

Origin and Evolution of Polycyclic Triterpene Synthesis

Carlos Santana-Molina ^{*},¹ Elena Rivas-Marin ¹, Ana M. Rojas ^{*},¹ and Damien P. Devos ^{*},¹

¹Centro Andaluz de Biología del Desarrollo (CABD)-CSIC, Junta de Andalucía, Universidad Pablo de Olavide, Seville, Spain

*Corresponding authors: E-mails: csantmol@gmail.com; a.rojas.m@csic.es; damienpdevos@gmail.com.

Associate editor: Fabia Ursula Battistuzzi

Abstract

Polycyclic triterpenes are members of the terpene family produced by the cyclization of squalene. The most representative polycyclic triterpenes are hopanoids and sterols, the former are mostly found in bacteria, whereas the latter are largely limited to eukaryotes, albeit with a growing number of bacterial exceptions. Given their important role and omnipresence in most eukaryotes, contrasting with their scant representation in bacteria, sterol biosynthesis was long thought to be a eukaryotic innovation. Thus, their presence in some bacteria was deemed to be the result of lateral gene transfer from eukaryotes. Elucidating the origin and evolution of the polycyclic triterpene synthetic pathways is important to understand the role of these compounds in eukaryogenesis and their geobiological value as biomarkers in fossil records. Here, we have revisited the phylogenies of the main enzymes involved in triterpene synthesis, performing gene neighborhood analysis and phylogenetic profiling. Squalene can be biosynthesized by two different pathways containing the HpnCDE or Sqs proteins. Our results suggest that the HpnCDE enzymes are derived from carotenoid biosynthesis ones and that they assembled in an ancestral squalene pathway in bacteria, while remaining metabolically versatile. Conversely, the Sqs enzyme is prone to be involved in lateral gene transfer, and its emergence is possibly related to the specialization of squalene biosynthesis. The biosynthesis of hopanoids seems to be ancestral in the Bacteria domain. Moreover, no triterpene cyclases are found in Archaea, invoking a potential scenario in which eukaryotic genes for sterol biosynthesis assembled from ancestral bacterial contributions in early eukaryotic lineages.

Key words: polycyclic triterpenes, squalene, carotenoids, lateral gene transfer, eukaryogenesis.

Introduction

Sterol lipids are polycyclic triterpenes predominantly found in the membranes of eukaryotic cells, where they influence fundamental properties of the membrane like its fluidity and permeability, as well as cell signaling (Sezgin et al. 2017). Given their importance, and their omnipresence in eukaryotes and their general absence in prokaryotes (with a few exceptions), sterol biosynthesis has historically been considered a feature that emerged during the development of the eukaryotic cell, eukaryogenesis (Cavaliere-Smith 2002). However, the origin of sterols remains an important unresolved issue in the field of evolutionary cell biology and the growing presence of sterol biosynthesis in bacteria is questioning the status of sterol as a eukaryotic landmark (Wei et al. 2016; Lee et al. 2018).

To address the origin of sterol, it is important to consider the representative polycyclic triterpenes in bacteria, the hopanoids. It has been shown that sterols and hopanoids can form a liquid-ordered phase in synthetic membranes, preferentially interacting with sphingolipids and lipid A, respectively, reflecting certain functional similarity (Sáenz et al. 2012, 2015). One aspect that has previously been given little consideration is the relationship that sterols and hopanoids hold with carotenoids, which are nonpolycyclic terpenes produced by eukaryotes, archaea, and bacteria, representing widespread, multifunctional and biotechnologically important isoprenoid derivatives (Rodríguez-Concepción et al. 2018). In

addition to their antioxidant activity, carotenoids also fulfill physiological functions similar to those of polycyclic triterpenes, such as the modulation of membrane fluidity and proton permeability (Gruszecki and Strzałka 2005; Kupisz et al. 2008; Grudziński et al. 2017). Both types of terpenoids, polycyclic and nonpolycyclic ones, are related biosynthetically (fig. 1).

Squalene (C₃₀) is the precursor of hopanoids and sterols, and its biosynthesis has traditionally been assigned to the action of Squalene synthase (Sqs), an enzyme present in all eukaryotes as well as in some archaea and bacteria. Sqs produces squalene in two steps, condensing two farnesyl-PP (C₁₅) molecules into presqualene diphosphate (PSPP), which is then reduced to squalene using NADPH (Nakashima et al. 1995). An alternative pathway was recently described in bacteria, based on the cooperation of three enzymes: HpnD, HpnC, and HpnE (Pan et al. 2015). First, HpnD condenses two farnesyl-PP molecules into PSPP and then, HpnC converts PSPP to hydroxysqualene (HSQ), which is then reduced to squalene by HpnE using FADH₂ (Pan et al. 2015). HpnC, HpnD, and Sqs are homologous isoprenoid-condensing enzymes that belong to the trans-isoprenyl diphosphate synthases head-to-head (Trans IPPS HH) family. Other members of the Trans IPPS HH family are involved in carotenoid precursor biosynthesis, such as CrtM enzyme that condenses two farnesyl-PP molecules (C₁₅) into diapophytoene (C₃₀), or CrtB that condenses two geranylgeranyl-PP (C₂₀) into phytoene (C₄₀). By contrast, HpnE is a member of the amino

© The Author(s) 2020. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Open Access

