinucleotide phosphate (NADPH). Hence, Sqs is functionally more diverged, compared to HpnCD and CrtBM. The origin of the SQS family (i.e. CrtBM) possibly dates back to the common ancestor of Gracillicutes in Bacteria (Santana-Molina et al. 2020). In contrast, the distribution of SQS proteins in Archaea is limited (e.g. haloarchaea), and thus a horizontal origin from Bacteria is inferred. The evolution of the SQS family has an important implication for the development of membrane regulation and phototrophy in Bacteria and is discussed in more details in sections Membrane organization and Phototrophy.

Cis-IPPS family

Cis-IPPS produces $C_{>30}$ long isoprenyl diphosphates in the *cis*-configuration via the HT condensation. While the *cis*-PT and *trans*-PT families do not share any sequence and structural similarity, they share a mechanistic similarity. For both enzyme families, diphosphate abstraction generates the electrophilic donor, which initiates prenyl transfer reactions. *Cis*-IPPS harbors the so-called the ζ-fold that is distinct from the α-fold in *trans*-IPPS, but also requires Mg^{2+} ions (Oldfield and Lin 2012, Grabińska et al. 2016, Chen et al. 2020a). *Cis*-IPPSs are soluble proteins, similar to *trans*-IPPS. The substrate of *cis*-IPPS is FPP in Bacteria and probably GGPP in Archaea, as is the case for long-chain *trans*-IPPS. Thus, *cis*-IPPS products do not have the all-*cis* configuration, but contain the *trans*-configuration that is derived from FPP or GGPP at one end of the molecule (Chen et al. 2020a) (Fig. 5). Bacterial *cis*-IPPS generally produces C_{55} undecaprenyl diphosphate (Manat et al. 2014), while archaeal *cis*-IPPS displays a wider product range, yielding $C_{30–70}$ chains (Eichler and Guan 2017). *Cis*-IPPS is ubiquitously distributed in Bacteria and Archaea and a phylogenetic analysis indeed suggests the presence of *cis*-IPPS in LUCA (Lombard 2016a). *Cis*-IPPS products serve as scaffolds for cell surface biogenesis such as peptidoglycans and N-glycosylated S-layers. *Trans*- and *cis*-IPPS do not share physiological functions, despite their catalyzing similar biochemical reactions. Perhaps, it is to avoid cross interference. Membrane modulation effects of *cis*-IPPS products (*cis*-polyprenols) have also been suggested (Hartley and Imperiali 2012). Similar to long-chain *trans*-IPPS, *cis*-IPPS utilize short-chain *trans*-IPPS products as the substrates and this implies that *cis*-IPPS evolved later than short-chain *trans*-IPPS.

In rare cases, *cis*-IPPS homologs are known to catalyze the so-called head-to-middle (HM) condensation, where the head of an isoprenyl diphosphate is attached to an internal carbon atom of another isoprenyl diphosphate molecule (Ogawa et al. 2016, Gao et al. 2018). For instance, a merochlorin biosynthesis enzyme Mcl22 conjugates DMAPP and GPP, yielding not FPP, but isosesquilavandulyl diphosphate (Gao et al. 2018) (Fig. 6). The HM condensation is nearly absent in Bacteria and Archaea, and thus is inferred to have evolved from conventional *cis*-IPPS only in specific lineages. In eukaryotes, several different types of HM condensation are known. They are generally catalyzed by *trans*-IPPS homologs (Thulasiram et al. 2007), but some proteins are yet to be characterized (e.g. highly branched isoprenoids) (Belt et al. 2000).

TES family

Besides the chain elongations described above (HT, HH, and HM condensations), an unusual radical-based chain elongation is

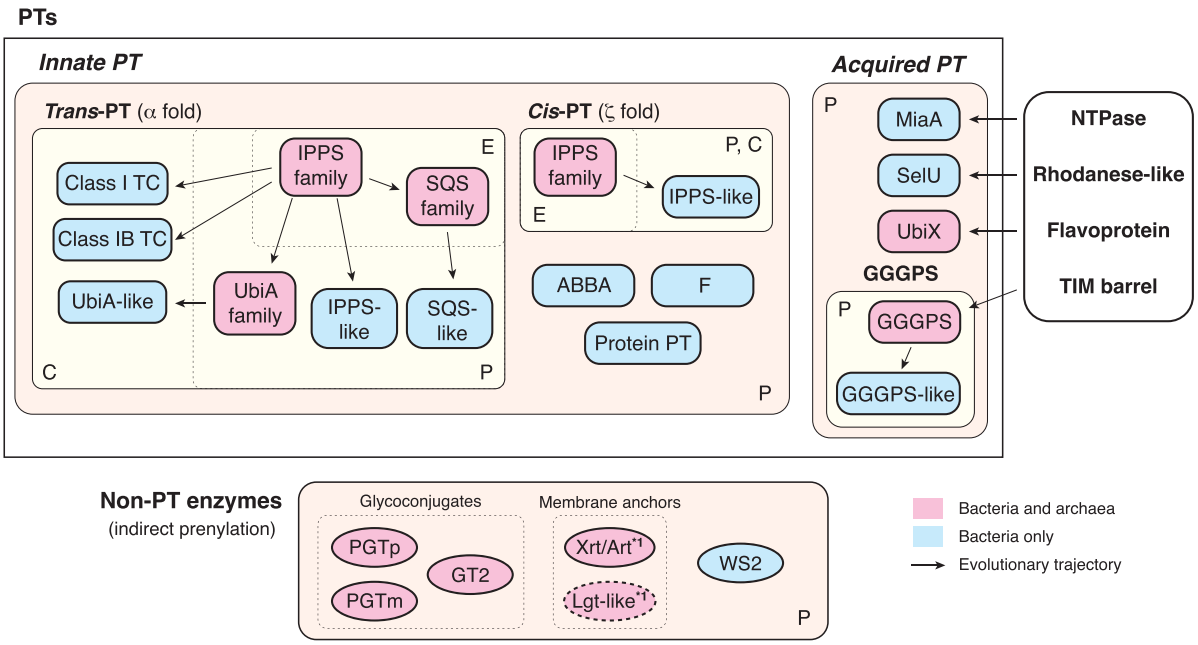

Figure 7. Evolutionary relationship of PTs and their catalytic functions. E, P, and C indicate chain elongation, prenylation, and cyclization. Non-PT enzymes that catalyze indirect prenylation reactions are also shown. *In vitro* enzymatic activities with non-natural substrates are not included. *1: the presence of functionally analogous proteins is predicted in Archaea, but is not experimentally confirmed. Abbreviations: NTPase, nucleoside triphosphatase and TIM, triosephosphate isomerase. See Fig. 4 for enzyme abbreviations.

known in Archaea. Tetraether synthase (TES) catalyzes the condensation of C_{20} prenyl chains to form C_{40} biphytanyl chains that constitute glycerol dibiphytanyl glycerol tetraethers (GDGTs) during the archaeal tetraether membrane biosynthesis (Lloyd et al. 2022, Zeng et al. 2022) (Fig. 6). TES is a radical SAM methyltransferase and conjugates the terminal methyl group of prenyl chains (Lloyd et al. 2022). The overall reaction is the tail-to-tail condensation, which is unique to TES. Both the substrate and the condensation product are the archaeal membrane itself.

Prenylation

Structural modifications, including prenylation and cyclization, are the main source for the enormous diversity of terpenoids. Prenylation is a process to transfer a prenyl group to a substrate. Prenylation modifies the physical and chemical properties of substrate compounds and is utilized to regulate their trafficking, sublocalization, stability, and function. Prenylation is observed for all major classes of organic compounds, including proteins, oligopeptides, saccharides, nucleic acids, lipids, and other small molecules such as porphyrins and quinones (Winkelblech et al. 2015). Prenylation is the stage to produce terpenoid-containing metabolites that are designed for specific physiological functions in primary/secondary metabolism.

Prenylation is catalyzed by various PT families, but many of them are evolutionarily related to one another (Fig. 7). In particular, the *trans*-PT superfamily that contains enzymes for chain elongation reactions also contains enzymes for prenylation (and cyclization) reactions. In this review, PTs are divided into innate and acquired groups for convenience (Fig. 7). Innate PTs are enzymes that are associated solely with terpenoid metabolism. Any evolutionary relationship to proteins that are not involved in terpenoid metabolism is not known (or not retained). In contrast, acquired PTs are enzymes that acquired the prenylation activity through functionally diverting from ancestral proteins that had functions irrelevant of terpenoid metabolism. It is noted that innate PTs are not necessarily older than acquired PTs. In fact, some acquired PTs probably have a comparably ancient origin to innate PTs, as discussed later. There are also some non-PT enzymes that indirectly perform prenylation reactions (Fig. 7). These enzymes do not catalyze prenylation itself, but transfer non-terpenoid substrates to acceptors that contain a prenyl moiety, including the archaeal lipid membrane (therefore, called membrane anchoring). Addressing the evolution of membrane anchoring pathways provide important clues for the status of ancestral cellular membranes and their interactions with substrate molecules (e.g. proteins and oligosaccharides).

Interestingly, the substrates for prenylation are predominantly *trans*-prenyl chains that are produced by *trans*-IPPS, apart from prenyl monomers. In contrast, prenylation of *cis*-prenyl chains that are produced by *cis*-IPPS is nearly absent. Instead, *cis*-prenyl chains are indirectly prenylated to other molecules by non-PT enzymes. The origin of this functional segregation is not known. Following sections describes individual PT and non-PT families.

Trans-PT superfamily (innate PTs)

UbiA family

This family is represented by 4-hydroxybenzoate PT (UbiA) during ubiquinone biosynthesis and is among the oldest PT families. UbiA proteins are transmembrane proteins, but are structurally similar to soluble *trans*-IPPS proteins, sharing a common ancestor (Li 2016) (Fig. 7). Thus, soluble *trans*-IPPS proteins seem to have adapted to hydrophobic membrane environments at least twice (UbiA and SQS families). Similar to SQS enzymes, the prenyl substrates of UbiA enzymes are the products of *trans*-IPPS. Prenylation occurs typically to aromatic substrates that

include quinones, phenazines, and porphyrins, but nonaromatic substrates are also prenylated, including carotenoids, glycerol phosphates, and riboses (Winkelblech et al. 2015, Yang et al. 2015, Li 2016) (Fig. 8). Most UbiA enzymes utilize FPP or longer, but some enzymes are known to use DMAPP and GPP (Zeyhle et al. 2014). For the former, prenylated compounds are generally involved in primary metabolism, including membrane organization, respiration, and phototrophy. In contrast, for the latter, prenylated compounds mostly engage in secondary metabolism. Overall, UbiA metabolites are both structurally and functionally diverse.

The UbiA family is one of the most widespread PT families in nature (Nowicka and Kruk 2010), together with *trans*- and *cis*-IPPSs. For instance, digeranylgeranylglyceryl phosphate synthase (DGGGPS) is a ubiquitous UbiA enzyme in Archaea and catalyzes the second of the two consecutive GGPP transfer reactions to glycerol-1-phosphate (G1P), yielding the precursor to archaeal membrane lipids (Koga and Morii 2007) (Fig. 8). The first GGPP transfer reaction is catalyzed by a non-UbiA enzyme (see section GGGPS family). For another example, quinone prenylation by UbiA enzymes is universally distributed in both Bacteria and Archaea, which reflects the ubiquity of quinone-based respirations in the two domains. The prenyl side chain length is typically C_{30}–C_{50} (Collins and Jones 1981, Nowicka and Kruk 2010, Elling et al. 2016). Quinone PTs consist of two distinct clades that are represented by two menaquinone biosynthesis enzymes: 1,4-dihydroxy-2-naphthoate PT (MenA) and 1,4-dihydroxy-6-naphthoate PT (MqnP) (Fig. 8). The presence of two clades reflects two distinct menaquinone biosynthesis pathways—the canonical (*o*-succinylbenzoate) pathway and the more recently described futalosine pathway (Dairi 2009). MenA is involved in the canonical pathway, while MqnP is hypothesized to take part in the futalosine pathway (Fig. 9). The MenA clade additionally contains 2-carboxy-1,4-naphthoquinone PT (cyMenA) that participates in phylloquinone biosynthesis in cyanobacteria. In contrast, the MqnP clade additionally contains UbiA for ubiquinone biosynthesis in proteobacteria and 4-hydroxybenzoate solanesyl transferase (HBST) for plastoquinone biosynthesis in cyanobacteria. Proteobacteria generally possess only UbiA, reflecting the replacement of menaquinone with ubiquinone in the entire lineage (Fig. 9).

Cyanobacteria seem to have switched their quinone repertoire from menaquinone to phylloquinone and plastoquinone in accordance with the evolution of oxygenic photosynthesis. The anaerobic sister clade of cyanobacteria (melainabacteria) are found to mostly lack MenA homologs, but instead have MqnP homologs (Table S1, Supporting Information). Cyanobacterial MqnP homologs (HBST) are distant from MqnP homologs in other bacteria, but instead are closer to UbiA (Fig. 8). This is consistent with the observation that cyanobacteria biosynthesize plastoquinone, instead of menaquinone. It was also recently inferred that ubiquinone biosynthesis evolved from the cyanobacterial plastoquinone biosynthesis (Degli Esposti 2017). Hence, plastoquinone biosynthesis in cyanobacteria seems to have emerged from the functional divergence of menaquinone biosynthesis in anaerobic cyanobacterial ancestors and further transformed into ubiquinone biosynthesis in proteobacteria (Ravcheev and Thiele 2016, Abby et al. 2020). In contrast, eukaryotic plastoquinone biosynthesis utilizes homogentisate solanesyl transferase (HST) that is evolutionarily distinct from cyanobacterial HBST (Figs. 8 and 9). Tocopherols, which are structurally related to plastoquinone (Block et al. 2013), are produced by homogentisate phytyltransferase (HPT) that is closely related to HST. Cyanobacteria additionally have a MenA homolog (cyMenA) that is involved in phylloquinone biosynthesis (Fig. 9). This enzyme may have a horizontal origin because melainabacteria generally do not have MenA.

In addition to DGGGPS and quinone PTs, there are many UbiA subfamilies that prenylate different substrates (Fig. 8). Heme O synthase (COX10) catalyzes the C_{15} or C_{20} prenylation of heme B to produce hemes A and O that are utilized for aerobic respiration (see also section Hemes) (Lübben and Morand 1994, Hederstedt 2012). COX10 homologs are widespread in Bacteria and Archaea, but are largely limited to aerobic species (Li 2016). Similarly, (bacterio)chlorophyll synthase (BCS) catalyzes the C_{15} or C_{20} prenylation of the (bacterio)chlorophyll chlorin ring. While chlorophylls mostly have a C_{20} chain, many bacteriochlorophylls have a C_{15} chain (Proctor et al. 2022). The distribution of BCS homologs is confined to photosynthetic organisms. Lycopene elongase (LyeJ) catalyzes the prenylation of C_{40} carotenoids to produce C_{50} elongated carotenoids by adding two prenyl monomers (Krubasik et al. 2001, Yang et al. 2015). LyeJ homologs are found in many species that are not known to produce carotenoids, and hence those homologs may have functions that are irrelevant to carotenoids. In contrast, decaprenylphosphoryl-β-D-5-phosphoribose synthase (DPPRS) is a rare example of *cis*-prenyl chain prenylation. DPPRS has an important role in the mycobacterial cell wall production (Li 2016). DPPRS homologs are widespread in Bacteria and Archaea, but the function of those homologs in non-mycobacterial species is unknown.

Interestingly, some UbiA proteins function as cyclases—GGPP/GFPP cyclase and *ent*-copalyl disposphate (*ent*-CPP) cyclase (Fig. 8). These enzymes are functionally similar to class I terpene cyclase (section Class I and IB TC families). UbiA-like cyclases are only sporadically distributed in Bacteria and seem to have secondarily evolved from UbiA enzymes that catalyze prenylation (see section UbiA family). In particular, *ent*-CPP cyclase is close to 1,6-dihydroxyphenazine geranyltransferase (CnqPT1) and may have functionally diverged from a CnqPT1-like PT. Phenazine prenylation is mostly performed by a different PT family (ABBA family), and thus CnqPT1 is a rare exception. Lastly, two unknown clusters that contain only uncharacterized proteins were found (U1 and U2 clusters; Fig. 8). Homologs from these clusters are distributed in several discrete bacterial and archaeal taxa. Overall, individual UbiA clades mostly contain both bacterial and archaeal species and a complex history of HGT is inferred.

Other trans-PT enzymes (trans-IPPS-like and SQS-like families)

Several additional *trans*-PT homologs are known to catalyze prenylation reactions. These enzymes participate in a diverse range of metabolic pathways and this review only briefly describes several well-known examples: ComQ, MoeN5, and XiaM (*trans*-IPPS homologs) and NzsG (SQS homolog). Tryptophan prenyltransferase ComQ catalyzes the GPP- or FPP-prenylation of heptapeptide pheromone ComX, which is part of the ComQXPA quorum-sensing system in Bacillales (Hirooka et al. 2020) (Fig. 10). In contrast, neocarazostatin biosynthesis enzyme NzsG, moenomycin biosynthesis enzyme MoeN5, and xiamycin biosynthesis enzyme XiaM are involved in DMAPP-, GPP-, and FPP-prenylations of antibiotic metabolites, respectively (Xu et al. 2012, Huang et al. 2015, Zhang et al. 2016, Kobayashi and Kuzuyama 2020). MoeN5 prenylation proceeds by a unique HM condensation of C_{10} and C_{15} prenyl moieties and a subsequent intramolecular rearrangement

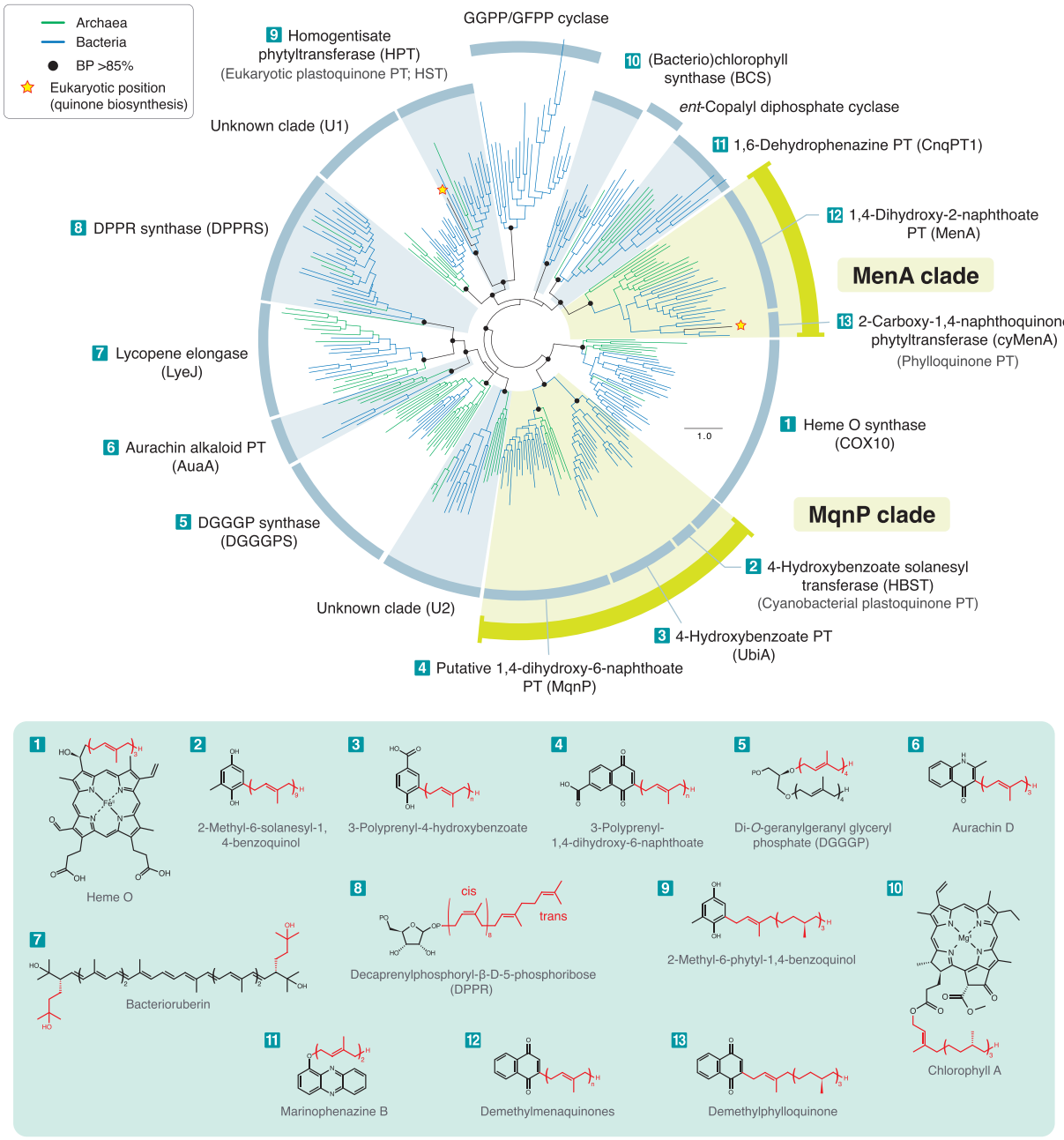

Figure 8. Maximum-likelihood phylogeny of UbiA enzymes and their product range (numbers 1–13). Clustering is solely based on the sequence similarity, while the majority of proteins are not characterized. Products of UbiA-like cyclases are shown in Fig. 14. Microbial quinone-associated UbiA families consist of two distinct clusters: MqnP and MenA clades. The tree contains two eukaryotic proteins as references (star marks). One is the plant HST (Q1ACB3) that belongs to the HPT family and is distinct from the cyanobacterial HBST that is included in the MqnP clade, but both HST and HBST engage in plastoquinone biosynthesis. The other is the plant MenA homolog (Q0WUA3) that is similar to the cyanobacterial MenA and both engage in phylloquinone biosynthesis. Protein sequences were retrieved from the NCBI database by Blastp search, using representative proteins (see below) as the query sequences with the cutoff threshold of e-value 10^{-5} and >60% query identity coverage. Taxonomically redundant sequences were excluded. The sequences were aligned using MUSCLE v. 3.8.31. Phylogenetic analyses were performed using IQ-TREE v. 2.0.6. Bootstrap values (BP) were based on 1000 Ultrafast bootstraps. The experimental settings and source data are in the supplementary information. Accession numbers for representative proteins in the tree (in the order of the assigned numbers): COX10, O31652; HBST, BAC91369; UbiA, P0AGK1; MqnP, O28106; U2, ETA71034; DGGGPS, BAA29095; AuaA, CAL48953; LyeJ, Q9HPD9; DPPRS, ABB44494; U1, BBM82675; HPT, BAC88224; GGPP/GFPP cyclase, URN16965 (StsC); BCS, AAC84024; *ent*-copalyl diphosphate cyclase, ACO31274 (PtmT1); CnqPT1, AIT42139; MenA, AAC76912; and cyMenA, BAC89519.

(Fig. 10). In this review, their taxonomic distribution and phylogenetic positions were examined (Fig. 11). While ComQ is found in Firmicutes, XiaM, MoeN5, and NzsG are all confined to Actinobacteria. ComQ, MoeN5, and XiaM likely evolved from bacterial short-chain *trans*-IPPS, while NzsG evolved from CrtB. CrtB from Actinobacteria form the sister clade to NzsG that is also found only in Actinobacteria. Thus, NzsG seems to have emerged by CrtB gene duplication within the phylum. These *trans*-IPPS and SQS homologs are examples of functional divergence of enzymes that are involved in primary metabolism toward secondary metabolism. There are several additional examples that are comprehensively reviewed elsewhere (Rudolf et al. 2021).

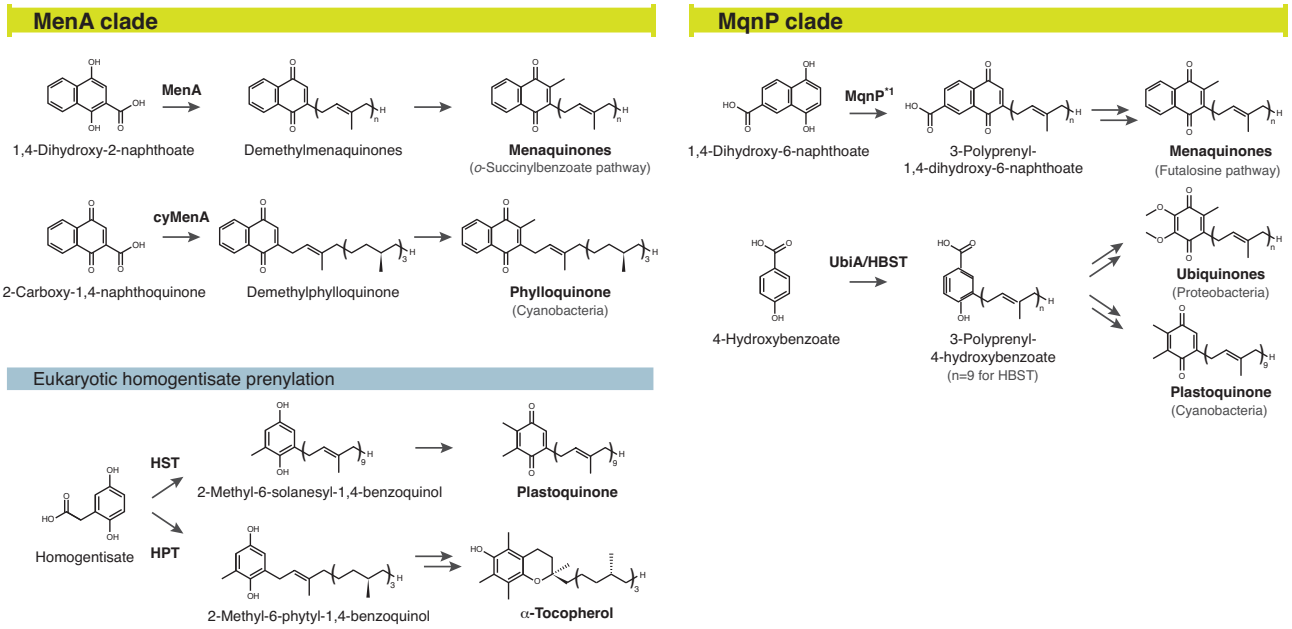

Figure 9. Major quinone prenylation pathways in Bacteria and Archaea. Eukaryotic prenylation pathway of homogentisate is also shown in comparison with the cyanobacterial plastoquinone pathway. ∗1: MqnP is proposed to catalyze the prenylation reaction in the figure, but it is not experimentally reconstituted yet (Joshi et al. 2018). Abbreviations: MenA, 1,4-dihydroxy-2-naphthoate PT; cyMenA, 2-carboxy-1,4-naphthoquinone PT; MqnP, putative 1,4-dihydroxy-6-naphthoate PT; UbiA, 4-hydroxybenzoate PT; HBST, 4-hydroxybenzoate solanesyl transferase; HST, homogentisate solanesyl transferase; and HPT, homogentisate phytyltransferase.