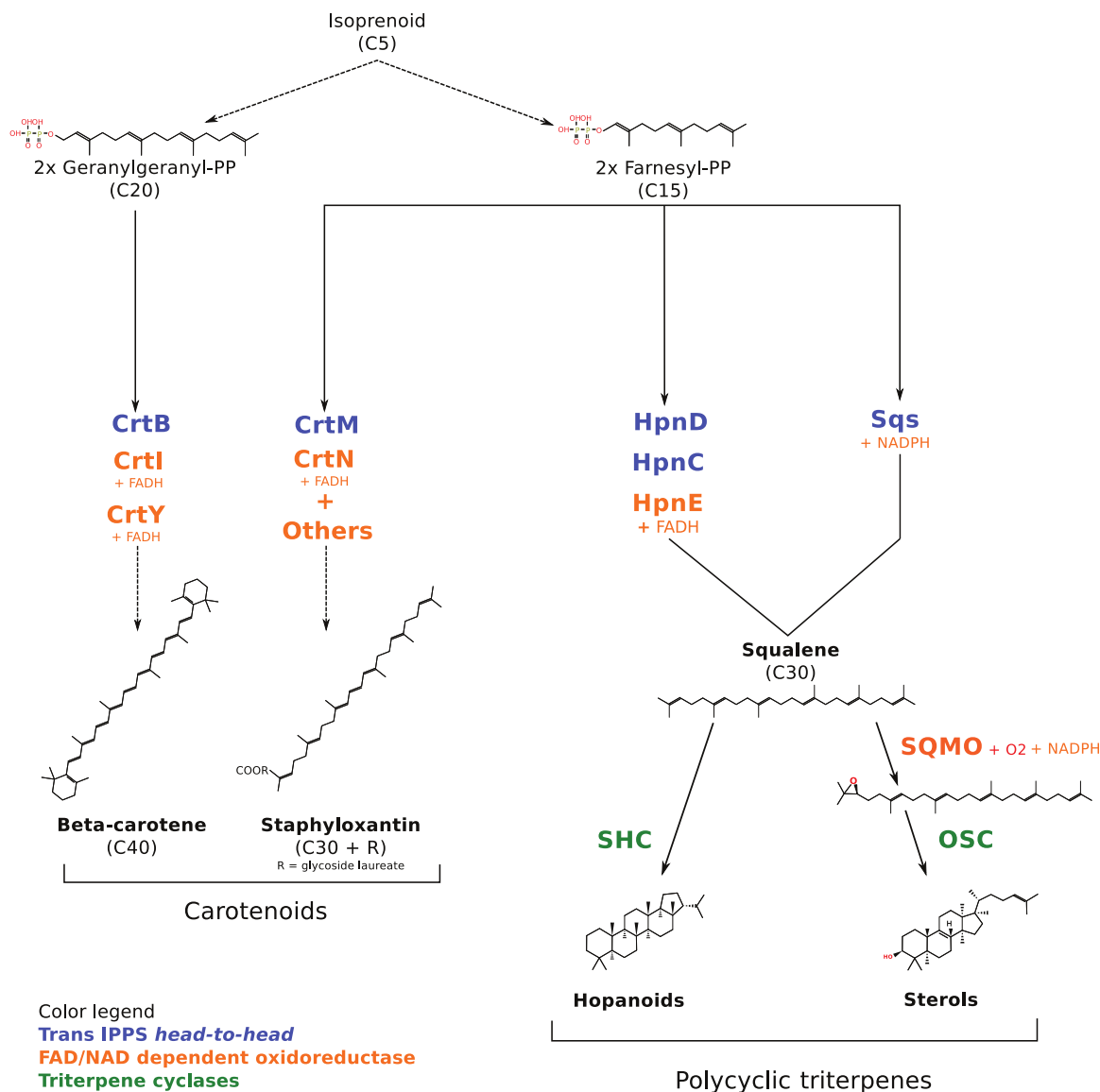


Fig. 1. Biosynthetic pathways for polycyclic triterpenes and carotenoids. Enzymes are colored according to their homology, with the exception of the FAD/NAD-dependent oxidoreductases where the CrtY and SQMO are not homologous to each other. Molecules are represented by their general structural scheme.

oxidoreductase family. CrtB and CrtM cooperate with other amino FAD/NAD-dependent oxidoreductases like CrtN, CrtD, and CrtI, homologs of HpnE, which convert these precursors into carotenoids like staphyloxanthin or β -carotene (Klassen 2010). Thus, the pathways of squalene and carotenoid biosynthesis are evolutionarily related. In hopanoid biosynthesis, squalene is directly cyclized into hopanoids by the squalene-hopene cyclase (SHC). For sterol biosynthesis, an additional step is required, whereby squalene is first converted into 2,3-epoxysqualene by the squalene monooxygenase (SQMO) and this intermediate is then cyclized into simplest sterols (like lanosterol, cycloartenol, or parkeol) by the oxidosqualene cyclase (OSC), homolog of SHC (fig. 1).

The linear and cyclic compounds have different physico-chemical characteristics (e.g., flexibility, rigidity, and stability) and thus might lead to different membrane properties. It has

been proposed that triterpene evolution has followed the laws of increasing complexity, suggesting that the linear compounds (e.g., carotenoids) preceded the polycyclic ones, such as hopanoids and sterols (Ourisson and Nakatani 1994). The increasing complexity of these molecules would have driven the appearance of more complex membrane characteristics (Ourisson and Nakatani 1994). Although the triterpene cyclase family have been previously studied (Fischer and Pearson 2007; Frickey and Kannenberg 2009), the origin of the squalene biosynthesis pathways (HpnCDE and Sqs) has not previously been considered. The existence of two pathways for squalene biosynthesis raises questions about the evolutionary relationship between HpnCDE, Sqs, and carotenoid synthases, and in particular, regarding the co-occurrence of both pathways in some genomes or the spread of triterpene cyclases concomitant with that of the squalene

biosynthesis pathways. These evolutionary conundrums must be resolved in order to fully understand how polycyclic triterpenes originated and evolved.