In vitro prenylation activity of trans-PT enzymes

Interestingly, some *trans*-PT enzymes that are not involved in prenylation *in vivo* are capable of catalyzing prenylation *in vitro*. Class I terpene cyclases are *trans*-PT enzymes that engage in cyclization in nature, but some members can catalyze the prenylation of unnatural substrates *in vitro* (He et al. 2020, Xu et al. 2021). It is currently not confirmed if prenylation also occurs *in vivo* by these cyclases. Together with the functional divergence from chain elongation to prenylation as described for ComQ, XiaM, and others in the previous section, these examples represent the functional plasticity of *trans*-PT proteins between chain elongation, prenylation, and cyclization. In fact, the transition between these three functions seems to have occurred many times for *trans*-PT proteins and others (Fig. 7).

Other innate PT families

Other innate PTs are mostly involved in secondary metabolism and seem to have late origins in specific microbial lineages (e.g. actinobacteria and cyanobacteria), apart from eukaryotes (e.g. fungi and plants) (Fig. 10). In contrast, protein PTs, which might be evolutionarily related to microbial terpene cyclases, play important roles in lipid–protein interactions in eukaryotes.

Cis-PT superfamily (cis-IPPS-like family)

Cis-PT enzymes that engage in prenylation are not common. Tuberculosinyladenosine synthase Rv3378c catalyzes the transfer of tuberculosinyl diphosphate (a bicyclic diterpene produced by class II terpene cyclase to adenosine, yielding tuberculosinyladenosine (Layre et al. 2014) (Fig. 10). This is the only prenylation activity that is experimentally confirmed for the *cis*-PT superfamily and is currently observed only in *Mycobacterium* species. The prenylation is associated with the host organism's virulence factor. The prenylation activity of *cis*-PT homologs is also predicted for the biosynthesis of nonribosomal oligopeptides in cyanobacteria (AgdJ; Pancrace et al. 2019). Rv3378c homologs are distributed outside the genus *Mycobacterium* as well. While their functions are not entirely understood, some homologs are possibly involved in unconventional diterpenoid biosynthesis (Stowell et al. 2022). For instance, a Rv3378c homolog from Chloroflexi participates in the *O*-methylkolavelool biosynthesis as a phosphatase (Nakano et al. 2015). In contrast, the glycerol *O*-prenylation activity is observed *in vitro* by an archaeal promiscuous *cis*-IPPS (Okada et al. 2021), but the presence of the prenylation activity with the native substrate is unclear.

ABBA family

The ABBA family is named after the presence of a characteristic $\alpha\beta\beta\alpha$ secondary structure element, which is also termed PT-barrel, among member proteins (Bonitz et al. 2011). The ABBA family comprises two distinct groups—the phenol/phenazine PT family (CloQ/NphB family) and the indole PT family (DMATS family). They have little sequence similarity, but share the same PT-barrel fold and a common ancestry has been suggested (Bonitz et al. 2011). The phenol/phenazine PT family prenylates various aromatic substrates, including phenols, phenazines, naphthalenes, and quinones, while the indole PT family prenylates mainly indole derivatives, including tryptophan residues in oligopeptides (Winkelblech et al. 2015) (Fig. 10). Prenylated quinones produced by ABBA enzymes are distinct from those produced by UbiA enzymes (e.g. NphB; Fig. 10) (Kuzuyama et al. 2005). ABBA enzymes are generally found in secondary metabolism. ABBA enzymes are soluble and active without divalent cations in most cases, unlike membrane-bound UbiA proteins that require Mg^{2+} ions for their catalysis (Tello et al. 2008). Prenylation occurs mostly to the aromatic ring via a mechanism similar to the Friedel–Crafts alkylation. However, nonaromatic moieties can also be prenyl acceptors. C_5 and C_{10} prenylations are dominant, although prenylations with a longer chain are also known (Takahashi et al. 2010). This contrasts with UbiA enzymes that prenylate mostly $C_{>15}$ chains.

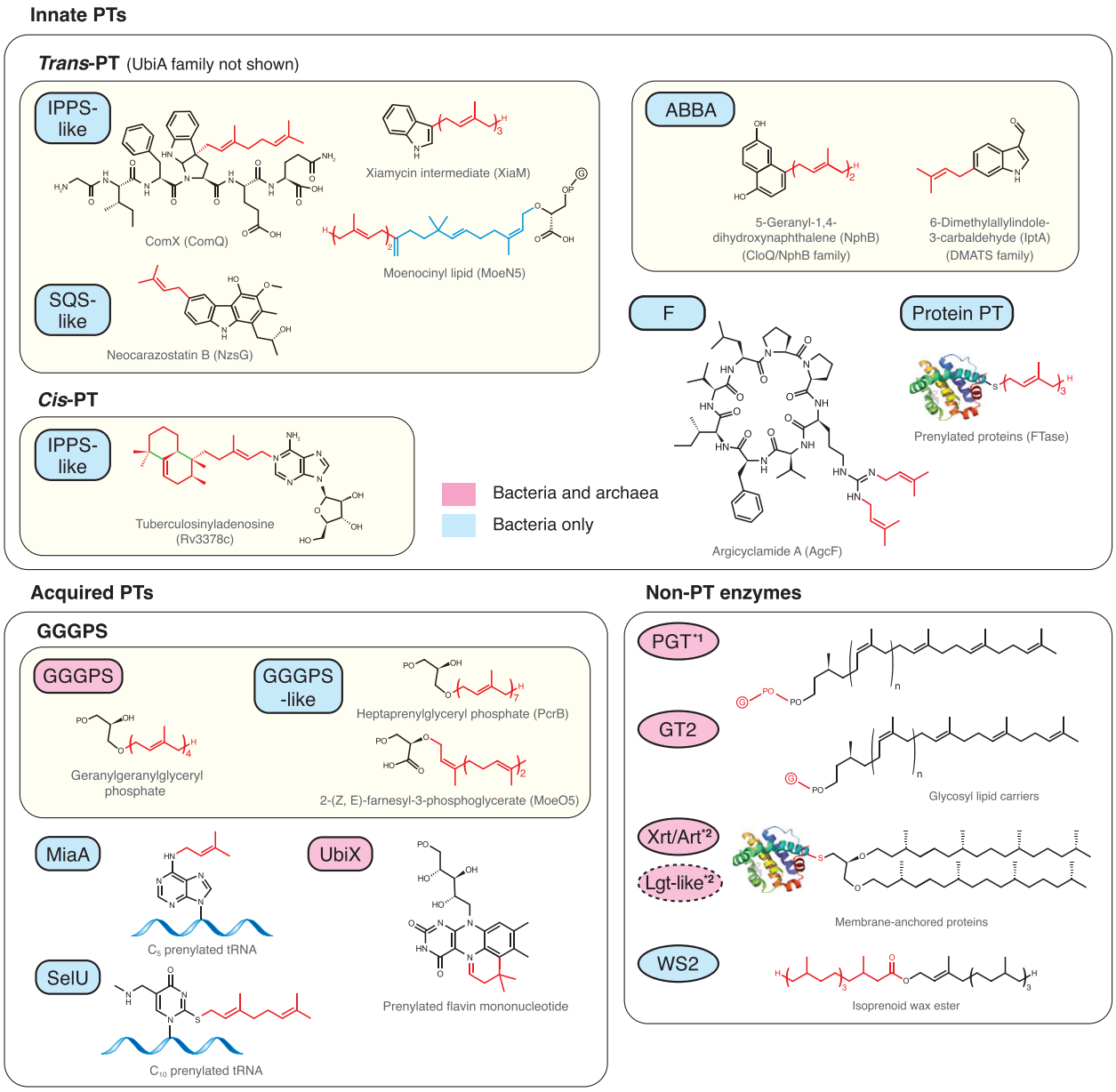

Figure 10. Representative prenylated products by innate PTs, acquired PTs and non-PT enzymes, excluding the UbiA family (shown in Fig. 8). Enzyme names are in brackets after compound names, in case the name is different from the enzyme family name. The directly or indirectly prenylated moiety is shown in red. The blue color indicates that the moiety is prenylated in a previous enzymatic step by MoeO5 (GGGPS-like in this figure). The green color indicates that the C–C bonds are created by a previous cyclization step by Rv3377c (class II terpene cyclase). ∗1: PGT contains two nonhomologous enzyme families—polytopic and monotopic PGTs. ∗2: only the membrane anchoring to the archaeal membrane can be seen as prenylation, while the anchoring in the bacterial membrane equals to fatty acylation. Also, Lgt-based membrane anchoring (lipobox pathway) is not yet identified in Archaea, although the presence of an analogous pathway is predicted. Abbreviations and accession numbers for representative proteins: ComX, quorum-sensing signal peptide; ComQ, tryptophan PT (P33690); XiaM, xiamycin biosynthesis enzyme M (CCH63739); MoeN5, moenomycin biosynthesis enzyme N5 (ABJ90163); NzsG, carbazole biosynthesis enzyme G (BBF24926); Rv3378c, tuberculosinyl adenosine transferase (P9WJ61); NphB, naphterpin biosynthesis enzyme B (Q4R2T2); CloQ, chlorobiocin biosynthesis enzyme Q (Q8GHB2); IptA, indole PT A (D6RT90); DMATS, dimethylallyltryptophan synthase; AgcF, argicyclamide biosynthesis enzyme F (BCR82584); FTase, farnesyl transferase (Q9LX33); PcrB, heptaprenylglyceryl phosphate synthase (O34790); MoeO5, moenomycin biosynthesis enzyme O5 (A0A011); G in a circle, glycosyl group; OP, phosphate group; and P in a circle, phosphate head group. See Supplementary Data 1 for the complete list of PT enzymes that are described in the review.

F family

F enzymes, which are named after the F genes for cyanobactin biosynthesis, are known for oligopeptide prenylation that is enzymatically distinct from protein prenylation (see the next section) (Zheng et al. 2022). Oligopeptide substrates include both ribosomal and nonribosomal peptides. Ribosomal peptides, or formally ribosomally synthesized and post-translationally modified peptides (RiPPs) (Arnison et al. 2013), are produced by the usual translation process by ribosome, while nonribosomal peptides are produced by the enzymatic activity of nonribosomal peptide synthetase (NRPS) (Wenski et al. 2022). For both groups, individual amino acid residues can be modified, including prenylation, fatty acylation, and others (McIntosh et al. 2009, Wenski et al. 2022). Lipidation generally increases hydrophobicity, stability and

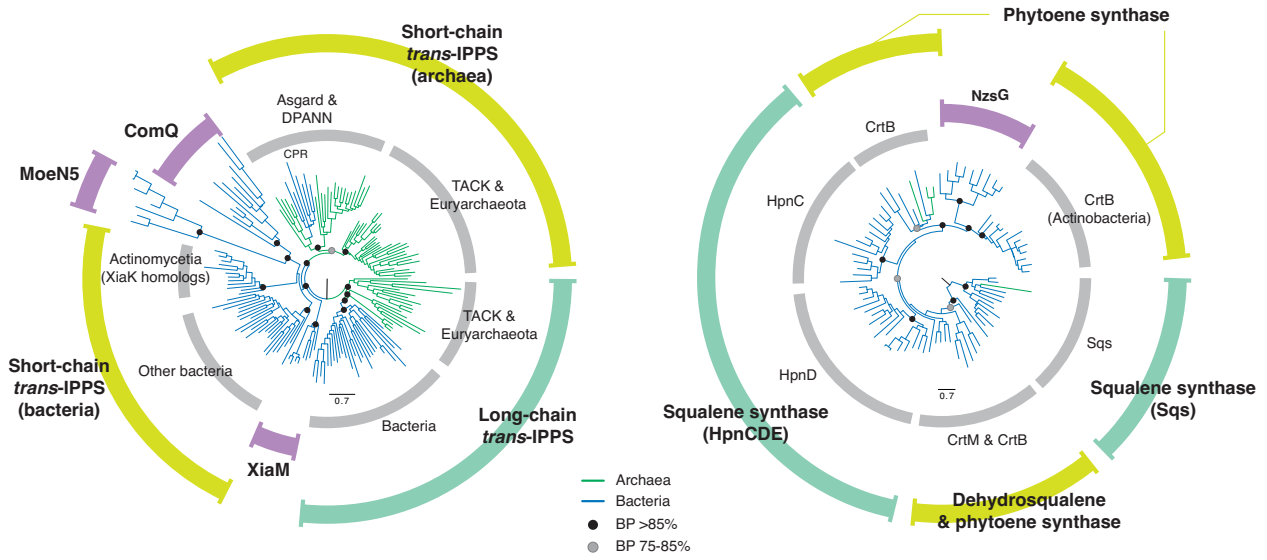

Figure 11. Maximum-likelihood phylogeny of PTs from the *trans*-IPPS and SQS families. Yellow and green colors indicate the ancestral functions of *trans*-IPPS/SQS proteins, while purple indicates divergent functions. Protein sequences were retrieved from the NCBI database by Blastp search, using representative proteins (see below) as the query sequences with the cutoff threshold of e-value 10^{-5} and >60% query identity coverage. Taxonomically redundant sequences were excluded. The sequences were aligned using MUSCLE v. 3.8.31. Phylogenetic analyses were performed using IQ-TREE v. 2.0.6. Bootstrap values (BP) were based on 1000 Ultrafast bootstraps. The experimental settings and source data are in the supplementary information. Accession numbers for representative proteins in the trees: short-chain *trans*-IPPS, P22939 (bacteria) and AAY79519 (archaea); long-chain *trans*-IPPS, P0AD57 (bacteria) and AAB89695 (archaea); ComQ, P33690; MoeN5, ABJ90163; XiaK, CCH63737; XiaM, CCH63739; HpnC, ABE49736; HpnD, ABE49737; CrtM, ACA82621; CrtB, BAC89685; Sqs, BAC91999; and NzsG, BBF24926. Abbreviation: XiaK, xiamycin biosynthesis enzyme K (short-chain *trans*-IPPS).

decreases polarity of substrate peptides. Oligopeptide prenylation is catalyzed by not only the F family, but also the ABBA and *trans*-PT (ComQ) families (Fig. 10). The prenylation activity of F enzymes has been extensively studied for cyanobactin, which is a class of RiPPs found in many cyanobacteria (Sivonen et al. 2010). Prenyl donors are in most cases DMAPP, but GPP is also used. Currently, prenylation is observed for tyrosine, tryptophan, serine, threonine, and arginine (Schultz et al. 2008, McIntosh et al. 2011, Zheng et al. 2022). An example for arginine prenylation is shown in Fig. 10 (Argicyclamide biosynthesis enzyme; AgcF). F enzymes are structurally similar to indole PTs in the ABBA family (Zheng et al. 2022). However, the characteristic PT barrel is absent in F enzymes, and thus the evolutionary relationship of the F family to the ABBA family is not clear.

Protein PT

Protein prenylation is observed for C_{15} and C_{20} chains (Jiang et al. 2018) (Fig. 10). All known protein prenylations are catalyzed by homologous PTs: farnesyl transferase (FTase) and geranylgeranyl transferases 1 and 2 (GGT-1 and GGT-2). These protein PTs are ubiquitously distributed in eukaryotes, but are nearly absent in Bacteria and Archaea. Instead, Bacteria and Archaea are observed or predicted to have an alternative (indirect) prenylation pathway by non-PT enzymes (see section Xrt/Art and Lgt-like families) (Eichler and Adams Michael 2005). Interestingly, the β subunit of protein PT is structurally similar to class II terpene cyclase that is widely distributed in Bacteria (Cao et al. 2010), although it is not clear if there is a direct evolutionary relationship between protein PT and class II terpene cyclase (see section Class II TC family).

Eukaryotic protein prenylation occurs mostly to a specific group of proteins: Ras superfamily (Marshall 1993). Ras proteins play important roles in cellular organization and signaling. Ras proteins are widely distributed in eukaryotes and are inferred to have been present in the last eukaryotic common ancestor (Rojas et al. 2012). Although the presence of Ras proteins is limited to eukaryotes, Ras-like homologs that lack the prenylation site have recently been discovered in Asgard archaea (Surkont and Pereira-Leal 2016). It is, thus inferred that eukaryotic Ras proteins are derived from archaeal Ras-like proteins, while prenylation of Ras proteins evolved during eukaryogenesis (Surkont and Pereira-Leal 2016). The prenylation by FTase and GGT-1 occurs to the cysteine residue in the CaaX motif at the C-terminus of the targeted proteins, while the prenylation by GGT-2 occurs to the cysteine residue in the CC/CXC motif at the C-terminus (Jiang et al. 2018). FTase and GGT-1 catalysis requires Zn^{2+} ions. Protein prenylation is distinct from other prenylations in that prenyl chains are added to the S atom of the cysteine residue. Attached prenyl groups are not simply lipid anchors, but directly involved in protein–protein interactions and protein–membrane interactions (Maurer-Stroh et al. 2003).

Acquired PT families

PTs in this category are derived from large protein superfamilies that have an extremely ancient origin and had ancestral functions that were irrelevant of terpenoids. Known examples include triosephosphate isomerase (TIM) barrels (Lang et al. 2000), nucleoside triphosphatases (NTPases) (Leipe et al. 2002), rhodaneses (Bordo and Bork 2002), and flavin-dependent proteins (Macheroux et al. 2011) (Fig. 7). Acquired PTs likely evolved from functional divergence of those ancient proteins, by accommodating terpenoids as substrates. PTs with a TIM barrel domain are indispensable for the archaeal membrane formation. PTs that are derived from NTPases and rhodaneses are involved in nucleic acid modification, while flavin-dependent PTs are involved in cofactor modification. Except for TIM barrel-derived PTs, acquired PTs have a function that is not related to membranes.

GGGPS family (TIM barrel)

The geranylgeranylglyceryl phosphate synthase (GGGPS) family contains TIM barrel proteins that transfer prenyl groups to oxygen of triose phosphates (Doud et al. 2011) (Fig. 10). The representative enzyme GGGPS catalyzes the first of the two GGPP transfers to G1P during the archaeal membrane lipid biosynthesis, followed by the second GGPP transfer by DGGGPS (Peterhoff et al. 2014). While DGGGPS is membrane-bound, GGGPS is cytosolic (Chen et al. 1993). GGGPS are universally distributed in Archaea, together with DGGGPS. Distant homologs of GGGPS (GGGPS-like proteins) are found in some bacteria, including heptaprenylglyceryl phosphate synthase PcrB in Bacillales and moenomycin biosynthesis enzyme MoeO5 in Streptomycetales (Fig. 10). PcrB catalyzes the transfer of C_{35} prenyl chains (heptaprenyl diphosphates) to G1P, instead of GGPP that is utilized in Archaea (Guldan et al. 2011). This is a rare example of utilizing long-chain isoprenyl diphosphates for a non-respiration purpose and also G1P that is rarely seen in Bacteria. However, the presence of PcrB-derived lipids in bacterial membranes is still not unclear. MoeO5 is involved in the transfer of FPP to 3-phosphoglyceric acid (Ostash et al. 2009). The presence of PcrB and MoeO5 homologs is limited to several discrete bacterial taxa, and thus these proteins likely diverged from more ancestral GGGPS. GGGPS may have been present in LUCA together with DGGGPS, although the evidence is not conclusive (see also section Membrane organization).

The utilization of C_{35} terpenoids (sesquarterpenoids) seems to be a characteristic of Bacillales (Ueda et al. 2015). Besides heptaprenylglyceryl phosphate and heptaprenyl menaquinone, Bacillales is also known to produce unique C_{35} pentacyclic compounds that are thought to be involved in the sporulation process to cope with environmental stress (see section Class II TC family) (Kontnik et al. 2008, Takigawa et al. 2010). It is not known what selective advantages sesquarterpenoids provide for host organisms, compared to other terpenoids.

MiaA and SelU families (NTPase and rhodanese)

tRNA dimethylallyltransferase (MiaA) and tRNA 2-selenouridine synthase (SelU) are involved in nucleic acid prenylation, although these two families are not homologous to each other (Fig. 10). MiaA catalyzes the transfer of DMAPP to the adenosine at position 37 of tRNAs for codons starting with uracil (Soderberg and Poulter 2000). The tRNA prenylation influences translational efficiency and fidelity. In eukaryotes, MiaA homologs are also involved in cytokinin (phytohormone) biosynthesis (Frébort et al. 2011). MiaA is universally distributed in Bacteria and is most likely ancestral in the domain. In contrast, MiaA is nearly absent in Archaea, and hence its presence in LUCA is not clear. Some bacteria also have SelU that transfers GPP to tRNA, which can affect codon bias and frameshift efficiency during translation as well as subcellular localization of RNA molecules (Dumelin et al. 2012). SelU has a high substrate specificity toward GPP and prenylations with other isoprenyl diphosphates have not been observed thus far (Szczupak et al. 2022). Both SelU homologs and GPP are not widely distributed in Bacteria and also are absent in Archaea. Therefore, it is currently inferred that SelU-based RNA modification evolved late in Bacteria.

UbiX family (Flavoprotein)

UbiX carries out the prenyl modification of FMN (White et al. 2015, Bloor et al. 2022) (Fig. 10). In Bacteria, UbiX adds DMAP or DMAPP to FMN and the prenylated FMN is used as the cofactor by 3-octaprenyl-4-hydroxybenzoate carboxylase (UbiD) to decarboxylate aromatic substrates during ubiquinone biosynthesis (Abby et al. 2020). Both UbiD and UbiX are flavoproteins. Ubiquinones are utilized for aerobic respiration and are found in proteobacteria (Pelosi et al. 2016) (Fig. 9). However, the UbiDX pair is in fact widespread in Bacteria including both aerobic and anaerobic species (Wang et al. 2018). The function of UbiDX is not fully understood yet, but the carboxylation/decarboxylation activity for both aromatic and nonaromatic compounds has been observed (Luo et al. 2010, Marshall et al. 2017).

UbiDX have a different function in Archaea. Archaeal UbiD homologs are in fact AMPD that participates in the MVA pathway during prenyl monomer biosynthesis (Hayakawa et al. 2018) (Route I; Fig. 2). Archaeal UbiD homologs utilize prenylated FMN produced by UbiX, similar to the bacterial UbiDX. UbiDX homologs are prevalent in Archaea, but their substrates (*trans*-anhydromevalonate 5-phosphate) are nonaromatic. The near ubiquity of UbiDX in both Bacteria and Archaea suggests that these two proteins were present in LUCA (see also section Terpenomes in and before LUCA). The UbiX family (or the homo-oligomeric flavin-containing cysteine decarboxylase family) is an ancient protein family that is involved in multiple cofactor biosynthesis, such as coenzyme A and tetrahydromethanopterin (Kupke et al. 2000, McNamara et al. 2014).

Non-PT enzymes

There are some enzymes that catalyze non-prenyl transfer reactions, but their substrates or acceptors encompass a prenyl moiety, and thus prenylation is indirectly achieved (Fig. 10). These enzymes are not PTs, but technically are e.g. acyltransferases and glycosyltransferases. Enzymatic products of these non-PT enzymes are generally not included in the terpenome, but they have important physiological roles that are distinct from those of typical terpenoids. Hence, the diversity of non-PT enzymes that act on prenyl substrates is briefly described in this section. For instance, *cis*-prenyl chains are conjugated with sugars (glycosyl groups) and are utilized for cell surface polymer biosynthesis. Also, membrane anchoring of proteins and other molecules in Archaea is another example of indirect prenylation, where the substrate is attached to the polar head group of archaeal lipid molecules that contain a prenyl moiety. Further, some acyltransferases that transfer acyl groups are capable of additionally transferring prenyl groups.

GT and PGT families

Compared to the enormous diversity of *trans*-prenyl chain prenylation by various PT families, prenylation of *cis*-prenyl chains is nearly absent in nature, except for the rare example of DPPRS in the UbiA family. While *cis*-prenyl chains are not prenylated by PT, they are conjugated with carbohydrates by glycosyltransferase (GT) or phosphoglycosyl transferase (PGT) (Jones et al. 2009, Hartley and Imperiali 2012) (Fig. 12). Conjugation of prenyl groups with other molecules marks the branching point for *trans*- and *cis*-prenyl chains to connect to different metabolic networks. $C_{>30}$ *cis*-prenyl chains seem to be dedicated to sugar transport during cell envelope biogenesis and extracellular protein modification. Glycosylated *cis*-prenyl chains act as glycosyl lipid carriers and provide the scaffolds for the biosynthesis of cell envelope polymers such as peptidoglycans (polysaccharide cell walls), O-antigens and teichoic acids in Bacteria and glycoproteins in both Bacteria and Archaea through glycosylation (Silhavy et al. 2010, Albers and Meyer 2011). Two major modes of glycosylation are

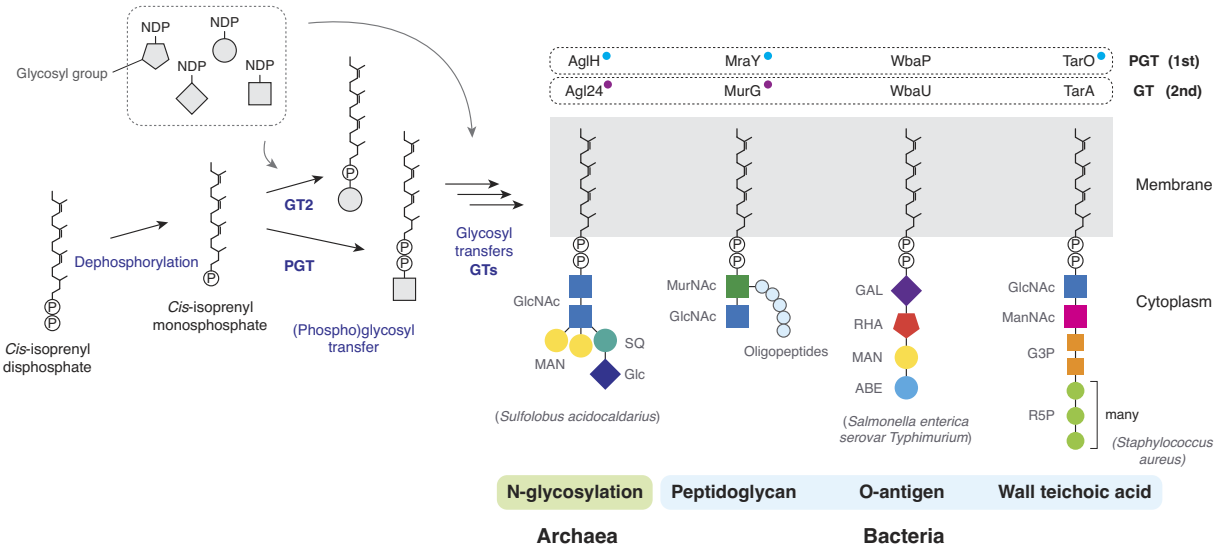

Figure 12. Schematic diagram of cell surface polymer biosynthesis (cytoplasmic phase). Glycosyl lipid carriers consist of the *cis*-isoprenyl chain moiety and the glycosyl moiety that are conjugated with each other via phosphate groups. The isoprene unit in the α-position is generally saturated for Archaea (as shown in the figure), while it remains unsaturated for Bacteria (Jones et al. 2009). The glycosyl moiety is first constructed in the cytoplasmic side of membranes, then flipped to the opposite side, further polymerized and/or modified, and added to pre-existing surface structures. Substantial structural variations are present in the glycosyl moiety and only representative examples are shown. Phosphoglycosyl transferases (PGTs) and glycosyltransferases (GTs) that catalyze the first two steps of the glycosyl oligomer formation are shown on top of individual molecules. Blue and purple circles next to the enzyme names indicate that the enzymes are homologous to each other and belong to the polytopic PGT family and the GT1/28 family, respectively. Polytopic PGT is not a GT, even though it is classified as Family 4 GT in InterPro (IPR000715). For the details of individual enzymes, see Swoboda et al. (2010), Jarrell et al. (2014), Egan et al. (2020), and Whitfield et al. (2020). Abbreviations: NDP, nucleotide diphosphate; P, phosphate; GT2, family 2 GT; GlcNAc, N-acetylglucosamine; MAN, mannose; SQ, 6-sulfoquinovose; Glc, glucose; MurNAc, N-acetylmuramic acid; GAL, galactose; RHA, Rhamnose; ABE, abequose; ManNAc, N-acetylmannosamine; G3P, glycerol-3-phosphate; R5P, ribitol-5-phosphate; AglH, GlcNAc–1–P transferase (archaeal glycosylation enzyme H); Agl24, GlcNAc transferase; MraY, MurNAc-1-P transferase; MurG, GlcNAc transferase; WbaP, GAL-1-P transferase; WbaU, RHA transferase; TarO, GlcNAc-1-P transferase; and TarA, ManNAc transferase. Accession numbers for representative proteins: AglH, P39465; Agl24, Q4J9C3; MraY, P0A6W3; MurG, P17443; WbaP, P26406; WbaU, P26402; TarO, X5EH87; TarA, Q2G2L3; and GT2, D4GUA0.

N- and O-glycosylation. N-glycosylation is the glycosyl transfer to the nitrogen atom of a protein and is predominantly observed in Archaea (Jarrell et al. 2014). In contrast, O-glycosylation, which is the glycosyl transfer to the oxygen atom of a protein, is more common in Bacteria (Nothaft and Szymanski 2010, Iwashkiw et al. 2013).

GTs and PGTs are membrane-associated proteins and transfer glycosyl and phosphoglycosyl groups, respectively, to *cis*-isoprenyl monophosphates that are derived from *cis*-isoprenyl diphosphates by dephosphorylation (Fig. 12). The glycosyl group is in turn provided from nucleotide diphosphate (NDP) sugars. (Phospho) glycosyl transfer to *trans*-prenyl chains is not known thus far. Currently, one GT family (Family 2; GT2) and two PGT families are known to catalyze the (phospho)glycosyl transfer to *cis*-prenyl substrates and each family is evolutionarily independent (Allen and Imperiali 2019). While GT2 encompasses enzymes of various functions, such as well-known chitin synthase and cellulose synthase (Taujale et al. 2020), a fraction of GT2 proteins transfer monosaccharides (glycosyl monomers) such as mannose and glucose to prenyl substrates. These monosaccharides are utilized to either extend pre-existing oligosaccharide moieties, or directly modify proteins.

In contrast, PGT proteins transfer phosphorylated glycosyl monomers (Fig. 10). PGTs are classified into monotopic and polytopic PGTs based on the topology of the catalytic core structure, but these two families are functionally overlapping (Allen and Imperiali 2019). PGTs catalyze the first step of cell envelope polymer biosynthesis, while subsequent additions of glycosyl monomers to pre-existing oligomers are performed by various GTs (Fig. 12). For instance, the initial step of O-antigen biosynthesis is catalyzed by a monotopic PGT (WbaP) (Whitfield et al. 2020). In contrast, the first step of peptidoglycan cell wall and wall teichoic acid biosynthesis is catalyzed by polytopic PGTs (MraY and TarO) (Swoboda et al. 2010, Egan et al. 2020). Similarly, the first step of the glycosyl moiety formation for N-glycosylation in Archaea is catalyzed by a polytopic PGT (AglH) (Jarrell et al. 2014). While monotopic PGTs are largely confined to Bacteria (O'Toole et al. 2021), polytopic PGTs are widely distributed in both Bacteria and Archaea. The evolution of polytopic PGT is closely associated with the origin of cell surface structures and is discussed in more details in section Cell surface biogenesis.

Xrt/Art and Lgt-like families

Many membrane-associated proteins are anchored to membrane lipids (membrane anchoring). Membrane anchoring is a type of protein lipidation and is an important process to modify the effects of protein–protein and protein–membrane interactions. Anchored proteins are classified in lipoproteins. Bacteria and Archaea are known or predicted to have membrane anchoring pathways that are distinct from the eukaryotic anchoring pathway and the protein lipidation pathway mediated by protein PTs (Eichler and Adams Michael 2005, Pohlschroder et al. 2018, Braun and Hantke 2019). In the microbial anchoring pathway, proteins are covalently anchored to the head group of membrane lipid molecules. Therefore, proteins are indirectly attached to either the prenyl group in the archaeal membrane, or the fatty acyl group in the bacterial membrane. The former can be seen as prenylation, while the latter equals to

fatty acylation. Two anchoring pathways are known thus far: lipobox and exosortase/archaeosortase pathways. However, both pathways are not universally distributed in Bacteria and Archaea, and thus the presence of those pathways in LUCA is not conclusive.

In the lipobox pathway, membrane anchoring of proteins (lipoprotein formation) is catalyzed by a set of enzymes—prolipoprotein diacylglyceryl transferase (Lgt), lipoprotein signal peptidase (Lsp), and in some cases lipoprotein *N*-acyltransferase (Lnt) (Nakayama et al. 2012). Lgt recognizes a unique signal peptide called the lipobox in the substrate proteins, which are eventually attached to the head group of membrane lipids (Fig. 10). The lipobox pathway is widely distributed in Bacteria and is likely ancestral in the domain (Sutcliffe et al. 2012). In contrast, the pathway is not identified in Archaea and no homologs for Lgt, Lsp, and Lnt have been detected in the domain thus far. Yet, lipobox-containing proteins are readily found in Archaea and the presence of lipoproteins has experimentally been suggested (Storf et al. 2010). This implies the presence of an unknown lipobox-based pathway to form lipoproteins in Archaea and perhaps even in LUCA. However, the presence of lipoproteins seems to be limited mostly to Euryarchaeota (Storf et al. 2010), and thus its ancestry in the last archaeal common ancestor is not clear. If lipoprotein formation is in fact an ancestral trait in LUCA, the observed difference between Bacteria and Archaea may reflect the adaptation of lipoprotein biosynthesis toward distinct lipid membranes in the two domains (Pohlschroder et al. 2018).

Another anchoring pathway is the exosortase (Xrt)/archaeosortase (Art) pathway. Xrts and Arts are transmembrane proteins and are distantly related to each other. Xrts are found in Bacteria and act on proteins that possess the C-terminal PEP motif, while Arts are found in Archaea and act on proteins that possess the C-terminal PGF motif (Haft et al. 2012). Similar to the lipobox pathway, substrate proteins are attached to the membrane lipid head group (Fig. 10). The distribution of Art is limited mostly to Euryarchaeota in Archaea, similar to the lipobox pathway (Pohlschroder et al. 2018). Hence, its ancestry in the archaeal common ancestor is not clear.

WS2 family

Acyltransferase (AT) is principally involved in fatty acylation. However, an AT enzyme (wax ester synthase 2; WS2) from a marine gammaproteobacterium produces prenyl wax esters by adding C_{16} acyl-CoA or C_{20} phytanoyl-CoA to C_{20} phytol (Holtzapple and Schmidt-Dannert 2007) (Fig. 10). WS2 can produce prenyl-only esters, hybrid esters of the acyl, and prenyl groups, and "normal" acyl-only esters. Accumulation of prenyl wax esters is observed in some marine bacteria, and thus this bifunctional AT may be more widespread in Bacteria (Röttig and Steinbüchel 2013). WS2 belongs to the AtfA (WS/DGAT) family that is involved in the acyl wax ester biosynthesis in Bacteria (Röttig and Steinbüchel 2013). Thus, WS2 likely functionally diverged from conventional AtfA enzymes.

Cyclization

Terpenoid cyclization is predominantly observed in Bacteria, apart from eukaryotes. The cyclization reaction is catalyzed by multiple evolutionarily independent groups of enzymes with distinct mechanisms from one another (Fig. 13). Historically, two types of cyclases are most extensively studied—class I and II terpene cyclases (TCs). These TCs catalyze the cyclization of *trans*-prenyl chains and produce the vast majority of the terpenoid structural diversity in nature. In particular, the class II cyclization of squalene and phytoene produces an important class of membrane-bound polycyclic terpenoids: hopanoids, steroids, and carotenoids. They play crucial roles in membrane organization and cell signaling (e.g. steroid hormones in eukaryotes) (Schiffer et al. 2019). There are several excellent reviews for the details about the biosynthesis and physiological functions of hopanoids, steroids and carotenoids (Nes 2011, Belin et al. 2018, Schiffer et al. 2019, Sandmann 2021). Instead, this review focuses on the evolutionary relationship of cyclases. In addition to class I and II TCs, there are many small cyclase families that have more recent origins but distinct reaction mechanisms. Only lycopene cyclase that catalyzes the carotenoid cyclization is found in both Bacteria and Archaea, although its origin is likely in Bacteria. Also, only one cyclase family (GrsAB) that cyclizes GDGTs is currently characterized only for Archaea. All other cyclases are specific to Bacteria. Hence, no cyclase was likely present in LUCA, contrary to enzymes for chain elongation and prenylation. In most cases, homologs of bacterial cyclases are also found in eukaryotes, although the products may be different. Similar to prenylation, cyclization can be catalyzed by both *trans*- and *cis*-PT enzymes and also several non-PT enzymes.

Trans-PT superfamily

TCs from the *trans*-PT superfamily exclusively produce secondary metabolites and their product range substantially varies depending on species. The physiological functions of cyclized products include cytotoxicity and antimicrobial and antiparasitic activities (Rudolf et al. 2021). The taxonomic distribution of TCs is mostly limited to aerobic and terrestrial species (e.g. Actinobacteria, Cyanobacteria, and Chloroflexi) (Yamada et al. 2015, Reddy et al. 2020). Hence, although the ultimate origin of those TCs may be ancient, its diversification was at the earliest only after Earth's first oxygenation around 2.4 billion years ago (Ga) (Lyons et al. 2014) and was more likely to be later.

Class I and IB TC families

Class I cyclization can be seen as an intramolecular prenyl transfer reaction. Class I TC is in fact evolutionarily related to *trans*-IPPS and retains two Asp-rich metal binding motifs (α-fold) (Fig. 14). Class I cyclization is known for C_{10}–C_{25} *trans*-prenyl chains (*trans*-isoprenyl diphosphates) in Bacteria (Dickschat 2016, Hou and Dickschat 2020) (Fig. 13). Class I cyclization of *cis*-isoprenyl diphosphates is currently observed only in eukaryotes (Gericke et al. 2020). Similarly, class I cyclization of $C_{>30}$ *trans*-chains is known only for eukaryotes thus far (Li and Gustafson 2021, Tao et al. 2022). In contrast, tetraprenyl-β-curcumene synthase BsuTS (YtpB) cyclizes C_{35} *trans*-heptaprenyl diphosphate by a class I-like mechanism in Bacillales (class IB or "large" TC; Sato et al. 2011, Fujihashi et al. 2018). Class IB TC is structurally similar to class I TC, although the sequence similarity is low (Rudolf and Chang 2020). Not all homologs of class IB TC perform cyclization (e.g. BalTS; Rudolf and Chang 2020).

Class I TCs are highly divergent in terms of sequence and function, but seem to share a common ancestry (Fig. 14). Also, class IB TCs form the sister clade to class I TCs. Thus, the vast majority of microbial class I-type cyclizations share a common origin, except for secondary-derived class I cyclases within the UbiA family (see the next section). Both class I and IB TCs are soluble proteins, similar to *trans*-IPPS, while SQS and UbiA family proteins are membrane-bound. Hence, class I and IB TCs may have evolved

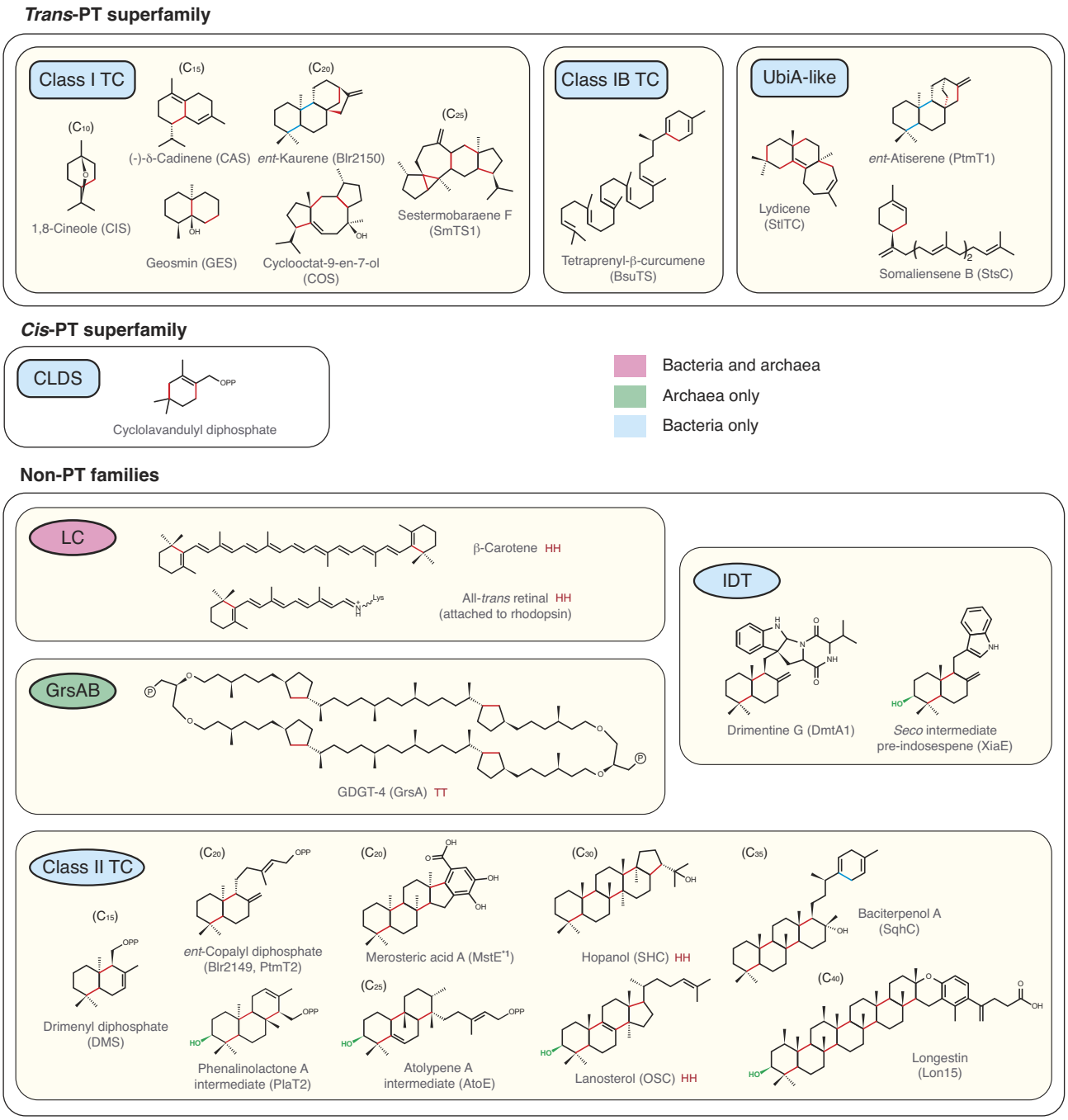

Figure 13. Terpenoids cyclized by PT and non-PT family enzymes. Direct cyclization products or final products of terpenoid gene clusters are shown. Cyclase names are in brackets after compound names, in case the name is different from the enzyme family name. Red lines indicate the C–C bonds created by cyclization reactions. Blue lines indicate the C–C bonds created by a previous cyclization step: Blr2149 (class II TC)/Blr2150 (class I TC), PtmT2 (class II TC)/PtmT1 (UbiA-like), and BsuTS (class IB TC)/SqhC (class II TC). Green colors indicate that the hydroxyl groups are derived from the epoxy groups in the substrates. HH and TT indicate that the precursor compounds are derived from the HH condensation and the TT condensation, respectively, while other products without a sign are derived from HT condensation precursors. *1: the canonical $\beta\gamma$ domains are not conserved and only the β domain is present in MstE. Abbreviations and accession numbers for representative proteins: CIS, 1,8-cineole synthase (B5GMG2); CAS, (-)-δ-cadinene synthase (B5GS26); GES, geosmin synthase (Q9X839); Blr2150, *ent*-kaurene synthase (Q45222); COS, cyclooctat-9-en-7-ol synthase (BAI44338); SmTS1, *S. mobaraensis* terpene synthase 1 (QTT76613); BsuTS (YtpB), tetraprenyl-β-curcumene synthase (O34707); StlTC, lydicene cyclase (WP_046929223); PtmT1, *ent*-atiserene synthase (ACO31274); StsC, somaliensenes cyclase (URN16965); CLDS, cyclolavandulyl diphosphate synthase (X5IYJ5); LC, lycopene cyclase (P21687; CrtY); lys, lysine residue; GrsA, GDGT ring synthase A (Q4J8I0); IDT, indole diterpene synthase; DmtA1, drimentine synthase (AVP32200); XiaE, xiamycin cyclase (I7KIT8); DMS, drimenyl diphosphate synthase (KKZ71921); Blr2149 and PtmT2, *ent*-copalyl diphosphate synthase (BAC47414 and ACO31276); PlaT2, phenalinolactone cyclase (EGJ74238); MstE, merosterol synthase (ATN39899); AtoE, atolypene cyclase (SED33305); SHC, squalene–hopene cyclase (P33247); OSC, oxidosqualene cyclase (P48449); SqhC, tetraprenyl-β-curcumene cyclase (Q796C3); and Lon15, longestin cyclase (BAF98632). See Supplementary Data 1 for the complete list of TC enzymes that are described in the review.

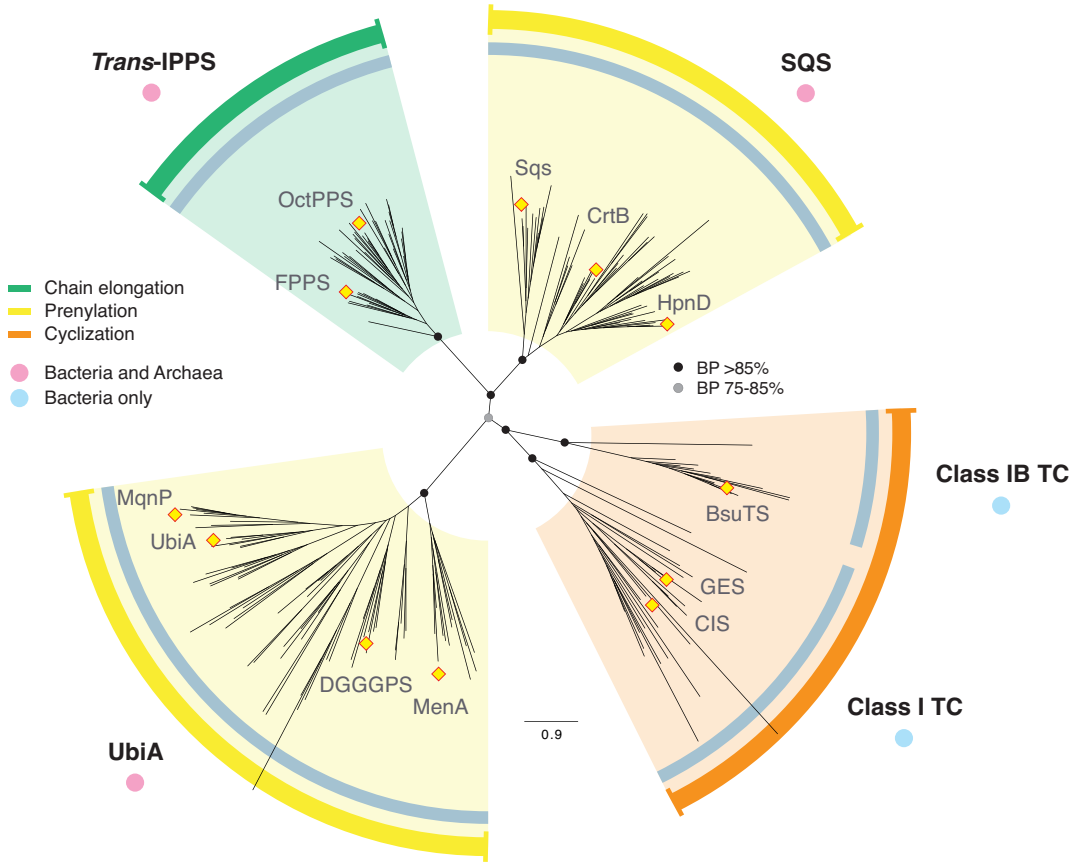

Figure 14. Evolutionary relationship of *trans*-PT superfamily proteins. Protein sequences for *trans*-IPPS, UbiA, and SQS were selected from the datasets in Figs 7 and 8. Functionally divergent proteins (ComQ, XiaM, MoeN5, and UbiA-like cyclase) are not included (see Figs 8 and 11). It is noted that *trans*-PT proteins share the catalytic α fold and are structurally homologous to each other, but the overall sequence similarity is nearly absent between clusters. Protein sequences for class I and IB TCs were retrieved from the NCBI database by PSI-Blastp search, using representative proteins (see below) as the query sequences with the cutoff threshold of e-value 10^{-5} and >60% query identity coverage. Taxonomically redundant sequences were excluded. The sequences were aligned using MUSCLE v. 3.8.31. Phylogenetic analyses were performed using IQ-TREE v. 2.0.6. Bootstrap values (BP) were based on 1000 Ultrafast bootstraps. The experimental settings and source data are in the supplementary information. Diamonds indicate the positions of representative proteins. Abbreviations and accession numbers for representative proteins: FPPS, farnesyl diphosphate synthase (P22939); OctPPS, octaprenyl diphosphate synthase (P0AD57); Sqs, squalene synthase (BAC91999); CrtB, phytoene synthase (BAC89685); HpnD, presqualene diphosphate synthase (ABE49737); BsuTS, tetraprenyl-β-curcumene synthase (O34707); GES, geosmin synthase (Q9X839); CIS, 1,8-cineole synthase (B5GMG2), MenA, 1,4-dihydroxy-2-naphthoate PT (AAC76912); DGGGPS, digeranylgeranylglyceryl phosphate synthase (BAA29095); UbiA, 4-hydroxybenzoate PT (P0AGK1); and MqnP, putative 1,4-dihydroxy-6-naphthoate PT (O28106).

from *trans*-IPPS. Class I TCs are found also in eukaryotes, in particular fungi and land plants (Quin et al. 2014, Pichersky and Raguso 2018). Bacterial and eukaryotic TCs diverge deeply with each other (Smanski et al. 2012), but it is hypothesized that plant TCs are derived from bacterial TCs (Cao et al. 2010). Some eukaryotes have an additional class I-type cyclase family within the *trans*-IPPS family (Wei et al. 2020).

UbiA family

While the UbiA family is primarily involved in the prenylation of various aromatic and nonaromatic substrates, some UbiA proteins are capable of cyclization through a class I TC-like mechanism. UbiA-like cyclases are membrane-bound, similar to prenylating UbiA enzymes, but do not always retain the canonical Asp-rich motifs (Rudolf and Chang 2020). UbiA-like cyclases form two separate clades among UbiA enzymes (Fig. 8), and hence those cyclases are suggested to have evolved twice independently from conventional PTs. UbiA-like cyclases are found to be involved in C_{20} and C_{25} terpenoid biosynthesis in some bacteria. For instance, StlTC and StsC from *Streptomyces* cyclize GGPP and GFPP, yielding tricyclic lydicene and monocyclic somaliensene, respectively (Yang et al. 2017, 2018) (GGPP/GFPP cyclase; Fig. 13). In contrast, PtmT1 from *Streptomyces platensis* spp. catalyzes the cyclization of *ent*-CPP, following the canonical class II TC cyclization of GGPP, and produce *ent*-atiserene (Smanski et al. 2011) (*ent*-CPP cyclase). *ent*-CPP is also cyclized by class I TCs, but the product is different (*ent*-kaurene; Fig. 13).

Cis-PT superfamily

Cis-PT enzymes are primarily involved in chain elongation, but one bacterial *cis*-PT enzyme is known to catalyze a class I TC-like cyclization (Rudolf and Chang 2020). Cyclolavandulyl diphosphate synthase (CLDS) from *Streptomyces sp. CL190* catalyzes both the HM condensation of two DMAPP molecules and its cyclization into a C_{10} monocyclic terpenoid (Ozaki et al. 2014) (Fig. 13). The cyclized product is further prenylated to other compounds and the prenylated products have an antimicrobial activity similar to the products of *trans*-PT-type TCs (Ozaki et al. 2014).