Given the oxygen requirement of SQMO, the origin of sterol must be linked to the appearance of this molecule on the Earth. By contrast, hopanoids can be produced in anoxic conditions, suggesting that SHC precedes OSC in evolutionary terms (Rohmer et al. 1979; Ourisson et al. 1987; Ourisson and Nakatani 1994). However, the differences in the molecular conformations of sterols and hopanoids have previously been interpreted as an indication that both kinds of enzyme evolved from a common ancestral intermediary (Fischer and Pearson 2007). Since the origin of sterols has been linked to the process of eukaryogenesis, their identification in some bacteria has historically been attributed to lateral gene transfer (LGT) from an ancestral eukaryote (Pearson et al. 2003; Desmond and Grimaldo 2009). However, the number of bacteria known to synthesize sterols is continuously increasing (Bird et al. 1971; Bode et al. 2003; Pearson et al. 2003; Desmond and Grimaldo 2009; Villanueva et al. 2014; Wei et al. 2016). The appearance of OSC/SQMO proteins is now considered to have occurred around 2.45 Ga, and the preferred interpretation is that the ancestors of these genes already existed in the eukaryotic stem lineage (Gold et al. 2017). Given the differences between bacterial and eukaryotic sterol structure (Banta et al. 2017) and production (Lee et al. 2018), it is reasonable to predict that bacterial sterols have distinct and unexplored functions that could provide some insight into the evolutionary pressure that prompted bacteria to acquire and retain a sterol biosynthesis pathway (Welander 2019). Hence, sterols appear to be more prominent in bacteria than previously thought, and more recent insights into the tree of life (Hug et al. 2016; Eme et al. 2017; Zaremba-Niedzwiedzka et al. 2017; Betts et al. 2018) suggest that now is an ideal time to reevaluate polycyclic triterpene evolution.

In contrast to earlier analyses, here we aimed to determine the origin of the most representative polycyclic triterpenes by reconstructing and comparing the phylogenies of the main enzymes involved in their synthesis, focusing particularly on members of the Trans IPPS HH family and triterpene cyclases. We discuss the evolution of squalene focused on HpnCDE and Sqs enzymes as well as the possible physiological roles provided by the different pathways in prokaryotes. On the other hand, the evolution of the triterpene cyclases displays complex evolution in Bacteria and Eukarya although its absence in Archaea is noteworthy. Together, this analysis supports a bacterial contribution to the origin of sterol biosynthesis genes, a phenomenon that has important implications for our models of eukaryogenesis.

Results and Discussion

The Squalene Biosynthesis Enzymes Evolved Following Two Distantly Related Routes

The phylogenetic reconstructions of the Trans IPPS HH define three main different classes (Sqs, CrtB/M, and HpnC/D; fig. 2). Two main issues are however to be considered. First, ancestral

rooting is problematic, and a long standing problem in phylogenetics (Gouy et al. 2015). We applied two widely used rooting methodologies (midpoint and branch-length estimation) that provided consistent topologies (fig. 2). Second, the order of the basal nodes can vary depending on many parameters (i.e., the evolutionary models or redundancy threshold, among others) affecting the topology of the tree as we observed in our analyses. In the case of the Trans IPPS HH, the large divergence time, the scarce signal of the informative regions with small size, and the relative speed of evolutionary rate amplified these aforementioned issues. Altogether, it is difficult to reconstruct the evolutionary relationships among these subfamilies. Hence, the order of the branches (specially at basal nodes) does not necessarily reflect the actual evolutionary trajectory. The composition of the main group of carotenoid synthases is consistent with that obtained in a prior analysis that only focused on enzymes responsible for carotenoid biosynthesis (Klassen 2010).

The vast majority of Trans IPPS HH enzymes are found in bacteria, although certain eukaryotic members of this family are related to endosymbiotic events. For example, CrtB in photosynthetic eukaryotes is of chloroplast origin and the NADH: Dehydrogenase ubiquinone complex I assembly factor 6 (not included in the tree) is of mitochondrial origin (supplementary text, [Supplementary Material online](#)). Other examples may derive from bacterial to eukaryotic LGT, such as CrtYB in Ascomycota (Sandmann 2002) that was later transferred to Aphididae (Metazoa: fig. 2; supplementary text, [Supplementary Material online](#)). Alternatively, archaeal CrtB is closely related to the bacterial type, although this relationship has yet to be strictly resolved (Klassen 2010).

The longest and best supported basal branch of the Trans IPPS HH tree is that which contains the Sqs subfamily, indicating that it is the most divergent class. Indeed, the Sqs subfamily seems to be characterized by an aspartic-rich region involved in Mg^{2+} binding, which differs from the other Trans IPPS HH subfamilies (DTxxD vs. DDxxD: [supplementary fig. S1, Supplementary Material online](#)). By contrast, HpnD and HpnC are phylogenetically closer to the CrtB and CrtM subfamilies. The genomic context of these different subfamilies of Trans IPPS HH shows that amino oxidase genes (like *hpnE*, *crtI*, or *crtN*) are usually found in the neighboring genes of *hpnD*, *hpnC*, and carotenoid synthases (both *crtB* and *crtM*), whereas they are infrequent in the neighboring genes of *sqs* (see [supplementary fig. S2, Supplementary Material online](#), for additional details). Thus, the phylogenetic proximity of HpnD, HpnC, and carotenoid synthases, and the more distant position of Sqs, is congruent within the genomic context. Protein searches of HpnE show that the closest homolog of these enzymes is the Z-carotene desaturase present in cyanobacteria and photosynthetic eukaryotes (see below), supporting the close relationship of HpnCDE enzymes with the ones involved in carotenoid production.