Non-PT cyclase families

LC family

Lycopene cyclase (LC) is involved in C_{40} carotenoid cyclization and consists of four subfamilies: CrtY, CrtL, CruA, and $CrtY_{cd}$ (Maresca et al. 2007, Klassen 2010) (Fig. 13). LC proteins are membrane-bound flavoproteins and are all distantly related to each other. However, it is not known if LC evolved once or multiple times. LC cyclizes lycopene that is a desaturated form of phytoene, which is in turn produced from GGPP by phytoene synthase (CrtB; SQS family). Some enzymes form only one cyclohexane ring at one end of the prenyl chain (γ-carotene), while others form two rings at both ends of the chain (α- and β-carotenes). The cyclization of dehydrosqualene (C_{30} equivalent of phytoene) is not known. Carotenoids are important light-harvesting pigments and also function as photoprotection and antioxidation agents (Sandmann 2021). Further, carotenoids have membrane regulation roles (Bramkamp and Lopez 2015). LC is found in phototrophic bacteria and archaea as well as some non-photosynthetic bacteria. However, the distribution in Archaea is mostly limited to haloarchaea (Sandmann 2021). Phylogenomic analyses suggest archaeal carotenoid biosynthesis is horizontally transferred from Bacteria (Klassen 2010). Carotenoids are derived from HH condensation products, and thus the evolution of LC is associated with that of SQS (see also section Phototrophy) (Fig. 5).

Enzymatic cleavage of C_{40} cyclized carotenoids by carotenoid oxygenase leads to the production of retinal (Daruwalla et al. 2020) (Fig. 13). This C_{20} apocarotenoid is embedded in the photosensing protein rhodopsin as the chromophore, constituting the retinal-based phototrophy system. Light absorption causes the isomerization of retinal from the all-*trans*-configuration to the 13-*cis*-configuration (Ruch et al. 2005). Retinal formation is observed for two nonhomologous dioxygenase families: Diox1 and Blh families. Diox1 enzymes cannot cleave bicyclic carotenoids such as carotene, but instead perform asymmetrical cleavages of various apocarotenoids, producing retinal among others (Ahrazem et al. 2016). In contrast, Blh enzymes perform the symmetrical cleavage of carotene, producing two molecules of retinal (Kim et al. 2009). The evolution of retinal and its phototrophy is discussed in further details in section Phototrophy.

GrsAB family

GDGT ring synthase (GrsAB) is only cyclase family that is currently characterized only for Archaea. GrsAB are responsible for the cyclization of archaeal biphytanyl lipid chains and hence the biosynthesis of GDGTs, together with TES (Zeng et al. 2019) (Fig. 13). Both TES and GrsAB are radical SAM-dependent enzymes. The cyclopentane ring formation by GrsAB is achieved through a free radical mechanism. The substrates of GrsAB are archaeal membranes themselves and thus GrsAB enzymes are likely membrane-bound, but this is not experimentally confirmed. The number of rings in GDGTs is associated with the growth temperature of host organisms and thus GDGTs are utilized as an environmental proxy (Schouten et al. 2013).

Class II TC family

Class II cyclization is known for C_{15}–C_{40} *trans*-prenyl chains (Rudolf et al. 2021, Pan et al. 2022) (Fig. 13). Class II TC initiates the cyclization by protonating the terminal double bond or the epoxy group, in concert with the cascade of C–C bond formation, generating polycyclic structures in a single enzymatic step (Christianson 2017). Class II TCs cyclize either HT condensation products (C_{15}–C_{25} and C_{35}–C_{40} *trans*-isoprenyl diphosphates and their derivatives), or HH condensation products (C_{30} squalene and its epoxide derivative oxidosqualene). Class II cyclization of the C_{30} HT condensation product (hexaprenyl diphosphate) is not known. Class II cyclization of C_{15}–C_{25} *trans*-isoprenyl diphosphates leaves the diphosphate group intact, and thus the reaction products are usually further modified, including dephosphorylation, class I cyclization and prenylation to other compounds (Peters 2010, Dickschat 2019, Kim et al. 2019, Vo et al. 2022). In contrast, the diphosphate end of C_{35} and C_{40} *trans*-isoprenyl diphosphates are modified prior to class II cyclization by class IB cyclase (C_{35}) or UbiA-type PT (C_{40}) (Dairi 2013, Sato 2013). Squalene–hopene cyclase (SHC) and oxidosqualene cyclase (OSC) are two major C_{30} TCs that are most extensively studied among class II TCs. These enzymes are integral membrane proteins, which is consistent with the hydrophobic nature of their substrates (Wendt et al. 1997, Thoma et al. 2004). By contrast, C_{20} TCs and their substrates are soluble (Rudolf et al. 2016). C_{20} and C_{30} cyclizations are most common, while other cyclizations are rare (Sato 2013, Ozaki et al. 2018, Kim et al. 2019, Pan et al. 2022).

Cyclized products of C_{15}–C_{40} individual terpene classes have unique physiological functions. C_{15}–C_{25} products are typically involved in cell signaling and antimicrobial activity (Rudolf et al. 2021). In contrast, C_{30} products seem to be specialized in membrane regulation. Class II cyclization of squalene and oxidosqualene generates well-known hopanoids and steroids (Xu et al. 2004). In eukaryotes, steroids are additionally utilized as signaling agents (steroid hormones) (Schiffer et al. 2019). C_{35} cyclization is rare, but baciterpenol in Bacillales is possibly involved in the sporulation process (Sato et al. 2011, Rudolf and Chang 2020). The interesting relationship between Bacillales species and C_{35} sesquarterpenoids is already described above (section GGGPS family). C_{40} cyclization is also rare and is known only for the antibiotic agent longestin (KS-505a) (Ishiyama et al. 2008, Dairi 2013).

All class II TCs are homologous to each other, but the evolutionary relationship between different terpene classes is not well-understood. Class II TC is not homologous to class I TC that belongs to the *trans*-PT superfamily and thus has a distinct evolutionary origin in Bacteria. The characteristic $(\alpha/\alpha)_6$ barrel fold (β-fold) in class II TC has a structural similarity to several other enzyme families that are not necessarily involved in terpenoid metabolism—protein PT, nisin cyclase, endoglucanase, and bacterial cellulase (Alzari et al. 1996, Park et al. 1997, Sakon et al. 1997, Li et al. 2006). It is not known if these enzymes are evolutionarily related to one another. Blast search retrieves a large number of divergent homologs in both Bacteria and Archaea (Fig. 15 for selected sequences), but there is no amino acid residue that is conserved among all homologs. For instance, an early-branching archaeon *Methanopyrus kandleri* possesses as many as seven homologs (see Table S2, Supporting Information). However, no archaeal species is known to produce cyclized terpenoids and it is not likely that these distant homologs are involved in terpenoid cyclization. Regardless of the function, the near ubiquity of β-fold-bearing proteins in both Bacteria and Archaea suggests that a β-fold-like motif was present in LUCA. Class II TCs form several distinct clades and it is currently unclear if class II cyclization evolved once or multiple times independently (Fig. 15). C_{30} and C_{35} TCs are monophyletic, and thus a single origin for these two TC families is inferred. In contrast, C_{20} TCs are found in two separate clades and thus may have evolved twice independently.

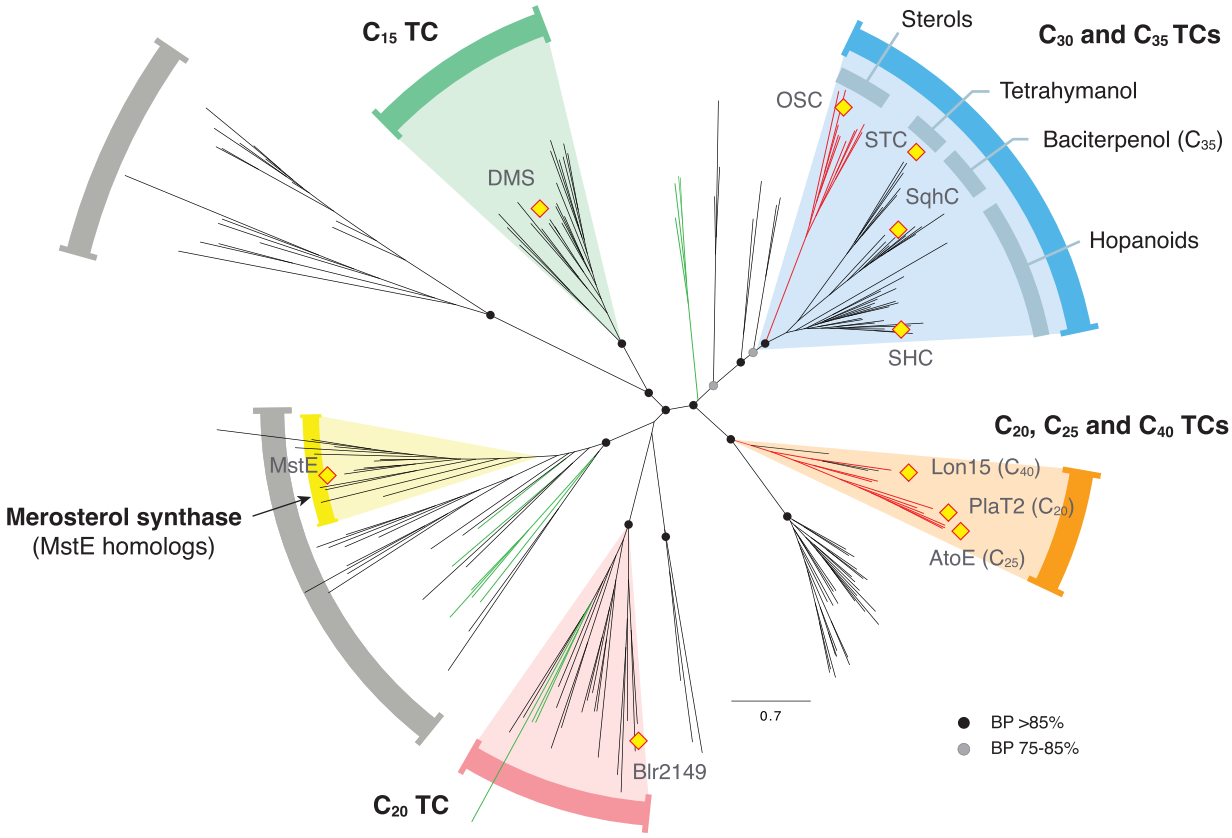

Figure 15. Evolutionary relationship of class II TCs and selected homologs. The product profile is also shown for C_{30} and C_{35} TCs. Gray-curved bars indicate that proteins consist of solely the β domain, while all other proteins consist of the $\beta\gamma$ or $\beta\gamma$-like domains. Red lines indicate that host organisms additionally have FMO. Green lines indicate archaeal proteins, while black lines indicate bacterial proteins. The function of archaeal proteins in the C_{20} TC cluster is unknown. The taxonomic distribution of MstE outside cyanobacteria is not known, and thus the clustering of MstE (light yellow area) is arbitrarily limited to cyanobacteria and several species that have close homologs. Protein sequences were retrieved from the NCBI database by PSI-Blastp search, using representative proteins (see below) as the query sequences with the cutoff threshold of e-value 10^{-5} and >60% query identity coverage. Taxonomically redundant sequences were excluded. The sequences were aligned using MUSCLE v. 3.8.31. Phylogenetic analyses were performed using IQ-TREE v. 2.0.6. Bootstrap values (BP) were based on 1000 Ultrafast bootstraps. The experimental settings and source data are in the supplementary information. The multiple sequence alignment was performed only for the β domain that is shared by all divergent homologs. C_{15} TCs that have a divergent β domain and an adjacent haloacid dehalogenase-like domain are not included since the β domain has nearly no sequence homology to other TCs. Diamonds indicate the positions of representative proteins. Abbreviations and accession numbers for representative proteins: DMS, drimenyl diphosphate synthase (KKZ71921; C_{15} TC); MstE, merosterol synthase (ATN39899); Blr2149, *ent*-copalyl diphosphate synthase (BAC47414; C_{20} TC); PlaT2, phenalinolactone cyclase (EGJ74238; C_{20} TC); AtoE, atolypene cyclase (SED33305; C_{25} TC); Lon15, longestin cyclase (BAF98632; C_{40} TC); OSC, oxidosqualene cyclase (P48449; C_{30} TC); STC, squalene–tetrahymanol cyclase (ORY20771; C_{30} TC); SqhC, tetraprenyl–β-curcumene cyclase (Q796C3; C_{35} TC); and SHC, squalene–hopene cyclase (P33247; C_{30} TC).

Class II TCs have complex modular architectures that contain up to three domains—α, β, and γ. All known class II TCs possess the β domain that comprises the β-fold and the catalytic site. The majority of microbial TCs ($C_{>20}$ TCs) additionally have the γ domain that is structurally homologous to the β domain, and thus likely evolved by a gene duplication from an ancestral β domain (Christianson 2017). However, some TCs have unconventional modular architectures. Merosterol biosynthesis enzyme MstE in cyanobacteria consists only of the β domain (Moosmann et al. 2020) (Fig. 13). Also, some C_{15} TCs have a divergent β domain that is connected to a haloacid dehalogenase-like domain (Vo et al. 2022), while other C_{15} TCs have the canonical $\beta\gamma$ domains (Pan et al. 2022). In contrast, some eukaryotic TCs combine the $\beta\gamma$ domains with the α domain that comprises the α fold as seen in class I TC (Cao et al. 2010, Jia et al. 2022). The $\alpha\beta\gamma$ architecture can work as the bifunctional class I + II TC. Other eukaryotic TCs lost one or two of the three domains and secondarily transformed into class I or class II TCs (α or $\alpha\beta$) (Cao et al. 2010, Oldfield and Lin 2012). The $\alpha\beta\gamma$ architecture is not known in Bacteria.

C_{30} TC (SHC and OSC) is the most widespread and likely have the oldest origin among all class II TCs. In contrast, the distribution of other TCs is confined mostly to several specific taxa, similar to class I TCs. SHC is inferred to be the oldest among C_{30} TCs and its evolution seems to be largely in parallel with SQS (section SQS family) (Santana-Molina et al. 2020). In contrast, OSC evolved from an ancestral SHC in aerobic environments (Hoshino and Gaucher 2021). The OSC substrate oxidosqualene is produced through the epoxidation of squalene by FAD-dependent monooxygenase (FMO). This step requires molecular oxygen and FMO and OSC are in fact distributed only in aerobic species (Summons et al. 2006). Interestingly, the combination of epoxidation and class II cyclization is observed for not only C_{30} terpenoids, but also C_{20}, C_{25}, and C_{40} terpenoids (Smanski et al. 2012, Dairi 2013, Kim et al. 2019) (Fig. 13). FMO and class II TC proteins for C_{20}, C_{25}, C_{30}, and C_{40} terpenoids are homologous between different terpenoid

Table 1. Terpenoid biosynthesis enzyme families that are involved in primary metabolism. Enzymes that are not discussed in this review are not included (e.g. prenyl reductase, desaturase). In most cases, short-chain indicates C_{15} or C_{20}.

Cellular process	Domain	*Trans*-PT	*Cis*-PT	Others	Compounds involved
Respiration	A and B	Short-chain *trans*-IPPS long-chain *trans*-IPPS UbiA (MenA/MqnP)			Quinones, methanophenazine, and hemes
Membrane formation	A	Short-chain *trans*-IPPS UbiA (DGGGPS)		GGGPS	Archaeol and GDGTs
Membrane regulation	B	Short-chain *trans*-IPPS SQS		Class II TC LC	Hopanoids, steroids, and carotenoids
Phototrophy	A and B	Short-chain *trans*-IPPS SQS UbiA (BCS)		LC	(Bacterio)chlorophylls, retinal and carotenoids
Cell envelope formation	A and B	Short-chain *trans*-IPPS	*cis*-IPPS	PGTp/PGTm GT2	Glycosyl lipid carriers

Abbreviations: PT, prenyltransferase; A, Archaea; B, Bacteria; IPPS, isoprenyl diphosphate synthase, UbiA, 4-hydroxybenzoate PT family; MenA, 1,4-dihydroxy-2-naphthoate PT; MqnP, putative 1,4-dihydroxy-6-naphthoate PT; DGGGPS, digeranylgeranylglyceryl phosphate synthase; GGGPS, geranylgeranylglyceryl phosphate synthase; GDGTs, glycerol dibiphytanyl glycerol tetraethers; SQS, squalene synthase family; Hpn, hopanoid biosynthesis enzyme; Sqs, squalene synthase; TC, terpene cyclase; LC, lycopene cyclase; Crt, carotenoid biosynthesis enzyme; BCS, (bacterio)chlorophyll synthase; PGTp, polytopic phosphoglycosyl transferase; PGTm, monotopic PGT; and GT2, family 2 glycosyltransferase.

classes for each enzyme. Thus, the tandem epoxidation and cyclization pathway seems to share a common ancestry. Class II TCs that have an associated FMO are found in two different clades (red lines; Fig. 15) and this suggests a horizontal transfer of the FMO and class II TC gene pair.

IDT family

The indole diterpene synthase (IDT) family is involved in indole terpenoid biosynthesis in some fungi and bacteria (Rudolf and Chang 2020). IDT enzymes catalyze the class II-type cyclization. Interestingly, some (but not all) IDT enzymes are involved in the tandem epoxidation and cyclization reactions, which are analogous to FMO and class II TC. Further, IDT proteins are integral membrane proteins, similar to C_{30} class II TCs. Yet, IDT proteins do not have a sequence similarity to class II TC. Thus, the apparent mechanistic similarity between IDT and class II TC proteins may be due to convergent evolution, or alternatively IDT proteins may be too diverged to detect the homology. For instance, XiaE from *Streptomyces sp. HKI0576* is involved in the biosynthesis of xiamycin A (indole sesquiterpenoid), in tandem with XiaL that acts as FMO (Xu et al. 2012) (Fig. 13). Contrarily, DmtA1 from *Streptomyces youssouensis OUC6819* is involved in drimentine biosynthesis without substrate epoxidation.

Other secondary metabolite-related cyclases

There are many more examples of TCs that have distinct cyclization mechanisms and evolutionary origins. For instance, vanadium haloperoxidases perform a cyclization and halogenation of terpenoid substrates simultaneously by a class II TC-like mechanism (Winter et al. 2007). Also, Stig cyclases perform the Cope rearrangement that is rare in nature, followed by the cyclization of terpenoid substrates (Li et al. 2015). These novel enzymes are generally distributed in a small number of bacterial taxa (commonly in Actinobacteria and Cyanobacteria), which is a general feature of cyclases for secondary metabolites. Although these enzymes are not discussed in this review, they demonstrate the biochemical versatility of terpenoid biosynthesis. Interested readers are referred to recent reviews (Baunach et al. 2014, Rudolf and Chang 2020). The presence of yet-unknown cyclases is also suggested. C_{35} heptaprenylcycline is a monocyclic terpenoid that is biosynthesized by cyclization of saturated *cis*-prenyl chains in mycobacterium species (Sato 2013). No conventional class I and II TCs are found in these species and hence the presence of a novel cyclase is predicted.

Terpenoids at the origin of primary metabolism

Many fundamental cellular processes are associated with terpenoid biosynthesis because main metabolites include terpenoids and/or terpenoid-attached compounds. Hence, addressing the evolution of terpenoid biosynthesis provides important information about the evolution of those terpenoid-associated cellular processes. Interestingly, primary metabolites that contain terpenoids are all membrane-associated, and hence a strong connection between terpenoids and cellular membranes is envisaged (Table 1). The majority of terpenoids that engage in primary metabolism are produced by *trans*-PT enzymes, which reflects the functional versatility of this superfamily. This contrasts with *cis*-PT enzymes that are confined to cell envelope formation. It is unknown if this distinction is caused by either biochemical properties or physiological constraints of enzymatic products. Following sections discuss the evolution of important primary metabolism from the viewpoint of terpenoid evolution.

Respiration

Quinones and methanophenazine

Terpenoids take part in respiration in the form of prenylated quinones and methanophenazine (Table 1 and Fig. 16). Prenylated quinones primarily function as electron carriers for both aerobic and anaerobic respirations, but in fact have multiple physiological roles (Franza and Gaudu 2022). Menaquinones are the most widely distributed quinones in both Bacteria and Archaea (Nowicka and Kruk 2010, Schoepp-Cothenet et al. 2013). Menaquinones are biosynthesized via two distinct pathways: the canonical *o*-succinylbenzoate pathway and the more recently described futalosine pathway (Dairi 2009) (Fig. 9). Both pathways are distributed in Bacteria and Archaea, but the canonical pathway is associated more with aerobic species, while the futalosine pathway is found more in anaerobic species (Zhi et al. 2014). Individual species generally have only one pathway. Menaquinone

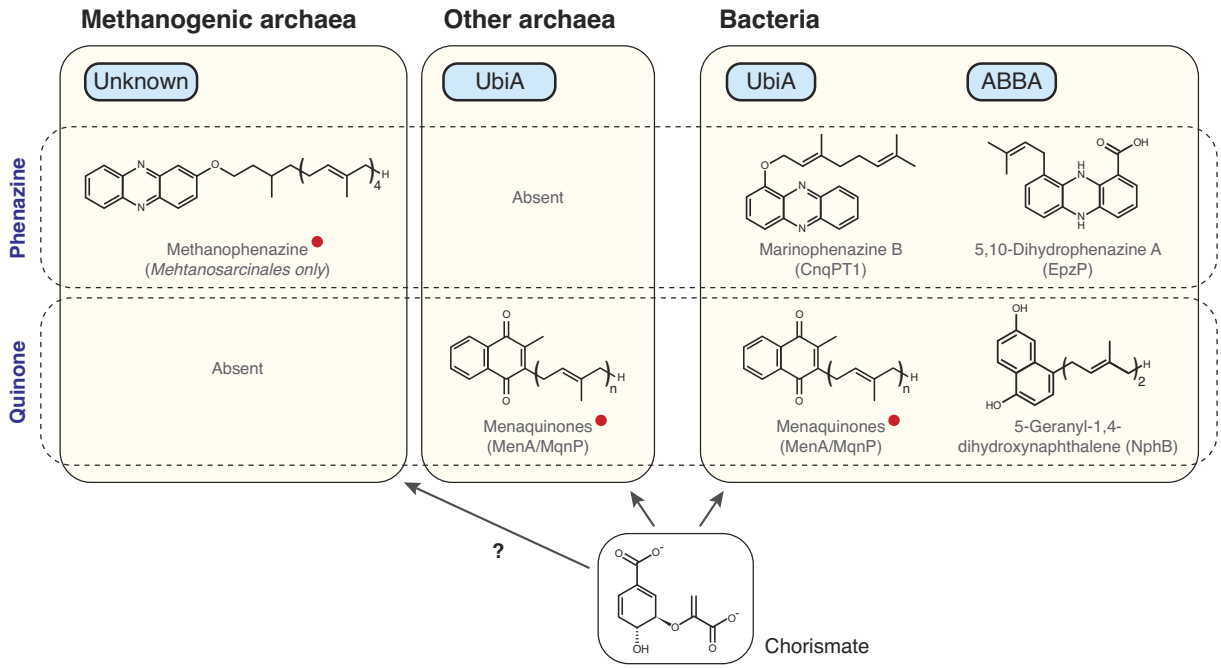

Figure 16. Prenylated quinones and phenazines in Bacteria and Archaea. These aromatic compounds are generally produced from chorismate, while the biosynthetic pathway for methanophenazine remains unclear. Red circles indicate that the compounds are utilized as electron carriers for respiration. Abbreviation and accession number: EpzP, 5,10-dihydrophenazine-1-carboxylate-9-dimethylallyltransferase (E5KWG9). See Figs 4 and 8 for other proteins.

precursors are prenylated by UbiA family enzymes in both pathways (MenA and possibly MqnP; Fig. 8) (Cotrim et al. 2017, Joshi et al. 2018). Previous phylogenetic analyses suggested that the canonical pathway was evolved in Bacteria and was later horizontally transferred to some archaea, while the futalosine pathway is more ancestral (Zhi et al. 2014). The distribution of MenA and MqnP is resurveyed in this review and the distribution of MenA/MqnP seems to be more complex than previously thought. For instance, MqnP is indeed widespread in some anaerobic lineages such as Clostridia and Melainabacteria in Bacteria, but is nearly absent in other anaerobic lineages such as Asgard and DPANN superphyla in Archaea. In contrast, MenA is found in those archaeal lineages, although its distribution is not universal. Overall, both MenA and MqnP are found in major bacterial and archaeal taxa, even though individual species generally possess only one of the two. Updated phylogenetic trees for MenA and MqnP are shown in Fig. 17. Bacteria and Archaea largely form separate clusters for both MenA and MqnP. Also, individual subclades within each domain roughly follow the species tree for both Bacteria and Archaea, despite the observation that the trees are not rooted between Bacteria and Archaea and there also seems to have been a substantial amount of HGT. Overall, the possibility that both enzymes were present in LUCA is not excluded.

It remains challenging to decisively infer the presence (or absence) of quinones and a quinone-based respiration system in LUCA. Although MenA and MqnP are widespread in both Bacteria and Archaea, several early-branching archaeal lineages lack quinones and enzymes that are necessary for quinone prenylation—long-chain *trans*-IPPS, MenA, and MqnP (Fig. 17). For instance, the majority of DPANN archaea that are potentially the deepest-branching species in the archaeal domain lack respiration-related genes (Dombrowski et al. 2019). This may be explained by their symbiotic lifestyle and genome reduction. However, several early-branching anaerobic lineages in Euryarchaeota also entirely lack quinones (Elling et al. 2016). They are mainly methanogens (*Methanopyri*, *Methanococci*, and *Methanobacteria*), but also include other species (*Thermococci*). These organisms rely on quinone-independent primordial respiratory systems (in particular groups 3 and 4 [NiFe] hydrogenases) (Peters et al. 2015). Interestingly, in aerobic organisms, group 4 [NiFe] hydrogenases adapted to accommodate quinones and transformed into complex I in the aerobic electron transport chain (Schut et al. 2013, 2016, Yu et al. 2021).

The near ubiquity of quinones in Bacteria and Archaea suggests that quinones were present and involved in anaerobic respiration well before the evolution of complex I in aerobic respiration. However, from the bioenergetics perspective, quinones are not preferred electron carriers on the early Earth, due to the absence of suitable terminal electron acceptors with a high electrochemical potential that can oxidize quinones. As discussed above, the phylogenetic tree topology for long-chain *trans*-IPPS, MenA, and MqnP is consistent with the scenario that quinone biosynthesis was largely vertically inherited in Bacteria and Archaea. However, the absence of corresponding genes in several early-branching archaeal lineages and also the possible presence of quinone-independent respiration systems in early life render the presence of quinones in LUCA inconclusive. Quinones perhaps had a function that was irrelevant to respiration in early life, but this remains speculative.