Thus, the enzymes involved in squalene biosynthesis appear to have evolved following two distantly related routes: the HpnCDE proteins closely related to the carotenoid biosynthesis and the divergent Sqs enzymes. Given the prior analyses of carotenoid biosynthesis enzymes (Klassen 2010),

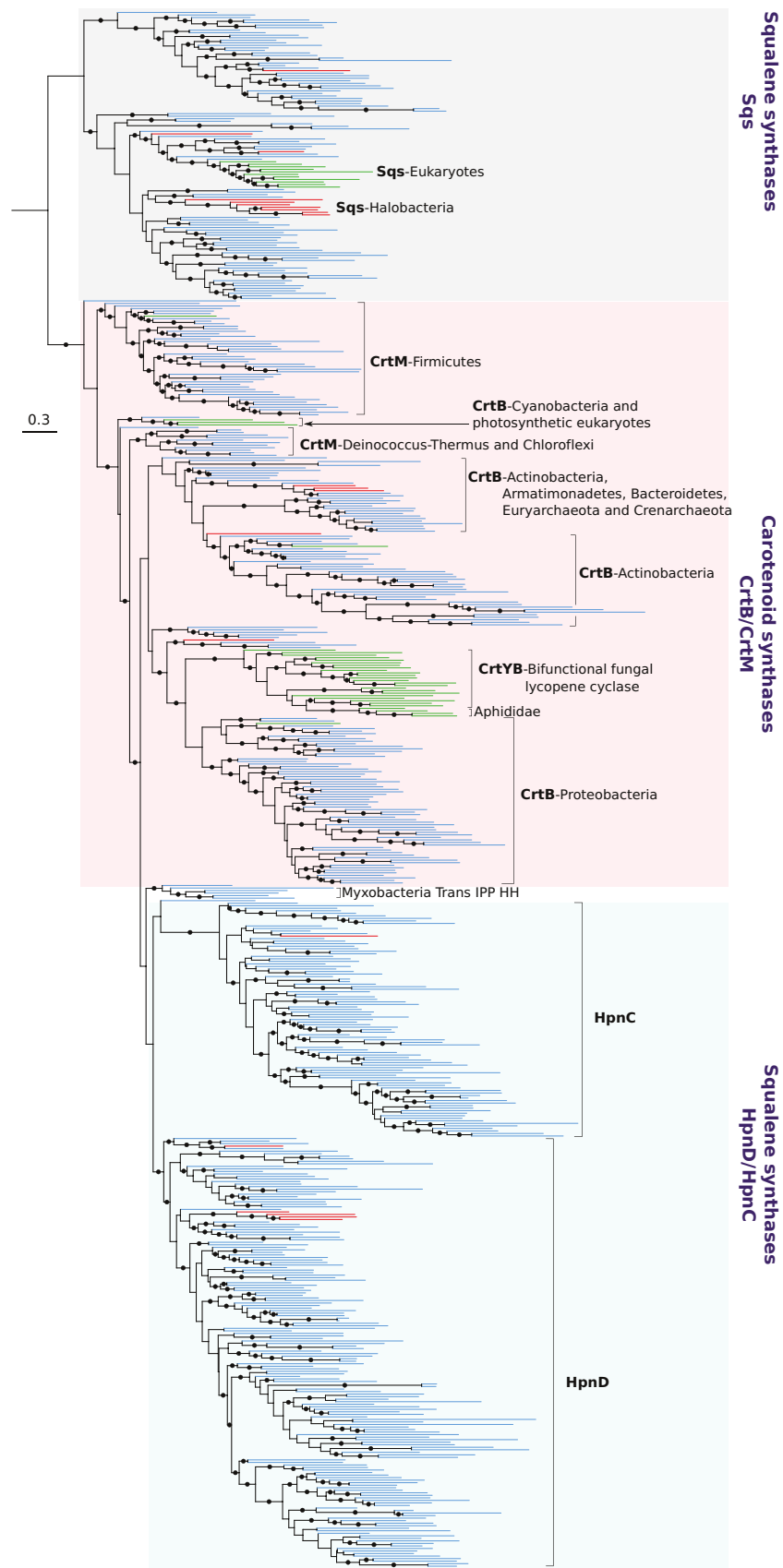


Fig. 2. ML reconstruction of the Trans IPPS HH family. The data set was reduced to 50% redundancy, obtaining a final alignment of 282 amino acid positions. The tree was calculated using IQ-TREE, with the LG + F + R10 model. Branches are colored according to the respective domain: blue, bacteria; red, archaea; and green, eukaryotes. The gray background represents the Sqs subfamily and the red background represents the diverse carotenoid synthase subfamilies, most of them identified previously (Klassen 2010). Bootstrap values higher than 90% are indicated.

we focused specifically on HpnD, HpnC, and HpnE, and the Sqs, subfamilies.

The hpnCDE Genes Are Likely Ancient and Form a Versatile Patchwork Pathway

The co-occurrence of HpnD, HpnC, and HpnE enzymes in prokaryotic genomes is the most common distribution in our data set (supplementary fig. S3, Supplementary Material online). In addition, the rooted topologies of the HpnD, HpnC, and HpnE reconstructions provided globally congruent topologies (supplementary fig. S4, Supplementary Material online), further supported by the concatenated reconstruction of the HpnCDE enzymes (supplementary fig. S5, Supplementary Material online). Therefore, the coevolution of these three enzymes might be inferred, suggesting that HpnCDE enzymes form a conserved biosynthetic patchwork in prokaryotes. Conversely, HpnD can often be found in the absence of the other enzymes, although the co-occurrence of HpnD and Sqs is also frequently observed (see below; supplementary fig. S3 and table S1, Supplementary Material online). Other kind of co-occurrence is infrequent (found in $< \sim 200$ genomes out of the $\sim 24,000$ analyzed) and they are usually restricted to specific taxa (supplementary table S1, Supplementary Material online). Thus, in order to evaluate the origin of the *hpnCDE* gene set, we took HpnD as the reference as it is the most conserved in prokaryotes (fig. 3A).