Meanwhile, some methanogens (Methanosarcinales) utilize a unique prenylated phenazine compound called methanophenazine as the electron carrier for their respiration, instead of conventional quinones (Thauer et al. 2008) (Fig. 16). Methanophenazine is the only known non-bacterial phenazine, while phenazine biosynthesis is nearly exclusive to Bacteria, in particular soil species (Mavrodi et al. 2010, Thomashow 2013, Dar et al. 2020). While bacterial phenazines are involved in secondary metabolism, methanophenazine is

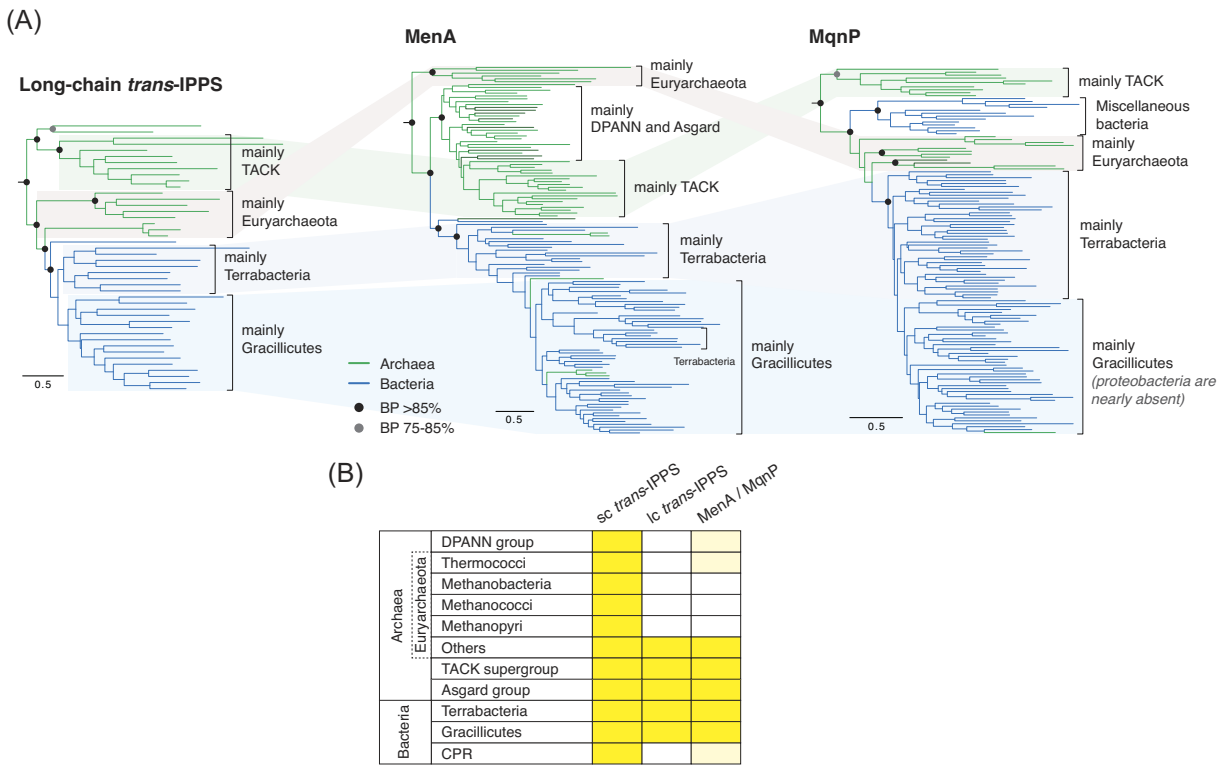

Figure 17. (A) Maximum-likelihood trees for three terpenoid biosynthesis enzymes that are associated with respirations—long-chain *trans*-IPPS, MenA, and MqnP. Outgroup is omitted for convenience. The dataset for long-chain *trans*-IPPS is identical to Fig. 11. Protein sequences were retrieved from the NCBI database by Blastp search, using representative proteins (see below) as the query sequences with the cutoff threshold of e-value 10^{-5} and >60% query identity coverage. Taxonomically redundant sequences were excluded. The sequences were aligned using MUSCLE v. 3.8.31. Phylogenetic analyses were performed using IQ-TREE v. 2.0.6. Bootstrap values (BP) were based on 1000 Ultrafast bootstraps. The source data are in the supplementary information. **(B)** Distribution of short-chain and long-chain *trans*-IPPSs and MenA/MqnP in Bacteria and Archaea. Light yellow boxes indicate its presence only in a small fraction of the taxon. Abbreviations: lc, long-chain; sc, short-chain; TACK, Thaumarchaeota, Aigarchaeota, Crenarchaeota, and Korarchaeota superphylum; and CPR, candidate phyla radiation. Abbreviations for representative proteins in the trees: long-chain *trans*-IPPS, P0AD57 (bacteria) and AAB89695 (archaea); MenA, AAC76912 (bacteria) and ABK77809 (archaea); MqnP, ABK52030 (bacteria) and O28106 (archaea).

a functional equivalent of respiratory quinones, and hence is an indispensable primary metabolite (Beifuss and Tietze 2005). Interestingly, methanophenazine has a C_{25} prenyl chain, unlike quinones that mostly have $C_{>30}$ chains. The prenylation mechanism of methanophenazine is not elucidated yet. ABBA enzymes that prenylate phenazines in Bacteria are not found in Archaea. In contrast, UbiA enzymes, which catalyze the prenylation of menaquinone precursors, are widely distributed in Archaea. In fact, divergent MqnP homologs are present in Methanosarcinales, although their function remains unknown (Supplementary Text 2; Table S3, Supporting Information). Hence, there is a possibility that methanophenazine prenylation proceeds not by ABBA enzymes, but by UbiA enzymes, similar to CnqPT1 that is a rare example of phenazine-prenylating UbiA enzyme (Zeyhle et al. 2014) (Fig. 16). Also, the IPPS enzyme that produces C_{25} prenyl chains for methanophenazine is inferred to be derived from archaeal long-chain *trans*-IPPS that produces $C_{>30}$ prenyl chains for quinones (Ogawa et al. 2010), instead of conventional short-chain *trans*-IPPS. These observations suggest that the prenyl chain formation and the prenylation of methanophenazine originates in menaquinone biosynthesis in Archaea. This implies that methanophenazine evolved later than quinones, contrary to a previous inference that methanophenazine-based respiration represents an intermediate state for the respiratory evolution of group 4 [NiFe] dehydrogenase from quinone-independent ancestral forms toward quinone-dependent modern forms (complex I; Schut et al. 2016). The biosynthetic pathway of methanophenazine is enigmatic and it is unknown if methanogens perform the bacterial-type phenazine biosynthesis or not. Understanding the archaeal phenazine biosynthesis pathway and its evolutionary relationship to the bacterial pathway would be beneficial to constrain the evolutionary origin of methanophenazine and cytochrome-coupled methanogenesis (Thauer et al. 2008).

Hemes

Terpenoids take part in respiration also in the form of prenylated hemes (Table 1 and Fig. 8). Hemes function as prosthetic groups for membrane-bound heme–copper oxidases that reduce molecular oxygen to water, being the last enzymatic complexes in aerobic respiration (known as complex IV). Heme–copper oxidases are evolutionarily related to nitric oxide reductases that reduce NO to water and N_2O (Sousa et al. 2012, Glass et al. 2023). Both enzyme families utilize hemes, but prenylated hemes are found only in heme–copper oxidases. Two types of hemes—heme A and heme O—have prenyl side chains and are biosynthesized from non-prenylated heme B (Hederstedt 2012). All three types are distributed in Bacteria, while Archaea possesses variants of hemes A and O (Lübben and Morand 1994). Only heme A is found in

Eukarya. Heme–copper oxidases are generally associated with (micro)aerobic organisms. Consistently, homologs of the heme prenylation enzyme (UbiA family; Fig. 8) is present mostly in aerobic species. However, it is currently not known if heme prenylation emerged in parallel with the evolution of heme–copper oxidases.

Membrane organization

Terpenoids are involved in membrane organization in both Bacteria and Archaea, but in two different ways. The archaeal membrane is composed of terpenoids, and thus the evolution of the entire archaeal domain is built on the evolution of two nonhomologous terpenoid biosynthesis enzymes: GGGPS and DGGGPS. These two enzymes catalyze the first and second successive prenyl group transfers to G1P during the archaeal membrane lipid biosynthesis. GGGPS is a cytoplasmic non-UbiA enzyme (Peterhoff et al. 2014), while DGGGPS is a membrane-bound UbiA enzyme (Koga and Morii 2007). The evolutionary history of GGGPS and DGGGPS remains unresolved, but these two enzymes are distributed not only in Archaea, but also in Bacteria (Peterhoff et al. 2014). Phylogenetic analyses suggest that archaeal and bacterial homologs separately cluster for both enzymes (Villanueva et al. 2021). However, while GGGPS and DGGGPS are universally distributed in archaea, they are also mostly confined to the FCB superphylum in Bacteria. Hence, the presence of these two enzymes in LUCA is not conclusive thus far (Coleman et al. 2019, Villanueva et al. 2021). In any case, GGGPS and DGGGPS has a comparably ancient origin to MenA and MqnP, even if they were not present in LUCA.

Contrary to membrane lipids in Archaea, membrane lipids in Bacteria are mostly fatty acid-based and thus terpenoids are not major structural components of the bacterial membrane. However, some bacteria are known to have a genetic repertoire to biosynthesize archaeal-type lipids that consist of G1P and ether-linked prenyl chains, although the presence of those lipids in membranes is still unclear. In Bacillales, GGGPS homolog PcrB catalyzes the transfer of C_{35} heptaprenyl diphosphate to G1P, yielding heptaprenylglyceryl phosphate (Guldan et al. 2011) (Fig. 10). In this pathway, the second prenylation does not occur, unlike the conventional archaeal lipid biosynthesis, but instead the remaining hydroxyl groups are acetylated. Also, bacteria that possess GGGPS and DGGGPS homologs (mainly FCB supergroup) potentially have the ability to biosynthesize archaeal-type lipids. *In vitro* characterizations of those homologs suggest that they are functional (Villanueva et al. 2021). Yet, the *in vivo* observation of archaeal-type lipid membranes is still absent in Bacteria thus far.

Apart from the terpenoids as structural components of membranes, terpenoids that are derived from squalene and phytoene (HH condensation products) play an important role in the regulation of fatty acyl membranes in Bacteria and Eukarya (HH-labelled compounds; Fig. 13). In modern biology, well-known membrane regulators include steroids, hopanoids and carotenoids (Cheng and Smith 2019). These compounds are all derived from the cyclization of squalene and phytoene by two distinct cyclase families: class II TC and LC (Table 1). Steroids play critical roles in membrane dynamics in nearly all eukaryotes (Sezgin et al. 2017), except for a few enigmatic steroid-deficient species (Takishita et al. 2017). The interactions between steroids and other membrane components, particularly sphingolipids, facilitate the mesoscopic-scale phase separation of membranes and enables host eukaryotic organisms to respond to external stimuli in a highly dynamic way (van Meer et al. 2008, Sezgin et al. 2017, Shaw et al. 2020). Steroids are further utilized as signaling agents (steroid hormones) in plants, fungi, and animals (Weete 1989, Schroepfer 2000, De Bruyne et al. 2014). Hopanoids are generally thought to be functional analogs of steroids and are observed to have similar phase separation effects (Sáenz et al. 2015). Carotenoids are also suggested to induce a heterogeneous membrane structure called functional membrane microdomains (FMMs) in Bacteria (Bramkamp and Lopez 2015, Lopez and Koch 2017). It is generally thought that steroids are eukaryotic-specific, while hopanoids are bacterial signatures (Summons et al. 2006). However, experimental and genetic data suggest that steroids are present also in some bacteria (Wei et al. 2016, Hoshino and Gaucher 2021), although their physiological function is not well-understood (Welander 2019). It is further suggested that steroid biosynthesis evolved in Bacteria and horizontally transferred to proto-eukaryotes during eukaryogenesis (Hoshino and Gaucher 2021).

In summary, the evolution of the HH condensation and subsequent cyclization was critical for the establishment of the modern fatty-acyl membrane regulation, culminating in the evolution of the eukaryotic-type dynamic membrane system (see also section Terpenomes in the pre-aerobic world). The SQS family that catalyzes the HH condensation of prenyl chains may have evolved largely concomitantly with class II TC in Bacteria. The origin of SQS and class II is at the latest in the common ancestor of Gracillicutes (Santana-Molina et al. 2020). In turn, the evolution of LC is no earlier than that of SQS and class II TC. Contrary to Bacteria and Eukarya, membrane regulation in Archaea remains enigmatic since HH condensation products are nearly absent in Archaea, except for haloarchaea and a few others that can utilize squalene and carotenoids (Salvador-Castell et al. 2019). Other terpenoid-derived compounds such as polyprenols and quinones are hypothesized to be potential membrane regulators, but the archaeal regulation mechanism remains largely unknown (Salvador-Castell et al. 2019).

Phototrophy

Currently known phototrophic systems utilize two types of chromophores—(bacterio)chlorophylls and retinal (Bryant and Frigaard 2006, Hohmann-Marriott and Blankenship 2011) (Figs 8 and 13). Both chlorophyll-based and retinal-based phototrophies require terpenoids. Chlorophyll-based phototrophy requires C_{15}–C_{20} prenyl chains and C_{40} carotenoids, while the retinal-based system additionally requires a cleavage compound of carotenoids (apocarotenoid) (Table 1).

In the chlorophyll-based system, prenyl chains are used to anchor (bacterio)chlorophylls in membranes. The presence of terpenoids may not be mandatory only to receive photons, which is performed by the porphyrin moiety of chlorophylls. However, photosynthesis is a membrane-associated process, and thus the membrane anchoring of (bacterio)chlorophylls by the prenyl moiety is critical for the appropriate arrangement of photosynthesis components. Carotenoids are involved in receiving additional light in parallel with (bacterio)chlorophylls and also have an important photoprotection role (Sandmann 2021). Similar to prenyl chains, carotenoids may not be mandatory for the functionality of photosynthesis. However, no phototrophic system without carotenoids is known in nature thus far (Hashimoto et al. 2016). Therefore, terpenoids are inferred to have been present at the origin of the extant forms of phototrophy, although this does not necessarily preclude the presence of extinct primitive photosystems without terpenoids. (Bacterio)chlorophyll-based phototrophy is sporadically distributed in a small number of discrete bacterial taxa, besides

photosynthetic eukaryotes, and thus its origin is most likely in Bacteria (Ward et al. 2019). (Bacterio)chlorophyll biosynthesis requires a UbiA enzyme that catalyzes the prenylation of chlorophyll precursors (BCS; Fig. 8). Carotenoid biosynthesis additionally requires a SQS enzyme to form phytoene and a LC enzyme to cyclize carotenoids. The SQS and LC families evolved within Bacteria, as discussed in the previous section. This constrains the evolutionary timing of the modern-type (bacterio)chlorophyll-based phototrophy.

In the retinal-based phototrophy, terpenoids play central roles because the chromophore is a terpenoid itself. Retinal-based phototrophy is much more widely distributed than the (bacterio)chlorophyll-based phototrophy (Béjà and Lanyi 2014, Brown 2014, Pinhassi et al. 2016). The phylogeny of light-receptor protein rhodopsin, where the retinal chromophore is embedded, suggests a substantial number of HGT between taxa, including archaea and viruses (Pinhassi et al. 2016). The simple architecture of the retinal-based phototrophy may have rendered associated genes highly mobile between taxa. The apparent wide distribution of rhodopsin in the three domains of life have provoked a hypothesis that phototrophy (light-driven ion pump) evolved early even in LUCA (Shalaeva et al. 2015, DasSarma and Schwieterman 2021). However, the distribution of rhodopsin in Archaea is sporadic (mostly haloarchaea and Marine Group II), and thus a horizontal origin of archaeal rhodopsins from Bacteria is more plausible. The distribution of rhodopsin is limited mostly to aerobic species even in Bacteria and the phylogenetic tree topology of rhodopsin is not consistent with the species tree of Bacteria and Archaea (Pinhassi et al. 2016). Hence, the origin of retinal-based phototrophy is more likely within Bacteria.

Retinal is an apocarotenoid that is biosynthesized through the enzymatic cleavage of C_{40} carotenoids (Fig. 13). This is catalyzed by carotenoid oxygenase, utilizing molecular oxygen (Blh family; Kim et al. 2009). Nonenzymatic cleavage of carotenoids may occur by reactive oxygen species (ROS) as observed in plant apocarotenoid productions (Felemban et al. 2019), but the microbial production of retinal by ROS is not well-understood thus far. Chromophores other than retinal are also not known for rhodopsin. Overall, it is possible that retinal and its phototrophy evolved only after the first oxygenation of Earth by the Great Oxidation Event (GOE) (Lyons et al. 2014). In some species, retinal may be produced by a non-Blh pathway (Nakajima et al. 2020, Chazan et al. 2022). For instance, Diox1 enzymes can produce retinal by the cleavage of other apocarotenoids (Ahrazem et al. 2016). In any case, currently known carotenoid cleavages, either enzymatically or nonenzymatically, are all oxygen-dependent. Even if molecular oxygen is not the defining factor for the origin of retinal, carotenoids as a whole are inferred to have evolved in Bacteria after the emergence of the HH condensation that is a prerequisite for the carotenoid precursor formation. Therefore, current data suggest that both (bacterio) chlorophyll-based and retinal-based phototrophies evolved in Bacteria only after the divergence of LUCA.

Cell surface biogenesis

Two major types of prenyl chains—*trans*- and *cis*-prenyl chains—are produced by *trans*- and *cis*-PT superfamilies, respectively. These two chains are dedicated to different cellular processes and seem to not interfere with each other (Table 1). *Cis*-prenyl chains are utilized as glycosyl lipid carriers in both Bacteria and Archaea, but for different physiological purposes between the two domains. While Archaea uses the glycosyl moiety for protein N-glycosylation (Jarrell et al. 2014), Bacteria utilizes the glycosyl moiety for the biosynthesis of cell envelope polymers, including peptidoglycans, O-antigens, and teichoic acids (Silhavy et al. 2010, Swoboda et al. 2010, Egan et al. 2020, Whitfield et al. 2020) (Fig. 12). Archaea mostly lacks polysaccharide-based cell walls, except for pseudomureins, but instead utilize S-layers that are commonly N-glycosylated as alternative cell walls (Albers and Meyer 2011).

Interestingly, the first two glycosyl transfer steps of N-glycosylation and peptidoglycan biosynthesis are homologous between Archaea and Bacteria. In these early steps, *cis*-prenyl chains (*cis*-isoprenyl monophosphates) successively conjugate with two glycosyl monomers that are derived from nucleotide sugars. The first step is the transfer of a phosphorylated glycosyl group to a *cis*-prenyl chain and is catalyzed by PGT enzymes, while the second step is the direct transfer of a glycosyl group to the first glycosyl group and is catalyzed by GT enzymes (Allen and Imperiali 2019) (section GT and PGT families). Bacterial PGT and GT enzymes (MraY and MurG) and archaeal enzymes (AglH and Agl24) are homologous to each other for individual families and phylogenetic analyses also supports the separation of the bacterial and archaeal clades for both families (Lombard 2016b, Meyer et al. 2022). Therefore, a *cis*-prenyl chain-based glycosyl conjugation system is inferred to have been present in LUCA. In Bacteria, one molecule of *N*-acetylglucosamine and one molecule of *N*-acetylmuramic acid form a disaccharide core in the first two steps (Fig. 12). The core conjugates with one another repeatedly and eventually form polysaccharide cell walls (peptidoglycan). In Archaea, some species similarly utilize *N*-acetylglucosamine as the first glycosyl monomer, but the composition of monomers for N-glycosylation are not universally conserved within the domain (Jarrell et al. 2014). Hence, the composition and the nature of ancestral glycoconjugates in LUCA is currently unclear, even though those conjugates were possibly involved in the formation of cell surface structures.

Synthesis—4 billion years of microbial terpenome evolution

The following sections aim to provide a comprehensive and coherent narrative of the evolution of terpenoids throughout the history of life, synthesizing all the evolutionary implications obtained in the previous sections. First, the current arguments about the enigmatic origin of protocells and lipid membranes are briefly summarized. Then, the potential relationship of terpenoids to the evolution early life and membranes are discussed. However, analyses of extant organisms do not provide direct information about pre-LUCA organisms and their biochemistry. Hence, any arguments about those earliest organisms remain indirect and thus speculative. Even for LUCA, many arguments remain inconclusive. Consequently, this section does not aim to provide a convincing picture about the nature of terpenoids in those ancient organisms, but provide several possible scenarios that are not excluded based on the current analyses. Lastly, the terpenome expansion after the divergence of Bacteria and Archaea, which is more clearly deduced from phylogenomic comparisons between the two domains, is summarized. It is noted that comprehensive analyses about the relationship of terpenoid biosynthesis to other carbon metabolic pathways and to the geochemical settings are beyond the scope of this review. These factors may have set the stages for the incessant diversification of terpenoid biosynthesis in a broader context. Integrating these additional perspectives will further enhance our understanding of early life evolution and are left for future studies.

Hypothetical abiotic terpenoid synthesis and protocell formation

There is currently no experimental confirmation of abiotic terpenoid production (Kirschning 2021). Similarly, there is no observation of terpenoids in abiotic sources in nature (Sephton 2002). Yet, some analogous reactions for terpenoid biosynthesis are known to proceed abiotically. For instance, C_5 isopentenol can conjugate into C_{10}–C_{15} isoprenols in the presence of the mineral clay montmorillonite K10 (Désaubry et al. 2003). Similarly, spontaneous vesicle formation is observed for C_{15} and C_{20} isoprenyl monophosphates (Streiff et al. 2007). C_5 prenyl monomer analogs (isopentenol and dimethylallyl alcohol) can also be produced by Prins-type reactions from formaldehyde and isobutene (Nakatani et al. 2014). However, the relevance of these reactions to the prebiotic Earth's environment is not clear.

Contrary to terpenoids, fatty acids are generally thought to be promising prebiotic precursors for the membranes of protocells (Deamer et al. 2002). These compounds are believed to have been provided through abiotic processes, including Miller–Urey-type electric discharges of atmosphere and Fischer–Tropsch-type reactions of CO_2/CO on meteorites and hydrothermal vents (Chen and Walde 2010). However, experimental and observational evidence for the abiotic formation of *long-chain* ($C_{>12}$) fatty acids is in fact scarce as is the case for abiotic terpenoid synthesis (Kirschning 2021). It is not fully validated if Fischer–Tropsch-type reactions can produce long-chain fatty acids under geologically plausible conditions (McCollom and Seewald 2007, McCollom 2013). Also, the presence of fatty acids on meteorites is either dismissed as terrestrial contamination (Sephton 2002), or observed only for short- to mid-chain lengths ($C_{<12}$) (Lai et al. 2019).

Therefore, there is no compelling evidence for the presence of any of the *modern-type* membrane building blocks in the prebiotic Earth thus far. The absence of clear evidence for lipid membrane-surrounded cells in and before LUCA even led to a hypothesis that LUCA did not have lipid membranes, but instead microscale compartments in hydrothermal systems functioned as primitive cells (Martin and Russell 2003, Martin et al. 2008, Lane et al. 2010). However, the universal presence of several membrane proteins, phosphate head groups, and membrane-associated cellular processes (e.g. ATP synthesis) in both Bacteria and Archaea suggests the presence of some forms of membranes in LUCA (Caforio and Driessen 2017). Experimental evidence suggests that mixing of different lipids, such as fatty acids and terpenoids, can stabilize the vesicle formation compared to membranes with a single lipid composition (Jain et al. 2014), even without phosphate head groups (Jordan et al. 2019). Hence, the emergence and early evolution of cellular membranes may have been a more complex process than previously thought, involving multiple types of lipid molecules.

Terpenomes in and before LUCA

The presence of terpenoids in pre-LUCA periods is supported by the universal presence of short-chain *trans*-IPPS in both Bacteria and Archaea and the phylogenetic relationship of bacterial and archaeal homologs (Lombard and Moreira 2011). The catalytic site resides in the well-conserved α domain and thus the evolution of the α fold would have been the critical moment for the emergence of $C_{>5}$ terpenoids in nature (α_{sc}; Fig. 18). The α-fold is inferred to have evolved by a fusion of further ancestral 4-helix bundle proteins (Christianson 2017). The first products of short-chain *trans*-IPPS were most likely FPP and/or GGPP. While these molecules are not stable in themselves, they were likely to be intermediates for other ancestral metabolic pathways, as observed in the modern terpenoid biochemistry. Prenyl monomers DMAPP and IPP were also presumably present, regardless of the synthetic route. The majority of terpenoid biosynthesis is performed by *trans*-PT enzymes, all of which evolved ultimately from short-chain *trans*-IPPS. The substrates of *trans*-PT enzymes are in most cases products of other *trans*-PT enzymes in earlier steps (dashed arrows; Fig. 18). Therefore, the substantial part of terpenoid biosynthesis is a result of the internal diversification of the *trans*-PT superfamily. In contrast, later parts of terpenoid biosynthesis (prenylation and cyclization) utilize additional enzymes (e.g. ABBA, class I and II TCs, and LC; Fig. 18). Those enzymes contribute to the diversity and complexity of the modern terpenome.

In modern organisms, terpenoids in primary metabolism are predominantly associated with cellular membranes. In the case of Archaea, the entire membrane is composed of terpenoid-based phospholipids. However, the earliest forms of terpenoids—prenyl monomers and $C_{<20}$ oligomers—are not directly associated with membranes, and hence the function of those earliest terpenoids is not clear. For instance, it was previously proposed that the conjugation of terpenoids and nucleic acids (i.e. prenylated nucleic acids) might represent a relic of prebiotic lipid–nucleic acid interactions (Scott 1997). In modern biology, prenyl conjugation of nucleic acids (tRNA) is observed for DMAPP and GPP and is catalyzed by two evolutionarily independent enzymes MiaA and SelU) (Soderberg and Poulter 2000, Dumelin et al. 2012). However, the distribution of these enzymes is confined to Bacteria and nucleic acid prenylation is not known in Archaea thus far. Although this does not necessarily preclude the presence of an analogous prenylation pathway in LUCA, evidence is currently absent to support the presence of interactions between terpenoids and nucleic acids before the emergence of Aacteria.

In contrast, C_5 prenylation of the cofactor FMN by the flavoenzyme UbiX is nearly ubiquitously distributed in both Bacteria and Archaea (Bloor et al. 2022). This suggests the presence of UbiX in LUCA, although the ancestral function of prenylated FMN is unknown (Fig. 18). FMN is among the oldest small molecules that potentially share a common ancestor with RNA molecules in the prebiotic world, together with other cofactors that harbor a common structural component (adenosine monophosphate handle) (Kirschning 2021). Therefore, the possibility of ancient interactions between prenyl monomers and cofactors, regardless of the presence of membranes, is at least not excluded. In turn, the biosynthesis of FPP and GGPP (but not shorter chains) by short-chain *trans*-IPPS may reflect the adaptation of terpenoid biosynthesis toward cellular membranes because only C_{15} or longer prenyl chains seem to have the capacity to form vesicles (Streiff et al. 2007). In this case, the evolution of short-chain *trans*-IPPS was the first step for terpenoids to participate in membrane organization. In other words, terpenoids in the modern sense originate in the functional deviation of prenyl monomers (and their precursors) that were present in a primordial chemical network outside the context of membrane organization.

The most ancestral route of the prenyl monomer formation pathway in Archaea (MVA pathway, Route I) is dependent on prenylated FMN, and therefore, prenyl monomers themselves (AMPD step; Fig. 2). This implies the presence of prenyl monomers already before the establishment of the MVA pathway. However, in this review, the presence of either the MVA or the MEP pathway in LUCA is not particularly supported, con-

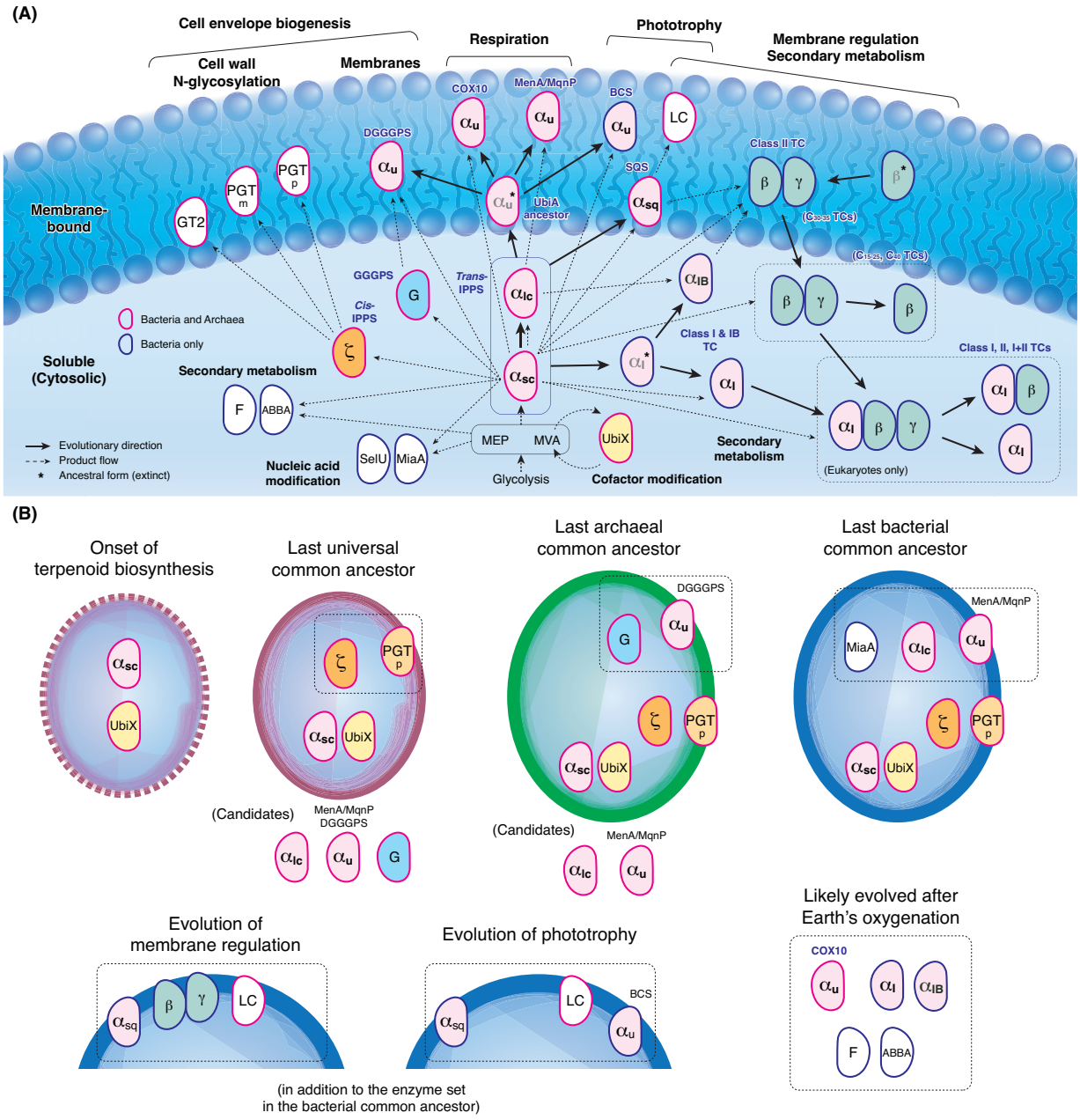

Figure 18. (A) Enzymatic expansion of microbial terpenoid biosynthesis. The evolutionary trajectory of terpenoid biosynthesis enzymes is shown by solid arrows, while the catalytic product flow is shown by dashed arrows. Proteins in the *trans*-PT superfamily (α-fold) are shown in pink to clarify their common ancestry. The direct ancestors of the UbiA family, the SQS family and the class I/IB TC family are not clear—likely either short-chain or long-chain *trans*-IPPS. The subcellular location of proteins is indicated by their position in the figure—membrane-bound or cytosolic. The subcellular location of proteins that possessed the extinct ancestral β domain ($\beta*$) is not clear. The presence of the MVA/MEP pathway in LUCA is not conclusive. Similarly, the evolutionary history of monotopic PGT and GT2 proteins is not elucidated yet. Also, it is beyond the scope of this review to infer the origin of glycolysis before LUCA. The greek letter nomenclature (α, β, γ, and ζ) is based on the conserved fold and is adapted from Lin and Oldfield (2012). The ε-fold for the SQS family is renamed as α_{SQ} to clarify the evolutionary link to other *trans*-PT proteins. Several small families are not shown (see Figs 10 and 13). Enzyme abbreviations: α_{SC}, short-chain *trans*-IPPS; α_{LC}, long-chain *trans*-IPPS; α_U, UbiA family; α_{SQ} (ε), SQS family; α_I, class I TC; α_{IB}, class IB TC; β and γ, two domains of class II TC; ζ, *cis*-IPPS; G, GGGPS; and PGTp/m, polytopic/monotopic PGT. **(B)** Enzyme repertoire in key ancient organisms. The membrane at the onset of terpenoid biosynthesis is represented by a dashed circle since its presence is not clear. At each evolutionary stage, newly evolved enzymes from the previous stage are highlighted by dashed squares. It is noted that the β domain may have been present in LUCA (section Class II TC family), but its function was probably not related to terpenoid biosynthesis and thus are not shown in the figure.

trary to previous arguments (section Monomer formation). The apparent no similarity between the two pathways hints at the possibility that the extant forms of prenyl monomer biosynthesis were in fact not evolved yet in LUCA. Hence, the absence of an established monomer biosynthesis pathway may also echo a primordial extinct form of prenyl monomer biochemistry that is implied by the presence of prenylated FMN. In this scenario, the chain elongation step by short-chain *trans*-IPPS evolved before a modern-type monomer biosynthesis pathway was consolidated.

Following short-chain *trans*-IPPS and UbiX, the evolution of additional terpenoid biosynthetic enzymes provides us with more clues to the evolutionary relationship between terpenoid biosynthesis and cellular membranes. *Cis*-IPPS that harbors the ζ-fold likely evolved before the divergence of LUCA (Fig. 18). The presence of *cis*-IPPS is indicative of cellular membranes, even though *cis*-IPPS itself is a soluble protein. *Cis*-IPPS produces *cis*-type prenyl chains, using the products of short-chain *trans*-IPPS. *Cis*-prenyl chains serve as membrane-bound scaffolds for the assembly of oligosaccharide chains, which are eventually utilized for the cell wall and other cell envelope structures in Bacteria and Archaea. For those processes, the presence of membranes is a prerequisite. The early steps of the lipid–oligosaccharide chain formation are homologous between Bacteria and Archaea. Hence, it is likely that LUCA had some forms of cell envelope structure that was built on cellular membranes. Long-chain *trans*-IPPS and UbiA family proteins that are utilized for respiration in modern organisms may additionally have been present in LUCA, or shortly after the divergence of LUCA. Both enzymes comprise the α domain and thus most likely emerged from short-chain *trans*-IPPS (Fig. 18). Long-chain *trans*-IPPS produces $C_{>30}$ prenyl chains that are dedicated to the formation of quinone side chains as membrane anchors. UbiA enzymes prenylate various substrates, but the oldest group in the family was likely involved in menaquinone prenylation, utilizing the products of long-chain *trans*-IPPS (Fig. 8). UbiA proteins themselves are membrane-bound, and thus the presence of membranes is prerequisite for their integrity and catalytic activity.

Hence, the presence of long-chain *trans*-IPPS, *cis*-IPPS, and UbiA enzymes all suggests the presence of membranes in host organisms, although this does not necessarily mean that terpenoids were structural components of membranes, as observed for modern archaea. In summary, the stepwise expansion of the early terpenoid biosynthesis machinery suggests that the most basal part of terpenoid biosynthesis—prenyl monomer formation—possibly predates the evolution of membrane-associated cellular processes that include respiration, cell envelope biosynthesis, N-glycosylation, and perhaps even cellular membrane itself, while the subsequent prenyl elongation (both *trans*- and *cis*-types) and in particular prenylation were largely built on the presence of cellular membranes.

Lipid composition of LUCA membranes

Even though the presence of cellular membranes is suggested in LUCA, the nature of ancestral membranes remains enigmatic. While Bacteria utilizes fatty acyl phospholipids, Archaea utilizes prenyl phospholipids. This dichotomy results in a long controversy about the lipid composition of the cellular membrane in LUCA. One hypothesis is that both fatty acids and terpenoids were present in LUCA (Koga 2011, Jain et al. 2014). The arguments mainly focus on the presence of prenyl lipid membranes, while the presence of fatty acyl membranes is taken almost for granted. However, there is surprisingly no strong evidence for the presence of $C_{>12}$ long-chain fatty acids in LUCA. Fatty acid biosynthesis is only sporadically distributed in Archaea, and hence a horizontal origin from Bacteria is suggested (Dibrova et al. 2014, Coleman et al. 2019). This implies even the total absence of long-chain fatty acids in LUCA. Similarly, transcriptional regulators for the fatty acid metabolism seem to be distinct between Bacteria and Archaea and their common ancestry is not supported (Wang et al. 2019). Additionally, not all fatty acid-processing proteins are present in Archaea (Lombard et al. 2012, Dibrova et al. 2014, Villanueva et al. 2017), including acyl carrier proteins (ACPs). For biochemical reactivity, both fatty acids and terpenoids need to be activated. Fatty acids are ACP-activated, while terpenoids are diphosphate-activated. ACP is a large conserved family of carrier proteins, typically consisting of 70–100 residues (Byers and Gong 2007) and is among key molecules in the bacterial metabolism. However, ACP is absent in the entire archaeal domain (Lombard et al. 2012), and thus the ACP-based fatty acid activation likely evolved only in Bacteria. Although a primitive fatty acid biosynthesis may have utilized an alternative activation mechanism such as coenzyme A (Genschel 2004), the observed discrepancy between Bacteria and Archaea implies that the pre-bacterial form of fatty acid biosynthesis, if existed, was different from the modern bacterial form.

The occurrence of fatty acid-processing proteins mostly only in Bacteria contrasts to the presence of basal terpenoid-processing proteins in both Bacteria and Archaea. Also, there is virtually no sign of the presence of long-chain fatty acids in the membrane-associated cellular processes in early life, unlike terpenoids that were likely involved in cell envelope biogenesis and perhaps also respiration. A straightforward interpretation of these observations is that fatty acid metabolism evolved later than terpenoid metabolism, even though fatty acids are generally believed to have a longer history in biology than terpenoids. One plausible reconciliation is that fatty acid metabolism may have gradually shifted from abiotic productions (Koonin and Martin 2005) to the modern protein-based biosynthesis. The apparent absence of an established fatty acid biosynthesis pathway in LUCA may suggest that the transition was still ongoing in LUCA and the pathway fully evolved after the divergence of Bacteria and Archaea, as is similarly speculated for prenyl monomer formation. Yet, the possibility that long-chain fatty acids were absent in LUCA is also not excluded based on the current observations. Overall, the presence of fatty acid *bio*-synthesis remains hypothetical before the emergence of Bacteria.

Terpenomes in the pre-aerobic world

The basal part of terpenoid biosynthesis is generally independent of molecular oxygen and the majority of basal terpenoid biosynthesis pathways appeared before the GOE. After the divergence of Bacteria and Archaea, the bacterial terpenome has constantly expanded until today, while the archaeal terpenome seems to largely retain its ancestral size, except for the specialization of its unique terpenoid-based membrane. Bacterial terpenome expansion generally consists of two major phases—the evolution of (1) terpenoid membrane regulators and (2) a diverse range of secondary metabolites. The first phase possibly occurred before the GOE, while the second phase occurred later. For both phases, the diversification of two functional domains—α and β domains—was critical (Fig. 18).

Important classes of membrane regulators—hopanoids, steroids, and carotenoids—are all built on the evolution of the HH condensation by SQS (α_{sq} or ε domain) and the following cyclization by class II TC ($\beta\gamma$ domains) and LC. The α_{sq} domain likely evolved from the ancestral α domain in *trans*-IPPS proteins. In turn, it is unclear if ancestral β domains initially had any terpenoid-associated functions, but a fusion of two β domains is inferred to have transformed into the canonical $\beta\gamma$ domains with a new catalytic site at the domain–domain interface (section Class II TC family) (Christianson 2017). This marked the onset of class II terpenoid cyclization. While modern class II TC acts on various $C_{>15}$ prenyl substrates, the oldest class II TC was likely specialized in squalene (Santana-Molina et al. 2020). Excluding steroids, the biosynthesis of hopanoids and carotenoids does not require molecular oxygen. Yet, there is currently no clear

evidence that hopanoids and carotenoids were present before the GOE. However, steroids are inferred to have already evolved shortly after the GOE (Gold et al. 2017), and thus the evolution of hopanoids and in particular carotenoids, which are inferred to have evolved before hopanoids and steroids, possibly predates the GOE. The evolutionary timing of the HH condensation (α_{sq}) and class II cyclization ($\beta\gamma$) seems to have been largely concomitant (Santana-Molina et al. 2020). HH condensation products are entirely hydrophobic and are embedded in membranes, unlike other terpenoid intermediates. Hence, the emergence of the HH condensation itself may have been associated with membrane regulation or membrane-associated cellular processes.

Hopanoids and carotenoids followed a different evolutionary history within individual classes. Hopanoids seem to largely retain their initial chemical diversity until today, since the hopanoid biosynthesis gene cluster is generally conserved among hopanoid-producing organisms (Belin et al. 2018). In contrast, carotenoids became associated with phototrophy, owing to the light-absorbing ability of conjugated double bonds that are produced by the desaturation activity of amino oxidase (section SQS family) (Hashimoto et al. 2016). Carotenoids were probably involved in the emergence of two extant forms of phototrophy – (bacterio)chlorophyll-based and retinal-based photosystems (section Phototrophy)—and were diversified through various lineage-specific modifications. The carbon skeletons of hopanoids, steroids, and carotenoids are highly recalcitrant to physical and chemical alterations over the geological time scale. Thus, those carbon skeletons can survive in the form of saturated or aromatized hydrocarbons for over hundreds of millions of years in sedimentary rocks. Today, these hydrocarbons are detected in geological samples as chemical fossils of ancient extinct organisms (e.g. Summons et al. 2021).

Terpenomes in the aerobic world

On the aerobic Earth, the bacterial terpenome experienced the second expansion phase through the evolution of terpenoid-associated secondary metabolites. This second expansion was driven by the further functional divergence of the α and β domains. The domain divergence can be seen in the evolution of class I and IB TCs (α domain) as well as class II TCs (β domain) that accommodate non-squalene substrates (Fig. 18). These TCs convert *trans*-prenyl substrates of various lengths into a wide variety of secondary metabolites. Interestingly, the association and dissociation of α, β, γ, and even other domains formed a complex, but unique molecular factory of terpenoids (e.g. $\alpha\beta$ and $\alpha\beta\gamma$ domains) (Faylo et al. 2021). The enzymatic reactions catalyzed by these domains do not require molecular oxygen. Thus, the terpenome expansion in the aerobic world seems to be linked more to the expansion of ecological niches and associated competitions among host organisms under aerobic conditions, than to the direct impact of oxygen on the chemical space of terpenoid biosynthesis.

In addition to the functional diversification of the pre-existing terpenoid biosynthesis machinery, various oxygen-utilizing enzymes were newly recruited for downstream terpenoid biosynthesis pathways. This also substantially contributed to the chemical diversity of terpenoids and had an impact on both primary and secondary metabolism. For instance, steroid biosynthesis evolved from hopanoid biosynthesis by the adaptation of monooxygenase toward squalene and the concomitant modification of class II TC to specialize in epoxidized substrates (Hoshino and Gaucher 2021). Steroid biosynthesis was a critical event for the establishment of the eukaryotic-specific dynamic membrane system and the evolution of the eukaryotic domain itself. The origin of retinal-based phototrophy is also possibly associated with the oxygenation of Earth because the currently known retinal biosynthesis requires dioxygenase, and hence molecular oxygen to produce retinal (Daruwalla et al. 2020), although the presence of a yet-unknown retinal biosynthesis pathway is suggested (Chazan et al. 2022). Many additional oxygen-utilizing enzymes introduced functional groups to the terpene carbon structures and enhanced the structural and physiological versatility of terpenoids (e.g. cytochrome P450) (Greule et al. 2018).

Also, while basal terpenoid biosynthesis does not require molecular oxygen, some enzymes are conversely susceptible to oxygen. Hence, there may have been an aerobic optimization of terpenoid biosynthesis in some aerobic bacteria and eukaryotes. Both the MEP and MVA pathways in the monomer formation stage contain oxygen-susceptible enzymes that lose the catalytic function in contact with molecular oxygen. The last two steps of the MEP pathway are catalyzed by two oxygen-susceptible enzymes (IspG and IspH) that contain Fe–S clusters (Zhao et al. 2013) (Fig. 2). Similarly, the fifth enzyme of the Route I MVA pathway (M5PDH) also contains a Fe–S cluster (Hayakawa et al. 2018). In contrast, in the diverged MVA pathways (Routes II, III, and IV), M5PDH is replaced by a GHMP enzyme that is oxygen-insusceptible. The adaptation of oxygen-insusceptible MVA pathway has been associated with the aerobic adaptation of host organisms toward eukaryogenesis, in response to Earth's oxygenation (Hoshino and Gaucher 2021).

Bacteria further acquired a large number of PTs and cyclases that engage in the modification of secondary metabolites (e.g. ABBA family, F family; Fig. 18). Secondary metabolites are generally associated with host organism's fitness (Firn and Jones 2000). It is generally thought that secondary metabolism is a late evolutionary trait in biology as a result of functional divergence from more ancestral primary metabolic pathways (Jenke-Kodama et al. 2008). Ribosomal and nonribosomal peptides and polyketides are common secondary metabolites in Bacteria (Wang et al. 2014). In particular, the polyketide biosynthesis pathway is largely homologous to the bacterial fatty acid biosynthesis pathway (Jenke-Kodama et al. 2005). Hence, the entire polyketide biosynthesis is inferred to have evolved within Bacteria. The distribution of secondary metabolites is mostly limited to terrestrial and aerobic species, although the marine realm is potentially a new frontier for yet-unknown secondary metabolites (Petersen et al. 2020). In recent years, some anaerobic bacteria are also found to possess the capacity to produce secondary metabolites (Letzel et al. 2013, 2014). Yet, the distribution of secondary metabolite biosynthesis genes is limited mostly to soil species, similar to aerobic species. In contrast, anaerobic species in harsh environments such as hot springs nearly completely lack those genes (Letzel et al. 2013).

Meanwhile, secondary metabolites are largely absent in Archaea and only a small number of examples are known thus far (e.g. archaeocins) (Besse et al. 2015, Charlesworth and Burns 2015). Recent genome mining suggests the presence of potential biosynthesis genes in Archaea (e.g. Chen et al. 2020b, Sharrar et al. 2020). However, the distribution of those genes is still largely limited to haloarchaea. PTs and cyclases that are associated with secondary metabolism are not found in Archaea and consistently, no terpenoid-associated secondary metabolites are detected in the domain. The overall scarcity of secondary metabolite biosynthesis genes in Archaea suggests their late horizontal origins from Bacteria, even if those genes are found in Archaea. Regardless of the ultimate origin, the appreciable diversification of secondary metabolites and associated terpenoids was most likely due to the

land colonization of aerobic bacteria, fungi, and land plants that established the modern terrestrial ecosystem. Therefore, the majority of the structural diversity of modern terpenoids, which are represented by various forms of mono-, sesqui-, di-, and sesterterpenoids, appeared relatively recently (i.e. Phanerozoic) in the entire history of terpenome evolution.

Conclusion and prospects

The microbial terpenome has a long evolutionary history and is deeply intertwined in the evolution of biochemistry in Bacteria and Archaea. The apparent enormous diversity of terpenoids long prevented us from grasping underlying principles that shaped the modern terpenome across the three domains of life. The complexity of terpenoid biosynthesis results from the combination of the functional diversification of ancestral terpenoid biosynthesis enzymes and the acquisition of novel enzymes. This review disentangled the complex history of the terpenome and shed light on the stepwise evolution of terpenoid biosynthesis. The comprehensive phylogenomic comparisons between Bacteria and Archaea about terpenoid biosynthesis reveal important similarities that reflect the ancient origin of terpenoid biosynthesis before the divergence of the two domains and also intriguing differences that are derived from domain-specific diversifications of terpenoids. Accelerating advancements of genomics now enable us to reconstruct the entire evolutionary history of individual classes of biomolecules, as exemplified for terpenoids in this review. Future work will include similar phylogenomic reconstructions for other biomolecules, including nucleic acids, saccharides, cofactors, and other lipids. Detailed stepwise comparisons between the reconstructed chronology of biomolecules and geochemical settings at different geological periods will ultimately provide us the whole picture for the evolutionary trajectory of biochemistry across 4 billion years of Earth's history. Additionally, the genetic repertoire for terpenoid biosynthesis, which is comprehensively described in this review, will enhance the identification of yet-uncharacterized terpenoids in nature and will also provide guides for directed evolution to engineer novel bioactive terpenoids.

Acknowledgments

Y.H. received funding from the DFG Priority program 2237. L.V. received funding from the Soehngen Institute for Anaerobic Microbiology (SIAM) through a Gravitation Grant (024.002.002) from the Dutch Ministry of Education, Culture and Science (OCW). The authors thank Professor Hisashi Hemmi for fruitful comments on the manuscript.

Supplementary data

Supplementary data is available at *FEMSRE Journal* online.

Conflict of interest statement. None declared.

Data availability

All study data are present in the article and/or supplementary information. Additional and raw data related to this paper are available from the authors on request.

References

Abby SS, Kazemzadeh K, Vragniau C *et al.* Advances in bacterial pathways for the biosynthesis of ubiquinone. *Biochim Biophys Acta Bioenerg* 2020;**1861**:148259.

Ahrazem O, Gómez-Gómez L, Rodrigo MJ *et al.* Carotenoid cleavage oxygenases from microbes and photosynthetic organisms: features and functions. *Int J Mol Sci* 2016;**17**:1781.

Albers S-V, Meyer BH. The archaeal cell envelope. *Nat Rev Microbiol* 2011;**9**:414–26.

Allen KN, Imperiali B. Structural and mechanistic themes in glycoconjugate biosynthesis at membrane interfaces. *Curr Opin Struct Biol* 2019;**59**:81–90.

Alzari PM, ne Souchon H, Dominguez R. The crystal structure of endoglucanase CelA, a family 8 glycosyl hydrolase from *Clostridium thermocellum*. *Structure* 1996;**4**:265–75.

Aoki M, Vinokur J, Motoyama K *et al.* Crystal structure of mevalonate 3,5-bisphosphate decarboxylase reveals insight into the evolution of decarboxylases in the mevalonate metabolic pathways. *J Biol Chem* 2022;**298**:102111.

Arnison PG, Bibb MJ, Bierbaum G,et al. Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat Prod Rep* 2013;**30**:108–60.

Azami Y, Hattori A, Nishimura H *et al.* (R)-mevalonate 3-phosphate is an intermediate of the mevalonate pathway in thermoplasma acidophilum. *J Biol Chem* 2014;**289**:15957–67.

Baunach M, Franke J, Hertweck C. Terpenoid biosynthesis off the beaten track: unconventional cyclases and their impact on biomimetic synthesis. *Angew Chem Int Ed* 2014;**9**:2604–26.

Beifuss U, Tietze M. Methanophenazine and other natural biologically active phenazines. In: Mulzer J (ed.), *Natural Products Synthesis II: Targets, Methods, Concepts*, Berlin, Heidelberg: Springer, 2005, 77–113.

Béjà O, Lanyi J. Nature's toolkit for microbial rhodopsin ion pumps. *Proc Natl Acad Sci USA* 2014;**111**:6538–9.

Belin BJ, Busset N, Giraud E *et al.* Hopanoid lipids: from membranes to plant–bacteria interactions. *Nat Rev Microbiol* 2018;**16**:304–15.

Belt ST, Allard WG, Massé G *et al.* Highly branched isoprenoids (HBIs): identification of the most common and abundant sedimentary isomers. *Geochim Cosmochim Acta* 2000;**64**:3839–51.

Besse A, Peduzzi J, Rebuffat S *et al.* Antimicrobial peptides and proteins in the face of extremes: lessons from archaeocins. *Biochimie* 2015;**118**:344–55.

Block A, Fristedt R, Rogers S,et al. Functional modeling identifies paralogous solanesyl-diphosphate synthases that assemble the side chain of plastoquinone-9 in plastids *J Biol Chem* 2013;**288**:27594–606.