The HpnD reconstruction and those for HpnC and HpnE have a sufficiently strong phylogenetic signal as to be able to infer their evolutionary history (fig. 3A, supplementary fig. S4 and text, Supplementary Material online). These phylogenies mostly define monophyletic groups with the exception of some discrete or paraphyletic groups, suggesting possible gains of these genes through LGT (supplementary text, Supplementary Material online). The presence of the HpnCDE enzymes in archaea (fig. 3A) is explained by LGT as it is restricted to the Euryarchaeota, Thermoplasmata-Poseidoniia class (according to GTDB taxonomy, Parks et al. 2018, and including Euryarchaeota Marine Group II: supplementary table S2, Supplementary Material online), which has been previously proposed to have received substantial genomic contributions from bacteria (Deschamps et al. 2014; López-García et al. 2015). The topology of most of the paraphyletic or discrete HpnD groups is also replicated in the HpnC and HpnE reconstructions (supplementary fig. S4, Supplementary Material online), supporting the idea that these genes were transferred together and that they coevolved. Other exceptions could indicate the loss of a phylogenetic signal due to the short length or functional promiscuity of these enzymes (supplementary text, Supplementary Material online), which indeed could influence the bootstrap supports and artifactual clustering of major clades. Combining the distribution of HpnCDE within the individual phyla with the congruent monophyletic groups in the phylogenies (supplementary table S2 and fig. S4, Supplementary Material online), we infer that the HpnCDE enzymes were ancestral in phyla like Actinobacteria, Acidobacteria (despite the paraphyly), Planctomycetes, Verrucomicrobia, Elusimicrobia, *Candidatus* Division Rokubacteria, NC10,

Nitrospira, and Proteobacteria (particularly in Alpha-, Beta-Gamma-, and Zeta-proteobacteria; for more detailed discussion, see supplementary text and table S2, Supplementary Material online). Therefore, acknowledging the low support in some clades but considering that the topologies of HpnCDE display deep similarities with a species tree (Hug et al. 2016), it is likely that the HpnCDE enzymes originated before the diversification of bacterial lineages mainly from Gracilicutes and some phyla from Terrabacteria, indicating that they were ancestral in Bacteria. Importantly, the *hpnD*, *hpnC*, and *hpnE* genes likely form a transcriptomic unit in most of these phyla (even if one of them is absent in some of the genome as is the case in Nitrospira), suggesting conserved cooperativity of the enzymes (supplementary fig. S5, Supplementary Material online). Other cases, such as the genomes that only contain HpnD, suggest a possible functional versatility (supplementary text, Supplementary Material online).

The incompleteness of the *hpnCDE* gene set in certain genomes makes it difficult to draw clear inferences about the origin of the three enzymes as a functional pathway. We mapped the other squalene biosynthesis related enzymes (HpnC, HpnE, and Sqs) onto the HpnD phylogeny, illustrating the irregular co-occurrence of these proteins (fig. 3A, also for HpnC and HpnE in supplementary fig. S4, Supplementary Material online). Although this pattern of co-occurrence of HpnCDE and Sqs was previously described in Acidobacteria (Damsté et al. 2017), it can apparently be extended to other phyla like Verrucomicrobia, Nitrospira, *Ca.* Rokubacteria, NC10, Elusimicrobia, Chloroflexi, Beta-Gamma-proteobacteria, and the euryarchaeal *Ca.* Poseidoniia class (for which complete genomes of some organisms are available: fig. 3A). However, the HpnD of those organisms that lack the HpnC and/or HpnE, branch with their relatives, as expected, and similar behavior is also observed for HpnC or HpnE (supplementary fig. S4, Supplementary Material online). When combined with the syntenic observations, these data suggest that the irregular distribution of *hpnCDE* gene set is mostly due to gene loss. Another possible explanation is that these enzymes have been transferred alone. Alternatively, the incompleteness of the HpnCDE pathway in some groups may be linked to the presence of Sqs. Indeed, the Sqs associated with the degenerated HpnCDE gene set usually cluster in the same branch of the phylogenetic reconstruction of Sqs (fig. 3B). No Sqs were found in the vicinity of HpnD (supplementary figs. S4 and S5, Supplementary Material online) suggesting that the acquisition of HpnD and Sqs is probably independent. Thus, given the topology of the monophyletic groups of HpnCDE reconstructions, it is most likely that the *hpnCDE* genes were already present in these genomes and that *sqs* were the mobile element (i.e., the result of LGT), possibly functionally displacing the HpnCDE through gene loss. These results could also illustrate a common evolutionary history for these events (the loss of *hpnCDE* genes associated with the presence of *Sqs* in the genomes), possibly related to the joined presence of these organisms in a specific environmental niche.