Bloor S, Michurin I, Titchiner GR *et al.* Prenylated flavins: structures and mechanisms. *FEBS J* 2022. https://doi.org/10.1111/febs.16371.

Bonitz T, Alva V, Saleh O *et al.* Evolutionary relationships of microbial aromatic prenyltransferases. *PLoS ONE* 2011;**6**:e27336.

Bordo D, Bork P. The rhodanese/Cdc25 phosphatase superfamily. *EMBO Rep* 2002;**3**:741–6.

Boucher Y, Doolittle W. The role of lateral gene transfer in the evolution of isoprenoid biosynthesis pathways. *Mol Microbiol* 2000;**37**:703–16.

Bramkamp M, Lopez D. Exploring the existence of lipid rafts in bacteria. *Microbiol Mol Biol Rev* 2015;**79**:81.

Braun V, Hantke K. Lipoproteins: structure, function, biosynthesis. In: Kuhn A (ed.), *Bacterial Cell Walls and Membranes*. Cham: Springer International Publishing, 2019, 39–77.

Brown LS. Eubacterial rhodopsins—unique photosensors and diverse ion pumps. *Biochim Biophys Acta* 2014;**1837**:553–61.

Bryant DA, Frigaard N-U. Prokaryotic photosynthesis and phototrophy illuminated. *Trends Microbiol* 2006;**14**:488–96.

Byers DM, Gong H. Acyl carrier protein: structure–function relationships in a conserved multifunctional protein family. *Biochem Cell Biol* 2007;**85**:649–62.

Caforio A, Driessen AJM. Archaeal phospholipids: structural properties and biosynthesis. *Biochim Biophys Acta Mol Cell Biol Lipids* 2017;**1862**:1325–39.
Cao R, Zhang Y, Mann F,et al. Diterpene cyclases and the nature of the isoprene fold. *Proteins* 2010;**78**:2417–32.
Castelle CJ, Banfield JF. Major new microbial groups expand diversity and alter our understanding of the tree of Life. *Cell* 2018;**172**:1181–97.
Castelle CJ, Méheust R, Jaffe AL *et al.* Protein family content uncovers lineage relationships and bacterial pathway maintenance mechanisms in DPANN Archaea. *Front Microbiol* 2021;**12**: 34140936.
Charlesworth JC, Burns BP. Untapped resources: biotechnological potential of peptides and secondary metabolites in archaea. *Archaea* 2015;**2015**:282035.
Chazan A, Rozenberg A, Mannen K *et al.* Diverse heliorhodopsins detected via functional metagenomics in freshwater actinobacteria, chloroflexi and archaea. *Environ Microbiol* 2022;**24**:110–21.
Chen A, Dale Poulter C, Kroon PA. Isoprenyl diphosphate synthases: protein sequence comparisons, a phylogenetic tree, and predictions of secondary structure. *Protein Sci* 1994;**3**:600–7.
Chen A, Poulter CD. Purification and characterization of farnesyl diphosphate/geranylgeranyl diphosphate synthase. A thermostable bifunctional enzyme from *Methanobacterium thermoautotrophicum*. *J Biol Chem* 1993;**268**:11002–7.
Chen A, Zhang D, Poulter CD. (S)-geranylgeranylglyceryl phosphate synthase. Purification and characterization of the first pathway-specific enzyme in archaebacterial membrane lipid biosynthesis. *J Biol Chem* 1993;**268**:21701–5.
Chen IA, Walde P. From self-assembled vesicles to protocells. *Cold Spring Harb Perspect Biol* 2010;**2**:a002170.
Chen C-C, Zhang L, Yu X *et al.* Versatile cis-isoprenyl diphosphate synthase superfamily members in catalyzing carbon–carbon bond formation. *ACS Catal* 2020a;**10**:4717–25.
Chen R, Wong HL, Kindler GS *et al.* Discovery of an abundance of biosynthetic gene clusters in Shark Bay microbial mats. *Front Microbiol* 2020b;**11**:1950.
Cheng X, Smith JC. Biological membrane organization and cellular signaling. *Chem Rev* 2019;**119**:5849–80.
Cherian S, Ryu SB, Cornish K. Natural rubber biosynthesis in plants, the rubber transferase complex, and metabolic engineering progress and prospects. *Plant Biotechnol J* 2019;**17**: 2041–61.
Christianson DW. Structural and chemical biology of terpenoid cyclases. *Chem Rev* 2017;**117**:11570–648.
Coleman GA, Davín AA, Mahendrarajah TA *et al.* A rooted phylogeny resolves early bacterial evolution. *Science* 2021;**372**:eabe0511.
Coleman GA, Pancost RD, Williams TA. Investigating the origins of membrane phospholipid biosynthesis genes using outgroup-free rooting. *Genome Biol Evol* 2019;**11**:883–98.
Collins MD, Jones D. Distribution of isoprenoid quinone structural types in bacteria and their taxonomic implication. *Microbiol Rev* 1981;**45**:316–54.
Cotrim CA, Weidner A, Strehmel N *et al.* A distinct aromatic prenyltransferase associated with the Futalosine pathway. *ChemistrySelect* 2017;**2**:9319–25.
Dairi T. An alternative menaquinone biosynthetic pathway operating in microorganisms: an attractive target for drug discovery to pathogenic *Helicobacter* and *Chlamydia* strains. *J Antibiot* 2009;**62**:347–52.
Dairi T. Biosynthetic genes and enzymes of isoprenoids produced by actinomycetes. In: Bach JT, Rohmer M (eds), *Isoprenoid Synthesis in Plants and Microorganisms: New Concepts and Experimental Approaches* New York: Springer, 2013, 29–49.
Dar D, Thomashow LS, Weller DM *et al.* Global landscape of phenazine biosynthesis and biodegradation reveals species-specific colonization patterns in agricultural soils and crop microbiomes. *Elife* 2020;**9**:e59726.
Daruwalla A, Zhang J, Lee Ho J *et al.* Structural basis for carotenoid cleavage by an archaeal carotenoid dioxygenase. *Proc Natl Acad Sci USA* 2020;**117**:19914–25.
DasSarma S, Schwieterman EW. Early evolution of purple retinal pigments on Earth and implications for exoplanet biosignatures. *Int J Astrobiol* 2021;**20**:241–50.
De Bruyne L, Höfte M, De Vleesschauwer D. Connecting growth and defense: the emerging roles of brassinosteroids and gibberellins in plant innate immunity. *Mol Plant* 2014;**7**:943–59.
Deamer D, Dworkin JP, Sandford SA *et al.* The first cell membranes. *Astrobiology* 2002;**2**:371–81.
Degli Esposti M. A journey across genomes uncovers the origin of ubiquinone in cyanobacteria. *Genome Biol Evol* 2017;**9**: 3039–53.
Dellas N, Thomas ST, Manning G *et al.* Discovery of a metabolic alternative to the classical mevalonate pathway. *Elife* 2013;**2**: e00672.
Désaubry L, Nakatani Y, Ourisson G. Toward higher polyprenols under 'prebiotic' conditions. *Tetrahedron Lett* 2003;**44**:6959–61.
Dibrova DV, Galperin MY, Mulkidjanian AY. Phylogenomic reconstruction of archaeal fatty acid metabolism. *Environ Microbiol* 2014;**16**:907–18.
Dickschat JS. Bacterial diterpene biosynthesis. *Angew Chem Int Ed* 2019;**58**:15964–76.
Dickschat JS. Bacterial terpene cyclases. *Nat Prod Rep* 2016;**33**:87–110.
Dombrowski N, Lee J-H, Williams TA *et al.* Genomic diversity, lifestyles and evolutionary origins of DPANN archaea. *FEMS Microbiol Lett* 2019;**366**:fnz008.
Doud EH, Perlstein DL, Wolpert M *et al.* Two distinct mechanisms for TIM barrel prenyltransferases in bacteria. *J Am Chem Soc* 2011;**133**:1270–3.
Dumelin CE, Chen Y, Leconte AM *et al.* Discovery and biological characterization of geranylated RNA in bacteria. *Nat Chem Biol* 2012;**8**:913–9.
Egan AJF, Errington J, Vollmer W. Regulation of peptidoglycan synthesis and remodelling. *Nat Rev Microbiol* 2020;**18**:446–60.
Eichler J, Adams Michael WW. Posttranslational protein modification in archaea. *Microbiol Mol Biol Rev* 2005;**69**:393–425.
Eichler J, Guan Z. Lipid sugar carriers at the extremes: the phosphodolichols archaea use in N-glycosylation. *Biochim Biophys Acta* 2017;**1862**:589–99.
Elling FJ, Becker KW, Könneke M *et al.* Respiratory quinones in archaea: phylogenetic distribution and application as biomarkers in the marine environment. *Environ Microbiol* 2016;**18**: 692–707.
Eme L, Spang A, Lombard J *et al.* Archaea and the origin of eukaryotes. *Nat Rev Microbiol* 2017;**15**:711.
Faylo JL, Ronnebaum TA, Christianson DW. Assembly-line catalysis in bifunctional terpene synthases. *Acc Chem Res* 2021; **54**:3780–91.
Felemban A, Braguy J, Zurbriggen MD *et al.* Apocarotenoids involved in plant development and stress response. *Front Plant Sci* 2019;**10**:1168.
Feng Y, Morgan RML, Fraser PD *et al.* Crystal structure of geranylgeranyl pyrophosphate synthase (CrtE) involved in cyanobacterial terpenoid biosynthesis. *Front Plant Sci* 2020;**11**:589.

Firn RD, Jones CG. The evolution of secondary metabolism – a unifying model. *Mol Microbiol* 2000;**37**:989–94.
Frank A, Groll M. The methylerythritol phosphate pathway to isoprenoids. *Chem Rev* 2017;**117**:5675–703.
Franza T, Gaudu P. Quinones: more than electron shuttles. *Res Microbiol* 2022;**173**:103953.
Frébort I, Kowalska M, Hluska T *et al.* Evolution of cytokinin biosynthesis and degradation. *J Exp Bot* 2011;**62**:2431–52.
Fujihashi M, Sato T, Tanaka Y,et al. Crystal structure and functional analysis of large-terpene synthases belonging to a newly found subclass. *Chem Sci* 2018;**9**:3754–8.
Fujiwara S, Yamanaka A, Hirooka K *et al.* Temperature-dependent modulation of farnesyl diphosphate/geranylgeranyl diphosphate synthase from hyperthermophilic archaea. *Biochem Biophys Res Commun* 2004;**325**:1066–74.
Gao J, Ko T-P, Chen L *et al.* "Head-to-middle" and "Head-to-tail" cis-prenyl transferases: structure of isosesquilavandulyl diphosphate synthase. *Angew Chem Int Ed* 2018;**57**:683–7.
Genschel U. Coenzyme A biosynthesis: reconstruction of the pathway in archaea and an evolutionary scenario based on comparative genomics. *Mol Biol Evol* 2004;**21**:1242–51.
Gericke O, Hansen NL, Pedersen GB *et al.* Nerylneryl diphosphate is the precursor of serrulatane, viscidane and cembrane-type diterpenoids in Eremophila species. *BMC Plant Biol* 2020;**20**:91.
Glass JB, Elbon CE, Williams LD. Something old, something new, something borrowed, something blue: the anaerobic microbial ancestry of aerobic respiration. *Trends Microbiol* 2023;**31**:135–41.
Gold DA, Caron A, Fournier GP *et al.* Paleoproterozoic sterol biosynthesis and the rise of oxygen. *Nature* 2017;**543**:420.
Grabińska KA, Park EJ, Sessa WC. *cis*-prenyltransferase: new insights into protein glycosylation, rubber synthesis, and human diseases. *J Biol Chem* 2016;**291**:18582–90.
Greule A, Stok JE, De Voss JJ *et al.* Unrivalled diversity: the many roles and reactions of bacterial cytochromes P450 in secondary metabolism. *Nat Prod Rep* 2018;**35**:757–91.
Guldan H, Matysik F-M, Bocola M *et al.* Functional assignment of an enzyme that catalyzes the synthesis of an archaea-type ether lipid in bacteria. *Angew Chem Int Ed* 2011;**50**:8188–91.
Haft DH, Payne SH, Selengut JD. Archaeosortases and exosortases are widely distributed systems linking membrane transit with posttranslational modification. *J Bacteriol* 2012;**194**:36–48.
Harms MJ, Thornton JW. Evolutionary biochemistry: revealing the historical and physical causes of protein properties. *Nat Rev Genet* 2013;**14**:559–71.
Hartley MD, Imperiali B. At the membrane frontier: a prospectus on the remarkable evolutionary conservation of polyprenols and polyprenyl-phosphates. *Arch Biochem Biophys* 2012;**517**:83–97.
Hashimoto H, Uragami C, Cogdell RJ. Carotenoids and photosynthesis. In: Stange C (ed.), *Carotenoids in Nature: Biosynthesis, Regulation and Function*. Cham: Springer International Publishing, 2016, 111–39.
Hayakawa H, Motoyama K, Sobue F *et al.* Modified mevalonate pathway of the archaeon *Aeropyrum pernix* proceeds via *trans*-anhydromevalonate 5-phosphate. *Proc Natl Acad Sci USA* 2018;**115**:10034.
He H, Bian G, Herbst-Gervasoni CJ *et al.* Discovery of the cryptic function of terpene cyclases as aromatic prenyltransferases. *Nat Commun* 2020;**11**:3958.
Hederstedt L. Heme A biosynthesis. *Biochim Biophys Acta Bioenerg* 2012;**1817**:920–7.
Hirooka K, Shioda S, Okada M. Identification of critical residues for the catalytic activity of ComQ, a Bacillus prenylation enzyme for quorum sensing, by using a simple bioassay system. *Biosci Biotechnol Biochem* 2020;**84**:347–57.
Hohmann-Marriott M, Blankenship R. Evolution of photosynthesis. *Annu Rev Plant Biol* 2011;**62**:515–48.
Holtzapple E, Schmidt-Dannert C. Biosynthesis of isoprenoid wax ester in marinobacter hydrocarbonoclasticus DSM 8798: identification and characterization of isoprenoid coenzyme A synthetase and wax ester synthases. *J Bacteriol* 2007;**189**:3804–12.
Hoshino Y, Gaucher EA. Evolution of bacterial steroid biosynthesis and its impact on eukaryogenesis. *Proc Natl Acad Sci USA* 2021;**118**:e2101276118.
Hoshino Y, Gaucher EA. On the origin of isoprenoid biosynthesis. *Mol Biol Evol* 2018;**35**:2185–97.
Hou A, Dickschat JS. The biosynthetic gene cluster for sestermobaraenes—discovery of a geranylfarnesyl diphosphate synthase and a multiproduct sesterterpene synthase from *Streptomyces mobaraensis*. *Angew Chem Int Ed* 2020;**59**:19961–5.
Huang S, Elsayed Somayah S, Lv M,et al. Biosynthesis of neocarazostatin A reveals the sequential carbazole prenylation and hydroxylation in the tailoring steps. *Chem Biol* 2015;**22**:1633–42.
Ishiyama A, Otoguro K, Namatame M,et al. In vitro and in vivo antitrypanosomal activitiy of two microbial metabolites, KS-505a and alazopeptin. *J Antibiot* 2008;**61**:627–32.
Iwashkiw JA, Vozza NF, Kinsella RL *et al.* Pour some sugar on it: the expanding world of bacterial protein O-linked glycosylation. *Mol Microbiol* 2013;**89**:14–28.
Jain S, Caforio A, Driessen AJM. Biosynthesis of archaeal membrane ether lipids. *Front Microbiol* 2014;**5**:641.
Jarrell KF, Ding Y, Meyer BH *et al.* N-linked glycosylation in archaea: a structural, functional, and genetic analysis. *Microbiol Mol Biol Rev* 2014;**78**:304–41.
Jenke-Kodama H, Müller R, Dittmann E. Evolutionary mechanisms underlying secondary metabolite diversity. In: Petersen F, Amstutz R (eds), *Natural Compounds as Drugs*. **Vol. 1**. Basel: Birkhäuser, 2008, 119–40.
Jenke-Kodama H, Sandmann A, Müller R *et al.* Evolutionary implications of bacterial polyketide synthases. *Mol Biol Evol* 2005;**22**:2027–39.
Jia Q, Brown R, Köllner TG *et al.* Origin and early evolution of the plant terpene synthase family. *Proc Natl Acad Sci USA* 2022;**119**:e2100361119.
Jiang H, Zhang X, Chen X *et al.* Protein lipidation: occurrence, mechanisms, biological functions, and enabling technologies. *Chem Rev* 2018;**118**:919–88.
Jones MB, Rosenberg JN, Betenbaugh MJ *et al.* Structure and synthesis of polyisoprenoids used in N-glycosylation across the three domains of life. *Biochim Biophys Acta Gen Subj* 2009;**1790**:485–94.
Jordan SF, Nee E, Lane N. Isoprenoids enhance the stability of fatty acid membranes at the emergence of life potentially leading to an early lipid divide. *Interface Focus* 2019;**9**:20190067.
Joshi S, Fedoseyenko D, Mahanta N *et al.* Novel enzymology in futalosine-dependent menaquinone biosynthesis. *Curr Opin Chem Biol* 2018;**47**:134–41.
Kim S-H, Lu W, Ahmadi MK *et al.* Atolypenes, tricyclic bacterial sesterterpenes discovered using a multiplexed in vitro Cas9-TAR gene cluster refactoring approach. *ACS Synth Biol* 2019;**8**:109–18.
Kim Y-S, Kim N-H, Yeom S-J *et al. In vitro* characterization of a recombinant blh protein from an uncultured marine bacterium as a beta-carotene 15,15'-dioxygenase. *J Biol Chem* 2009;**284**:15781–93.

Kirschning A. Coenzymes and their role in the evolution of life. *Angew Chem Int Ed* 2021;**60**:6242–69.
Klassen JL. Phylogenetic and evolutionary patterns in microbial carotenoid biosynthesis are revealed by comparative genomics. *PLoS ONE* 2010;**5**:e11257.
Kobayashi M, Kuzuyama T. Recent advances in the biosynthesis of carbazoles produced by actinomycetes. *Biomolecules* 2020;**10**: 1147.
Koga Y, Morii H. Biosynthesis of ether-type polar lipids in archaea and evolutionary considerations. *Microbiol Mol Biol Rev* 2007;**71**:97–120.
Koga Y. Early evolution of membrane lipids: how did the lipid divide occur?. *J Mol Evol* 2011;**72**:274–82.
Kontnik R, Bosak T, Butcher RA *et al.* Sporulenes, heptaprenyl metabolites from *Bacillus subtilis* spores. *Org Lett* 2008;**10**: 3551–4.
Koonin EV, Martin W. On the origin of genomes and cells within inorganic compartments. *Trends Genet* 2005;**21**:647–54.
Krubasik P, Kobayashi M, Sandmann G. Expression and functional analysis of a gene cluster involved in the synthesis of decaprenoxanthin reveals the mechanisms for C50 carotenoid formation. *Eur J Biochem* 2001;**268**:11432736.
Kupke T, Uebele M, Schmid D *et al.* Molecular characterization of lantibiotic-synthesizing enzyme EpiD reveals a function for bacterial dfp proteins in coenzyme A biosynthesis. *J Biol Chem* 2000;**275**:31838–46.
Kuzuyama T, Noel JP, Richard SB. Structural basis for the promiscuous biosynthetic prenylation of aromatic natural products. *Nature* 2005;**435**:983–7.
Lai JCY, Pearce BKD, Pudritz RE *et al.* Meteoritic abundances of fatty acids and potential reaction pathways in planetesimals. *Icarus* 2019;**319**:685–700.
Lane N, Allen JF, Martin W. How did LUCA make a living? Chemiosmosis in the origin of life. *Bioessays* 2010;**32**:271–80.
Lang D, Thoma R, Henn-Sax M *et al.* Structural evidence for evolution of the β/α barrel scaffold by gene duplication and fusion. *Science* 2000;**289**:1546–50.
Lange B, Rujan T, Martin W *et al.* Isoprenoid biosynthesis: the evolution of two ancient and distinct pathways across genomes. *Proc Natl Acad Sci USA* 2000;**97**:13172–7.
Layre E, Lee HJ, Young DC,et al. Molecular profiling of mycobacterium tuberculosis identifies tuberculosinyl nucleoside products of the virulence-associated enzyme Rv3378c. *Proc Natl Acad Sci USA* 2014;**111**:2978–83.
Leipe DD, Wolf YI, Koonin EV *et al.* Classification and evolution of P-loop gtpases and related ATPases11Edited by J. Thornton. *J Mol Biol* 2002;**317**:41–72.
Letzel A-C, Pidot SJ, Hertweck C. A genomic approach to the cryptic secondary metabolome of the anaerobic world. *Nat Prod Rep* 2013;**30**:392–428.
Letzel A-C, Pidot SJ, Hertweck C. Genome mining for ribosomally synthesized and post-translationally modified peptides (RiPPs) in anaerobic bacteria. *BMC Genom* 2014;**15**:983.
Leys D. Flavin metamorphosis: cofactor transformation through prenylation. *Curr Opin Chem Biol* 2018;**47**:117–25.
Li B, Yu JPJ, Brunzelle JS *et al.* Structure and mechanism of the lantibiotic cyclase involved in nisin biosynthesis. *Science* 2006;**311**: 1464–7.
Li K, Gustafson KR. Sesterterpenoids: chemistry, biology, and biosynthesis. *Nat Prod Rep* 2021;**38**:1251–81.
Li S, Lowell AN, Yu F *et al.* Hapalindole/ambiguine biogenesis is mediated by a cope rearrangement, C–C bond-forming cascade. *J Am Chem Soc* 2015;**137**:15366–9.
Li W. Bringing bioactive compounds into membranes: the UbiA superfamily of intramembrane aromatic prenyltransferases. *Trends Biochem Sci* 2016;**41**:356–70.
Liang P-H, Ko T-P, Wang AHJ. Structure, mechanism and function of prenyltransferases. *Eur J Biochem* 2002;**269**:3339–54.
Lloyd CT, Iwig DF, Wang B *et al.* Discovery, structure and mechanism of a tetraether lipid synthase. *Nature* 2022;**609**:197–203.
Lohr M, Schwender J, Polle JEW. Isoprenoid biosynthesis in eukaryotic phototrophs: a spotlight on algae. *Plant Sci* 2012;**185–186**: 9–22.
Lombard J, López-García P, Moreira D. An ACP-independent fatty acid synthesis pathway in archaea: implications for the origin of phospholipids. *Mol Biol Evol* 2012;**29**:3261–5.
Lombard J, Moreira D. Origins and early evolution of the mevalonate pathway of isoprenoid biosynthesis in the three domains of life. *Mol Biol Evol* 2011;**28**:87–99.
Lombard J. Early evolution of polyisoprenol biosynthesis and the origin of cell walls. *PeerJ* 2016b;**4**:e2626.
Lombard J. The multiple evolutionary origins of the eukaryotic N-glycosylation pathway. *Biol Direct* 2016a;**11**:36.
Lopez D, Koch G. Exploring functional membrane microdomains in bacteria: an overview. *Curr Opin Microbiol* 2017;**36**:76–84.
Lübben M, Morand K. Novel prenylated hemes as cofactors of cytochrome oxidases. archaea have modified hemes A and O. *J Biol Chem* 1994;**269**:21473–9.
Luo Y, Li W, Ju J,et al. Functional characterization of TtnD and TtnF, unveiling new insights into tautomycetin biosynthesis. *J Am Chem Soc* 2010;**132**:6663–71.
Lyons T, Reinhard C, Planavsky N. The rise of oxygen in Earth's early ocean and atmosphere. *Nature* 2014;**506**:307–15.
Macheroux P, Kappes B, Ealick SE. Flavogenomics – a genomic and structural view of flavin-dependent proteins. *FEBS J* 2011;**278**:2625–34.
Manat G, Roure S, Auger R *et al.* Deciphering the metabolism of undecaprenyl-phosphate: the bacterial cell-wall unit carrier at the membrane frontier. *Microb Drug Resist* 2014;**20**:199–214.
Mann FM, Thomas JA, Peters RJ. Rv0989c encodes a novel (E)-geranyl diphosphate synthase facilitating decaprenyl diphosphate biosynthesis in *Mycobacterium tuberculosis*. *FEBS Lett* 2011;**585**:549–54.
Maresca JA, Graham JE, Wu M *et al.* Identification of a fourth family of lycopene cyclases in photosynthetic bacteria. *Proc Natl Acad Sci USA* 2007;**104**:11784–9.
Markov GV, Laudet V. Small molecules as products of evolution. *Curr Biol* 2022;**32**:R100–5.
Marshall CJ. Protein prenylation: a mediator of protein-protein interactions. *Science* 1993;**259**:1865–6.
Marshall SA, Payne KAP, Fisher K *et al.* The UbiX flavin prenyltransferase reaction mechanism resembles class I terpene cyclase chemistry. *Nat Commun* 2019;**10**:2357.
Marshall SA, Payne KAP, Leys D. The UbiX-UbiD system: the biosynthesis and use of prenylated flavin (prFMN). *Arch Biochem Biophys* 2017;**632**:209–21.
Martin W, Baross J, Kelley D *et al.* Hydrothermal vents and the origin of life. *Nat Rev Microbiol* 2008;**6**:805–14.
Martin W, Russell M. On the origins of cells: a hypothesis for the evolutionary transitions from abiotic geochemistry to chemoautotrophic prokaryotes, and from prokaryotes to nucleated cells. *Philos Trans R Soc Lond, B, Biol Sci* 2003;**358**:59.
Matsumi R, Atomi H, Driessen A *et al.* Isoprenoid biosynthesis in archaea–biochemical and evolutionary implications. *Res Microbiol* 2011;**162**:39–52.