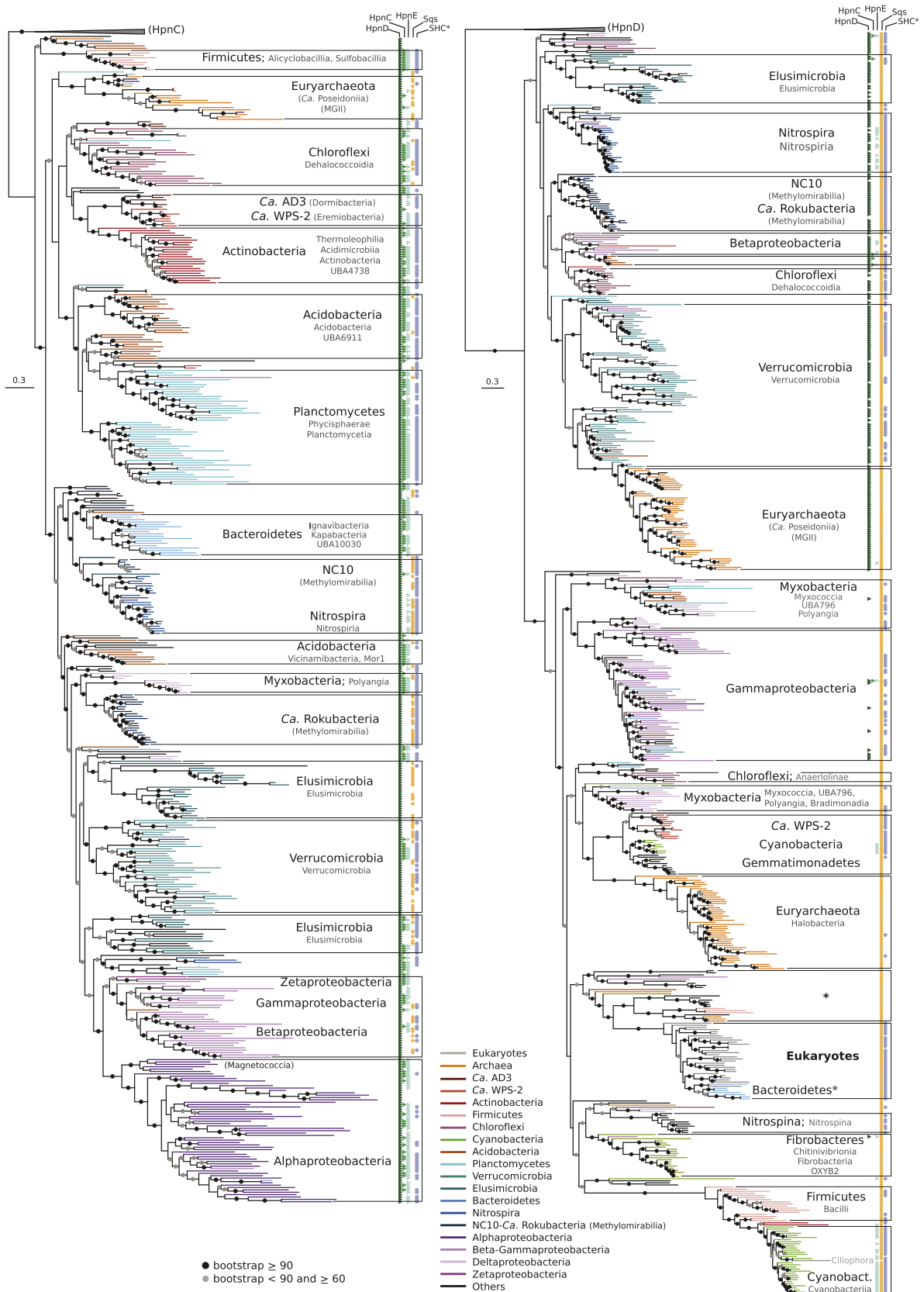


Fig. 3. Phylogeny and co-occurrence of the two pathways of squalene synthetic genes mapped to the phylogenetic trees. The data set was reduced by progressive redundancy of the phylum and branches, colored according to the prokaryotic phyla. Smaller fonts indicate taxonomical classes, and those between brackets the taxonomical classes designated by GTDB. (A) ML reconstruction of HpnD rooted to the HpnC subfamily. The final alignment contained 268 amino acid positions and the tree was calculated by IQ-TREE, with the LG + F + R10 model. (B) ML reconstruction of Sqs rooted to the HpnD subfamily. The final alignment contained 313 amino acid positions, and the tree was calculated using IQ-TREE with the LG + F + R9

Therefore, the phylogenetic reconstructions, together with the distribution of these genes and their synteny, suggest that *hpnCDE* genes arose and potentially assembled as a functional pathway early in bacterial evolution. On the other hand, the HpnD, HpnC, and HpnE enzymes might act independently as they do not always co-occur. Alterations of the HpnCDE pathway (i.e., gene losses) suggest recurrent deviations of HpnCDE enzymes into alternative roles, possibly involving terpenoid metabolism and even carotenoid synthesis. The fact that the HpnD and HpnC subfamilies cluster together phylogenetically (fig. 2) adopt a common arrangement in the genome, and their phylogenetic reconstructions are congruent (supplementary figs. S4 and S5, Supplementary Material online), suggests that both enzymes are derived from an ancestral duplication followed by specialization. HpnD was probably the ancestral form as it mediates the initial and crucial step (condensing two farnesyl-PP), and it is more conserved than HpnC.

Sqs Is of Bacterial Origin and Has Been Transferred Multiple Times

In contrast to the HpnCDE enzymes that are only present in Bacteria and specific Euryarchaeota, Sqs sequences can be found in the three domains of life. The phylogeny of Sqs establishes diverse paraphyletic groups of bacteria and archaea, suggesting multiple LGT events in prokaryotes (fig. 3B), as opposed to the apparently single origin in eukaryotes. Hence, and concomitant with sterol biosynthesis (Gold et al. 2017), Sqs were probably already present in the Last Eukaryotic Common Ancestor. Notably, and with only one eukaryotic exception, the Sqs of some members of the Ciliophora branch within the group of Cyanobacteria/Firmicute Sqs (fig. 3B), both belonging to the Terrabacteria supergroup (Battistuzzi and Hedges 2009). This result sheds light on the origin of tetrahymanol biosynthesis in Ciliophora, to date considered to be a late bacterial contribution through LGT (Frickey and Kannenberg 2009; Takishita et al. 2012). In fact, the eukaryotic tetrahymanol cyclases also branch close to the Firmicute SHCs (Terrabacteria: fig. 4) and thus, the source of the Ciliophora Sqs and tetrahymanol-cyclase genes is related to the Terrabacteria group.

The presence of Sqs in Archaea is restricted to Euryarchaeota–Halobacteria class, and the previously indicated euryarchaeal *Ca. Poseidonii* class (including Euryarchaeota Marine Group II) and these sequences forming three distinct paraphyletic groups: one for Halobacteria and two for *Poseidonii* (fig. 3B). Thus, this discrete taxonomic distribution combined with the phylogeny suggest that the three groups of euryarchaeal Sqs appear to have originated from bacteria through independent LGT events, in accordance with previous observations for these specific euryarchaeota groups (López-García et al. 2015: a more detailed

discussion is offered in the supplementary text, Supplementary Material online).

Sqs are found in diverse bacterial groups, mostly showing a monophyletic pattern (fig. 3B). However, the mixed topology of the different taxonomic groups suggests that Sqs have been transferred multiple times between bacteria, as suggested previously for those genomes that also contain *hpnD* (see above). Likewise, the Sqs in various paraphyletic Proteobacteria groups are found in different genomic contexts, supporting the idea of various LGT events in Proteobacteria (particularly for Beta-proteobacteria, Gamma-proteobacteria, and Myxobacteria; supplementary fig. S6 and text, Supplementary Material online). Remarkably, Cyanobacteria and Firmicutes-Bacilli (such as Nitrospina or Fibrobacteres) only have Sqs but never HpnCDE (fig. 3B), suggesting a switch in squalene biosynthesis. Indeed, the Sqs from Cyanobacteria and Firmicutes are very divergent from the rest of the Sqs (supplementary fig. S1B and text, Supplementary Material online).

Thus, a eukaryotic origin of Sqs is unlikely because of the widespread distribution in bacteria, and the potential to be ancestral in some phyla such as Cyanobacteria. In addition, we infer that the presence of Sqs in Archaea is the result of LGT events. There is no support here for the vertical inheritance of Sqs between Archaea and Eukarya, and in fact, the “current” absence of Sqs homologs in Asgard argues against this phenomenon. Therefore, our data suggest that it is most likely the Sqs enzymes have a bacterial origin, and that their presence in Eukarya and Archaea is best explained by distinct LGT events.

The Enzymes Involved in Hopanoid and Sterol Biosynthesis Have Different Evolutionary Histories

Triterpene cyclases (SHC and OSC) are considered the evolutionary footprints of polycyclic triterpenes because they carry out the key cyclization of the precursor into the simplest polycyclic triterpene. Our phylogenetic results on polycyclic triterpene cyclases define three main groups: SHCs, OSCs, and an additional group of divergent SHCs (hereafter SHC-like: fig. 4 and extended in supplementary fig. S7, Supplementary Material online). The eukaryotic SHC-like sequences display a signal that is most coherent with LGT events from bacteria, as suggested previously (Frickey and Kannenberg 2009; Takishita et al. 2012; Li et al. 2018; Jia et al. 2019: supplementary text, Supplementary Material online). Eukaryotic OSCs have a single origin, whereas the various bacterial OSCs form different paraphyletic groups external and internal to eukaryotes (see below). Thus, polycyclic triterpenes are present in Eukaryotes and Bacteria but absent from Archaea, possibly reflecting the distinct compatibilities of polycyclic triterpenes with the chemistry of the bacteria–eukaryote and archaea membranes

FIG. 3. Continued

model. The * at the basal branches of the eukaryotic group indicates diverse prokaryotes with few members: Brachyspirae, Halanaerobiales, Zixibacteria, *Ca. Poseidonii*, Delongbacteria, and Bdellovibrio. The phylogenetic profiles of HpnD, HpnC, HpnE, Sqs, and SHC* are represented to the right of each tree. SHC* includes SHC, OSC, or similar. Bootstrap values higher than 90% and 60% are indicated by black and gray circles, respectively.