Maurer-Stroh S, Washietl S, Eisenhaber F. Protein prenyltransferases. *Genome Biol* 2003;**4**:212.
Mavrodi DV, Peever TL, Mavrodi OV,et al. Diversity and evolution of the phenazine biosynthesis pathway. *Appl Environ Microbiol* 2010;**76**:866–79.
McCollom TM, Seewald JS. Abiotic synthesis of organic compounds in deep-sea hydrothermal environments. *Chem Rev* 2007;**107**:382–401.
McCollom TM. Laboratory simulations of abiotic hydrocarbon formation in Earth's deep subsurface. *Rev Mineral Geochem* 2013;**75**:467–94.
McIntosh JA, Donia MS, Nair SK *et al.* Enzymatic basis of ribosomal peptide prenylation in cyanobacteria. *J Am Chem Soc* 2011;**133**:13698–705.
McIntosh JA, Donia MS, Schmidt EW. Ribosomal peptide natural products: bridging the ribosomal and nonribosomal worlds. *Nat Prod Rep* 2009;**26**:537–59.
McNamara DE, Cascio D, Jorda J *et al.* Structure of dihydromethanopterin reductase, a cubic protein cage for redox transfer. *J Biol Chem* 2014;**289**:8852–64.
Meyer BH, Adam PS, Wagstaff BA *et al.* Agl24 is an ancient archaeal homolog of the eukaryotic N-glycan chitobiose synthesis enzymes. *Elife* 2022;**11**:e67448.
Miziorko H. Enzymes of the mevalonate pathway of isoprenoid biosynthesis. *Arch Biochem Biophys* 2011;**505**:131–43.
Moody ERR, Mahendrarajah TA, Dombrowski N *et al.* An estimate of the deepest branches of the tree of life from ancient vertically evolving genes. *Elife* 2022;**11**:e66695.
Moosmann P, Ecker F, Leopold-Messer S *et al.* A monodomain class II terpene cyclase assembles complex isoprenoid scaffolds. *Nat Chem* 2020;**12**:968–72.
Nakajima Y, Kojima K, Kashiyama Y *et al.* Bacterium lacking a known gene for retinal biosynthesis constructs functional rhodopsins. *Microbes Environ* 2020;**35**:ME20085.
Nakano C, Oshima M, Kurashima N *et al.* Identification of a new diterpene biosynthetic gene cluster that produces O-methylkolavelool in *Herpetosiphon aurantiacus*. *ChemBioChem* 2015;**16**:772–81.
Nakatani Y, Ribeiro N, Streiff S *et al.* Search for the most 'primitive' membranes and their reinforcers: a review of the polyprenyl phosphates theory. *Orig Life Evol Biosph* 2014;**44**:197–208.
Nakayama H, Kurokawa K, Lee BL. Lipoproteins in bacteria: structures and biosynthetic pathways. *FEBS J* 2012;**279**:4247–68.
Nes WD. Biosynthesis of cholesterol and other sterols. *Chem Rev* 2011;**111**:6423–51.
Nishimura H, Azami Y, Miyagawa M *et al.* Biochemical evidence supporting the presence of the classical mevalonate pathway in the thermoacidophilic archaeon *Sulfolobus solfataricus*. *J Biochem* 2013;**153**:415–20.
Nothaft H, Szymanski CM. Protein glycosylation in bacteria: sweeter than ever. *Nat Rev Microbiol* 2010;**8**:765–78.
Nowicka B, Kruk J. Occurrence, biosynthesis and function of isoprenoid quinones. *Biochim Biophys Acta Bioenerg* 2010;**1797**:1587–605.
O'Toole KH, Imperiali B, Allen KN. Glycoconjugate pathway connections revealed by sequence similarity network analysis of the monotopic phosphoglycosyl transferases. *Proc Natl Acad Sci USA* 2021;**118**:e2018289118.
Ogawa T, Emi K-i, Koga K *et al.* A cis-prenyltransferase from *Methanosarcina acetivorans* catalyzes both head-to-tail and nonhead-to-tail prenyl condensation. *FEBS J* 2016;**283**:2369–83.
Ogawa T, Yoshimura T, Hemmi H. Geranylfarnesyl diphosphate synthase from methanosarcina mazei: different role, different evolution. *Biochem Biophys Res Commun* 2010;**393**:16–20.
Okada M, Unno H, Emi K-I *et al.* A versatile *cis*-prenyltransferase from *Methanosarcina mazei* catalyzes both C- and O-prenylations. *J Biol Chem* 2021;**296**:100679.
Oldfield E, Lin F-Y. Terpene biosynthesis: modularity rules. *Angew Chem Int Ed* 2012;**51**:1124–37.
Ostash B, Doud EH, Lin C *et al.* Complete characterization of the seventeen step moenomycin biosynthetic pathway. *Biochemistry* 2009;**48**:8830–41.
Ozaki T, Shinde SS, Gao L *et al.* Enzymatic formation of a skipped methyl-substituted octaprenyl side chain of longestin (KS-505a): involvement of homo-IPP as a common extender unit. *Angew Chem Int Ed* 2018;**57**:6629–32.
Ozaki T, Zhao P, Shinada T *et al.* Cyclolavandulyl skeleton biosynthesis via both condensation and cyclization catalyzed by an unprecedented member of the cis-isoprenyl diphosphate synthase superfamily. *J Am Chem Soc* 2014;**136**:4837–40.
Pan J-J, Solbiati JO, Ramamoorthy G *et al.* Biosynthesis of squalene from farnesyl diphosphate in bacteria: three steps catalyzed by three enzymes. *ACS Cent Sci* 2015;**1**:77–82.
Pan X, Du W, Zhang X *et al.* Discovery, structure, and mechanism of a class II sesquiterpene cyclase. *J Am Chem Soc* 2022;**144**:22067–74.
Pancrace C, Ishida K, Briand E,et al. Unique biosynthetic pathway in bloom-forming cyanobacterial genus microcystis jointly assembles cytotoxic aeruginoguanidines and microguanidines. *ACS Chem Biol* 2019;**14**:67–75.
Park H-W, Boduluri SR, Moomaw JF *et al.* Crystal structure of protein farnesyltransferase at 2.25 angstrom resolution. *Science* 1997;**275**:1800–5.
Pelosi L, Ducluzeau A-L, Loiseau L *et al.* Evolution of ubiquinone biosynthesis: multiple proteobacterial enzymes with various regioselectivities to catalyze three contiguous aromatic hydroxylation reactions. *Msystems* 2016;**1**:e00091–00016.
Peterhoff D, Beer B, Rajendran C *et al.* A comprehensive analysis of the geranylgeranylglyceryl phosphate synthase enzyme family identifies novel members and reveals mechanisms of substrate specificity and quaternary structure organization. *Mol Microbiol* 2014;**92**:885–99.
Peters JW, Schut GJ, Boyd ES *et al.* [FeFe]- and [NiFe]-hydrogenase diversity, mechanism, and maturation. *Biochim Biophys Acta* 2015;**1853**:1350–69.
Peters RJ. Two rings in them all: the labdane-related diterpenoids. *Nat Prod Rep* 2010;**27**:1521–30.
Petersen L-E, Kellermann MY, Schupp PJ. Secondary metabolites of marine microbes: from natural products chemistry to chemical ecology. In: Jungblut S, Liebich V, Bode-Dalby M (eds), *YOUMARES 9 – The Oceans: Our Research, Our Future: Proceedings of the 2018 Conference for YOUng MArine RESearcher in Oldenburg, Germany*. Cham: Springer International Publishing, 2020, p. 159–80.
Pichersky E, Raguso RA. Why do plants produce so many terpenoid compounds?. *New Phytol* 2018;**220**:692–702.
Pinhassi J, DeLong Edward F, Béjà O *et al.* Marine bacterial and archaeal ion-pumping rhodopsins: genetic diversity, physiology, and ecology. *Microbiol Mol Biol Rev* 2016;**80**:929–54.
Planavsky NJ, Crowe SA, Fakhraee M *et al.* Evolution of the structure and impact of Earth's biosphere. *Nat Rev Earth Environ* 2021;**2**:123–39.
Pohlschroder M, Pfeiffer F, Schulze S *et al.* Archaeal cell surface biogenesis. *FEMS Microbiol Rev* 2018;**42**:694–717.
Proctor MS, Sutherland GA, Canniffe DP *et al.* The terminal enzymes of (bacterio)chlorophyll biosynthesis. *R Soc Open Sci* 2022;**9**:211903.
Quin MB, Flynn CM, Schmidt-Dannert C. Traversing the fungal terpenome. *Nat Prod Rep* 2014;**31**:1449–73.

Ravcheev DA, Thiele I. Genomic analysis of the human gut microbiome suggests novel enzymes involved in quinone biosynthesis. *Front Microbiol* 2016;**7**:128.
Reddy GK, Leferink NGH, Umemura M *et al.* Exploring novel bacterial terpene synthases. *PLoS ONE* 2020;**15**:e0232220.
Rohmer M, Knani M, Simonin P *et al.* Isoprenoid biosynthesis in bacteria: a novel pathway for the early steps leading to isopentenyl diphosphate. *Biochem J* 1993;**295**(Pt 2):517–24.
Rojas AM, Fuentes G, Rausell A *et al.* The Ras protein superfamily: evolutionary tree and role of conserved amino acids. *J Cell Biol* 2012;**196**:189–201.
Röttig A, Steinbüchel A. Acyltransferases in bacteria. *Microbiol Mol Biol Rev* 2013;**77**:277–321.
Ruch S, Beyer P, Ernst H *et al.* Retinal biosynthesis in eubacteria: in vitro characterization of a novel carotenoid oxygenase from synechocystis sp. PCC 6803. *Mol Microbiol* 2005;**55**:1015–24.
Rudolf JD, Alsup TA, Xu B *et al.* Bacterial terpenome. *Nat Prod Rep* 2021;**38**:905–80.
Rudolf JD, Chang C-Y. Terpene synthases in disguise: enzymology, structure, and opportunities of non-canonical terpene synthases. *Nat Prod Rep* 2020;**37**:425–63.
Rudolf JD, Dong L-B, Cao H,et al. Structure of the ent-copalyl diphosphate synthase PtmT2 from *Streptomyces platensis* CB00739, a bacterial type II diterpene synthase. *J Am Chem Soc* 2016;**138**:10905–15.
Sáenz JP, Grosser D, Bradley AS *et al.* Hopanoids as functional analogues of cholesterol in bacterial membranes. *Proc Natl Acad Sci USA* 2015;**112**:11971–6.
Sakon J, Irwin D, Wilson DB *et al.* Structure and mechanism of endo/exocellulase E4 from *Thermomonospora fusca*. *Nat Struct Mol Biol* 1997;**4**:810–8.
Salvador-Castell M, Tourte M, Oger PM. In search for the membrane regulators of archaea. *Int J Mol Sci* 2019;**20**:4434.
Sandmann G. Carotenoid biosynthesis in the phylum actinobacteria. In: Misawa N (ed.), *Carotenoids: Biosynthetic and Biofunctional Approaches*. Singapore: Springer, 2021, 175–81.
Santana-Molina C, Henriques V, Hornero-Méndez D *et al.* The squalene route to C30 carotenoid biosynthesis and the origins of carotenoid biosynthetic pathways. *Proc Natl Acad Sci USA* 2022;**119**:e2210081119.
Santana-Molina C, Rivas-Marin E, Rojas AM *et al.* Origin and evolution of polycyclic triterpene synthesis. *Mol Biol Evol* 2020;**37**:1925–41.
Sato T, Yoshida S, Hoshino H *et al.* Sesquarterpenes (C_{35} terpenes) biosynthesized via the cyclization of a linear C_{35} isoprenoid by a tetraprenyl-β-curcumene synthase and a tetraprenyl-β-curcumene cyclase: identification of a new terpene cyclase. *J Am Chem Soc* 2011;**133**:9734–7.
Sato T. Unique biosynthesis of sesquarterpenes (C_{35} terpenes). *Biosci Biotechnol Biochem* 2013;**77**:1155–9.
Satta A, Esquirol L, Ebert BE *et al.* Molecular characterization of cyanobacterial short-chain prenyltransferases and discovery of a novel GGPP phosphatase. *FEBS J* 2022;**289**:6672–93.
Schiffer L, Barnard L, Baranowski ES *et al.* Human steroid biosynthesis, metabolism and excretion are differentially reflected by serum and urine steroid metabolomes: a comprehensive review. *J Steroid Biochem Mol Biol* 2019;**194**:105439.
Schoepp-Cothenet B, van Lis R, Atteia A *et al.* On the universal core of bioenergetics. *Biochim Biophys Acta Bioenerg* 2013;**1827**:79–93.
Schouten S, Hopmans EC, Sinninghe Damsté JS. The organic geochemistry of glycerol dialkyl glycerol tetraether lipids: a review. *Org Geochem* 2013;**54**:19–61.
Schroepfer GJ. Oxysterols: modulators of cholesterol metabolism and other processes. *Physiol Rev* 2000;**80**:361–554.
Schultz AW, Oh D-C, Carney JR *et al.* Biosynthesis and structures of cyclomarins and cyclomarazines, prenylated cyclic peptides of marine actinobacterial origin. *J Am Chem Soc* 2008;**130**:4507–16.
Schut GJ, Boyd ES, Peters JW *et al.* The modular respiratory complexes involved in hydrogen and sulfur metabolism by heterotrophic hyperthermophilic archaea and their evolutionary implications. *FEMS Microbiol Rev* 2013;**37**:182–203.
Schut GJ, Zadvornyy O, Wu C-H *et al.* The role of geochemistry and energetics in the evolution of modern respiratory complexes from a proton-reducing ancestor. *Biochim Biophys Acta* 2016;**1857**:958–70.
Scott AI. How were porphyrins and lipids synthesized in the RNA world?. *Tetrahedron Lett* 1997;**38**:4961–4.
Sephton MA. Organic compounds in carbonaceous meteorites. *Nat Prod Rep* 2002;**19**:292–311.
Sezgin E, Levental I, Mayor S *et al.* The mystery of membrane organization: composition, regulation and roles of lipid rafts. *Nat Rev Mol Cell Biol* 2017;**18**:361.
Shalaeva DN, Galperin MY, Mulkidjanian AY. Eukaryotic G protein-coupled receptors as descendants of prokaryotic sodium-translocating rhodopsins. *Biol Direct* 2015;**10**:63.
Sharrar AM, Crits-Christoph A, Méheust R *et al.* Bacterial secondary metabolite biosynthetic potential in soil varies with phylum, depth, and vegetation type. *Mbio* 2020;**11**:e00416–00420.
Shaw TR, Ghosh S, Veatch SL. Critical phenomena in plasma membrane organization and function. *Annu Rev Phys Chem* 2020;**72**:51–72.
Silhavy TJ, Kahne D, Walker S. The bacterial cell envelope. *Cold Spring Harb Perspect Biol* 2010;**2**:a000414.
Sivonen K, Leikoski N, Fewer DP *et al.* Cyanobactins—ribosomal cyclic peptides produced by cyanobacteria. *Appl Microbiol Biotechnol* 2010;**86**:1213–25.
Smanski MJ, Peterson RM, Huang SX *et al.* Bacterial diterpene synthases: new opportunities for mechanistic enzymology and engineered biosynthesis. *Cur Opin Chem Biol* 2012;**16**:132–41.
Smanski MJ, Yu Z, Casper J *et al.* Dedicated ent-kaurene and ent-atiserene synthases for platensimycin and platencin biosynthesis. *Proc Natl Acad Sci USA* 2011;**108**:13498–503.
Soderberg T, Poulter CD. *Escherichia coli* dimethylallyl diphosphate:tRNA dimethylallyltransferase: essential elements for recognition of tRNA substrates within the anticodon stem-loop. *Biochemistry* 2000;**39**:6546–53.
Sousa FL, Alves RJ, Ribeiro MA *et al.* The superfamily of heme–copper oxygen reductases: types and evolutionary considerations. *Biochim Biophys Acta Bioenerg* 2012;**1817**:629–37.
Storf S, Pfeiffer F, Dilks K *et al.* Mutational and bioinformatic analysis of Haloarchaeal lipobox-containing proteins. *Archaea* 2010;**2010**:410975.
Stowell EA, Ehrenberger MA, Lin Y-L *et al.* Structure-guided product determination of the bacterial type II diterpene synthase Tpn2. *Commun Chem* 2022;**5**:146.
Streiff S, Ribeiro N, Wu Z *et al.* "Primitive" membrane from polyprenyl phosphates and polyprenyl alcohols. *Chem Biol* 2007;**14**:313–9.
Summons RE, Bradley AS, Jahnke LL *et al.* Steroids, triterpenoids and molecular oxygen. *Philos Trans R Soc Lond, B, Biol Sci* 2006;**361**:951–68.
Summons RE, Welander PV, Gold DA. Lipid biomarkers: molecular tools for illuminating the history of microbial life. *Nat Rev Microbiol* 2021;**20**:174–85.

Surkont J, Pereira-Leal JB. Are there Rab gtpases in archaea?. *Mol Biol Evol* 2016;**33**:1833–42.
Sutcliffe IC, Harrington DJ, Hutchings MI. A phylum level analysis reveals lipoprotein biosynthesis to be a fundamental property of bacteria. *Protein Cell* 2012;**3**:163–70.
Swoboda JG, Campbell J, Meredith TC *et al.* Wall teichoic acid function, biosynthesis, and inhibition. *ChemBioChem* 2010;**11**: 35–45.
Szczupak P, Radzikowska-Cieciura E, Kulik K *et al. Escherichia coli* tRNA 2-selenouridine synthase SelU selects its prenyl substrate to accomplish its enzymatic function. *Bioorg Chem* 2022;**122**: 105739.
Szkopińska A, Płochocka D. Farnesyl diphosphate synthase; regulation of product specificity. *Acta Biochim Pol* 2005;**52**:45–55.
Tachibana A, Yano Y, Otani S *et al.* Novel prenyltransferase gene encoding farnesylgeranyl diphosphate synthase from a hyperthermophilic archaeon, *Aeropyrum pernix*. *Eur J Biochem* 2000;**267**: 321–8.
Taib N, Megrian D, Witwinowski J *et al.* Genome-wide analysis of the Firmicutes illuminates the diderm/monoderm transition. *Nat Ecol Evol* 2020;**4**:1661–72.
Takahashi S, Takagi H, Toyoda A *et al.* Biochemical characterization of a novel indole prenyltransferase from streptomyces sp. SN-593. *J Bacteriol* 2010;**192**:2839–51.
Takigawa H, Sugiyama M, Shibuya Y. C_{35}-terpenes from *Bacillus subtilis* KSM 6-10. *J Nat Prod* 2010;**73**:204–7.
Takishita K, Chikaraishi Y, Tanifuji G *et al.* Microbial eukaryotes that lack sterols. *J Eukaryot Microbiol* 2017;**64**:897–900.
Tao H, Lauterbach L, Bian G,et al. Discovery of non-squalene triterpenes. *Nature* 2022;**606**:414–9.
Taujale R, Venkat A, Huang L-C *et al.* Deep evolutionary analysis reveals the design principles of fold A glycosyltransferases. *Elife* 2020;**9**:e54532.
Tello M, Kuzuyama T, Heide L *et al.* The ABBA family of aromatic prenyltransferases: broadening natural product diversity. *Cell Mol Life Sci* 2008;**65**:1459–63.
Thauer RK, Kaster A-K, Seedorf H *et al.* Methanogenic archaea: ecologically relevant differences in energy conservation. *Nat Rev Microbiol* 2008;**6**:579–91.
Thoma R, Schulz-Gasch T, D'Arcy B *et al.* Insight into steroid scaffold formation from the structure of human oxidosqualene cyclase. *Nature* 2004;**432**:118–22.
Thomashow LS. Phenazines in the environment: microbes, habitats, and ecological relevance. In: Chincholkar S, Thomashow L (eds), *Microbial Phenazines: Biosynthesis, Agriculture and Health*. Berlin, Heidelberg: Springer, 2013, 199–216.
Thulasiram HV, Erickson HK, Poulter CD. Chimeras of two isoprenoid synthases catalyze all four coupling reactions in isoprenoid biosynthesis. *Science* 2007;**316**:73–76.
Ueda D, Yamaga H, Murakami M *et al.* Biosynthesis of sesterterpenes, head-to-tail triterpenes, and sesquarterpenes in *Bacillus clausii*: identification of multifunctional enzymes and analysis of isoprenoid metabolites. *ChemBioChem* 2015;**16**:1371–7.
van Meer G, Voelker D, Feigenson G. Membrane lipids: where they are and how they behave. *Nat Rev Mol Cell Biol* 2008;**9**:112–24.
Villanueva L, Schouten S, Damsté JSS. Phylogenomic analysis of lipid biosynthetic genes of Archaea shed light on the 'lipid divide'. *Environ Microbiol* 2017;**19**:54–69.
Villanueva L, von Meijenfeldt FAB, Westbye AB *et al.* Bridging the membrane lipid divide: bacteria of the FCB group superphylum have the potential to synthesize archaeal ether lipids. *ISME J* 2021;**15**:168–82.
Vo NNQ, Nomura Y, Kinugasa K *et al.* Identification and characterization of bifunctional drimenol synthases of marine bacterial origin. *ACS Chem Biol* 2022;**17**:1226–38.
Wallrapp FH, Pan J-J, Ramamoorthy G,et al. Prediction of function for the polyprenyl transferase subgroup in the isoprenoid synthase superfamily. *Proc Natl Acad Sci USA* 2013;**110**:E1196–202.
Wang H, Fewer David P, Holm L, Rouhiainen L *et al.* Atlas of nonribosomal peptide and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes. *Proc Natl Acad Sci USA* 2014;**111**:9259–64.
Wang K, Sybers D, Maklad HR *et al.* A TetR-family transcription factor regulates fatty acid metabolism in the archaeal model organism *Sulfolobus acidocaldarius*. *Nat Commun* 2019;**10**:1542.
Wang P-H, Khusnutdinova AN, Luo F *et al.* Biosynthesis and activity of prenylated FMN cofactors. *Cell Chem Biol* 2018;**25**:560–70.e566.
Ward LM, Cardona T, Holland-Moritz H. Evolutionary implications of anoxygenic phototrophy in the bacterial phylum candidatus eremiobacterota (WPS-2). *Front Microbiol* 2019;**10**.
Weete JD. Structure and function of sterols in fungi. Paoletti R, Kritchevsky D, (eds). *Adv Lipid Res* 1989;**23**:115–67.
Wei G, Eberl F, Chen X *et al.* Evolution of isoprenyl diphosphate synthase-like terpene synthases in fungi. *Sci Rep* 2020;**10**:14944.
Wei JH, Yin X, Welander PV. Sterol synthesis in diverse bacteria. *Front Microbiol* 2016;**7**:990.
Welander PV. Deciphering the evolutionary history of microbial cyclic triterpenoids. *Free Radical Biol Med* 2019;**140**:270–8.
Wendt KU, Poralla K, Schulz GE. Structure and function of a squalene cyclase. *Science* 1997;**277**:1811–5.
Wenski SL, Thiengmag S, Helfrich EJN. Complex peptide natural products: biosynthetic principles, challenges and opportunities for pathway engineering. *Synth Syst Biotechnol* 2022;**7**: 631–47.
White MD, Payne KAP, Fisher K,et al. UbiX is a flavin prenyltransferase required for bacterial ubiquinone biosynthesis. *Nature* 2015;**522**:502–6.
Whitfield C, Williams DM, Kelly SD. Lipopolysaccharide O-antigens—bacterial glycans made to measure. *J Biol Chem* 2020;**295**:10593–609.
Winkelblech J, Fan A, Li S-M. Prenyltransferases as key enzymes in primary and secondary metabolism. *Appl Microbiol Biotechnol* 2015;**99**:7379–97.
Winter JM, Moffitt MC, Zazopoulos E *et al.* Molecular basis for chloronium-mediated meroterpene cyclization – cloning, sequencing, and heterologous expression of the napyradiomycin biosynthetic gene cluster. *J Biol Chem* 2007;**282**:16362–8.
Xu B, Li Z, Alsup TA *et al.* Bacterial Diterpene Synthases Prenylate Small Molecules. *ACS Catal* 2021;**11**:5906–15.
Xu R, Fazio G, Matsuda S. On the origins of triterpenoid skeletal diversity. *Phytochemistry* 2004;**65**:261–91.
Xu Z, Baunach M, Ding L *et al.* Bacterial synthesis of diverse indole terpene alkaloids by an unparalleled cyclization sequence. *Angew Chem Int Ed* 2012;**51**:10293–7.
Yamada Y, Kuzuyama T, Komatsu M *et al.* Terpene synthases are widely distributed in bacteria. *Proc Natl Acad Sci USA* 2015;**112**:857–62.
Yang Y-l, Zhang S, Ma K,et al. Discovery and characterization of a new family of diterpene cyclases in bacteria and fungi. *Angew Chem Int Ed* 2017;**56**:4749–52.
Yang Y, Yatsunami R, Ando A *et al.* Complete biosynthetic pathway of the C_{50} carotenoid bacterioruberin from Lycopene in the extremely halophilic archaeon *Haloarcula japonica*. *J Bacteriol* 2015;**197**:1614–23.

Yang Y, Zhang Y, Zhang S *et al.* Identification and characterization of a membrane-bound sesterterpene cyclase from *Streptomyces somaliensis*. *J Nat Prod* 2018;**81**:1089–92.

Yu H, Schut GJ, Haja DK *et al.* Evolution of complex I-like respiratory complexes. *J Biol Chem* 2021;**296**:100740.

Zeng L, Dehesh K. The eukaryotic MEP-pathway genes are evolutionarily conserved and originated from Chlaymidia and cyanobacteria. *Bmc Genomics* 2021;**22**:137.

Zeng Z, Chen H, Yang H *et al.* Identification of a protein responsible for the synthesis of archaeal membrane-spanning GDGT lipids. *Nat Commun* 2022;**13**:1545.

Zeng Z, Liu X-L, Farley Kristen R *et al.* GDGT cyclization proteins identify the dominant archaeal sources of tetraether lipids in the ocean. *Proc Natl Acad Sci USA* 2019;**116**: 22505–11.

Zeyhle P, Bauer JS, Steimle M *et al.* A membrane-bound prenyltransferase catalyzes the O-prenylation of 1,6-dihydroxyphenazine in the marine bacterium *Streptomyces sp*. CNQ-509. *ChemBioChem* 2014;**15**:2385–92.

Zhang L, Chen C-C, Ko T-P,et al. Moenomycin biosynthesis: structure and mechanism of action of the prenyltransferase MoeN5. *Angew Chem Int Ed* 2016;**55**:4716–20.

Zhao L, Chang W-c, Xiao Y *et al.* Methylerythritol phosphate pathway of isoprenoid biosynthesis. *Annu Rev Biochem* 2013;**82**: 497–530.

Zheng Y, Cong Y, Schmidt EW *et al.* Catalysts for the enzymatic lipidation of peptides. *Acc Chem Res* 2022;**55**:1313–23.

Zhi X-Y, Yao J-C, Tang S-K *et al.* The futalosine pathway played an important role in menaquinone biosynthesis during early prokaryote evolution. *Genome Biol Evol* 2014;**6**:149–60.

Received 9 November 2022; **revised** 13 February 2023; **accepted** 28 February 2023