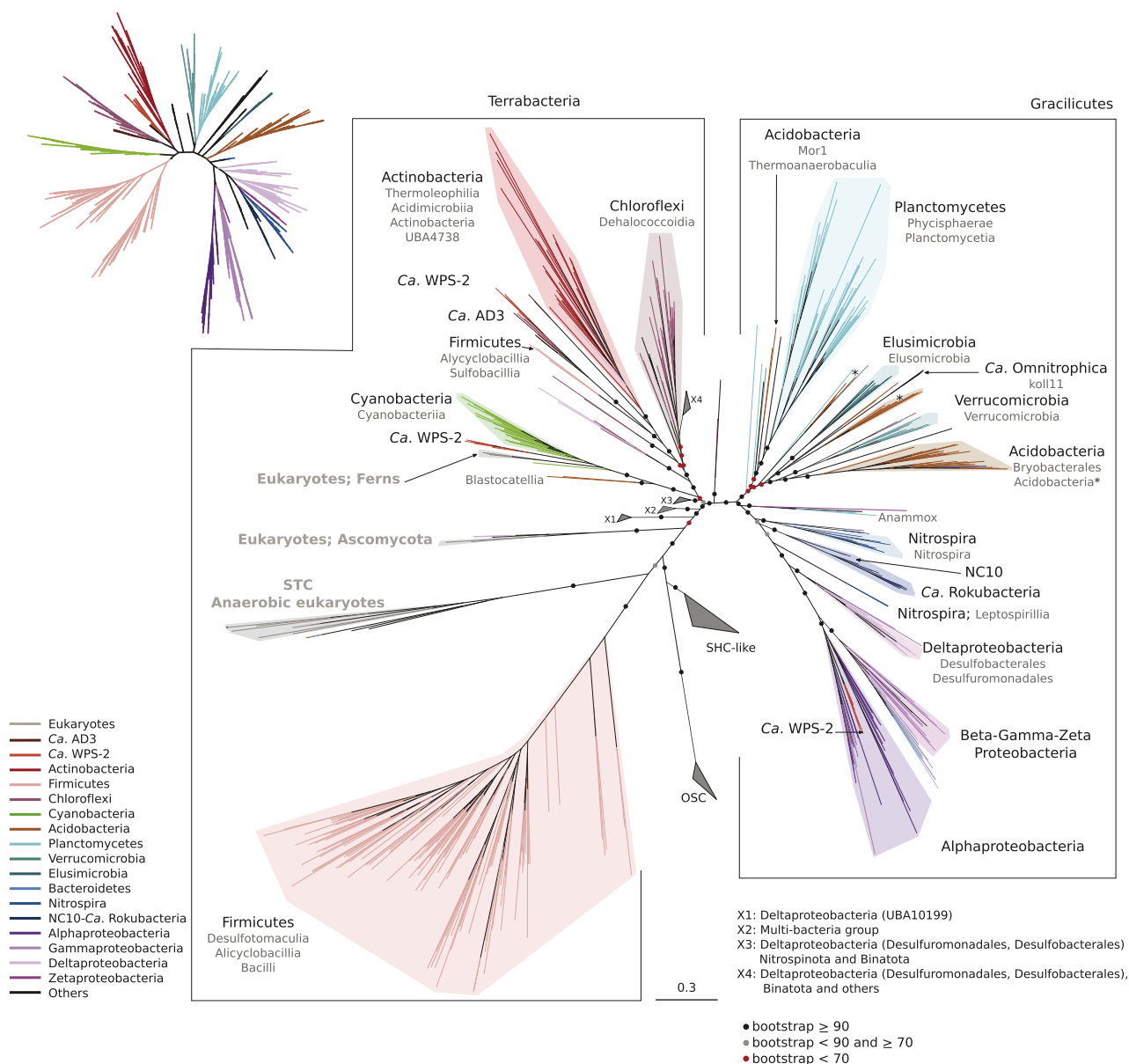


Fig. 4. ML reconstruction of the triterpene cyclase family to show the evolution of the closest homologs to SHC. Eukaryotic squalene-tetrahymanol cyclase is designated as STC. Progressive redundancy was used to reduce the data set by phylum, and the branches are colored according to the prokaryotic phyla. The final alignment contained 726 amino acid positions and the tree was calculated using IQ-TREE with the LG + F + R10 model. Only those bootstraps at basal nodes and at nodes defining main taxa are shown. Some branches have been collapsed and the rest of bootstraps are omitted to favor visualization, the full tree with all bootstraps indicated is available in [supplementary figure S7, Supplementary Material online](#). The inset tree indicates the topology of the species tree of the respective organisms bearing a triterpene cyclase to show the major congruent taxa with SHC phylogeny. This tree is provided GTDB database ([Parks et al. 2018](#)) and the branch-lengths of the inset tree were ignored to ease the visualization.

(G3P vs. G1P, or phospholipid vs. isoprenoids, among other features; [Peretó et al. 2004](#)).

Hopanoids Are Ancestral in Bacteria and They Followed a Heterogeneous Evolution

We selected the SHC/SHC-like and OSC branches in order to analyze each family independently. The closest homologs to SHC are present in 31 of the 130 phyla, with a similar taxonomical distribution to HpnCDE or Sqs that encompasses Proteobacteria, Nitrospira, Nitrospina, *Ca. Rokubacteria*, NC10, Planctomycetes, Verrucomicrobia, Elusimicrobia,

Acidobacteria, Cyanobacteria, *Candidatus* WPS-2 (designated as *Eremiobacteria* by GTDB), *Candidatus* AD3 (Dormibacteria), Actinobacteria, Chloroflexi, and Firmicutes ([fig. 4](#) and [supplementary table S2, Supplementary Material online](#)). The topology of the reconstruction of SHC homologs recovers the two major groups of bacteria, Gracilicutes and Terrabacteria (SHC was not found in Candidate Phyla Radiation bacteria: [fig. 4](#)). The phyla in these supergroups are mostly monophyletic, albeit with some exceptions, like Deltaproteobacteria (including Desulfobacterales and Desulfuromonadales) and *Ca. WPS-2*, whose sequences are

situated distally in the tree that is perhaps indicative of possible LGT events. Similarly, the sequences from Acidobacteria are paraphyletic yet closely related, suggesting the possible loss of the phylogenetic signal (a similar behavior was observed in HpnCD reconstructions). The topology of the major monophyletic groups in the SHC phylogeny was similar to the reconstruction of the phylogenetic marker RpoB (supplementary fig. S8A and B, Supplementary Material online), which is similar to the most accepted reconstruction of the tree of life (Hug et al. 2016; Parks et al. 2018: fig. 4 inset tree as a reference species tree provided by GTDB).

We identified divergent triterpene cyclase groups. The bifunctional triterpene/sesquiterpene cyclase (SqhC, also known as tetraprenyl- β -curcumene cyclase: Sato et al. 2011a, 2011b) was restricted to Firmicutes-Bacilli, with some exceptions in other classes, and seemed to show higher evolutionary rates than other triterpene cyclases or other house-keeping genes like *rpoB* (supplementary fig. S8C and text, Supplementary Material online). Other divergent triterpene cyclases, such as the SHC-like proteins, are apparently formed by different sequences mainly from Myxobacteria, Actinobacteria, and Planctomycetes, most likely representing artifactual clustering (supplementary figs. S9 and S10 and text, Supplementary Material online). The SHC-like proteins from Planctomycetes are the only ones that conserve critical catalytic residues and their genomic context suggest that they probably use different precursors to squalene (supplementary fig. S10 and text, Supplementary Material online). Thus, it is likely that the SHC-like enzymes from Planctomycetes produce polycyclic triterpenes other than hopanoids or sterols. Identifying these novel polycyclic triterpenes and establishing their situation within the geological record might help attribute specific geological fossils to specific taxonomic groups.

The global congruency of the SHC phylogeny with a species tree suggests a vertical evolution of SHC in the bacterial domain, with subsequent divergence (e.g., the SqhC in Firmicutes). Vertical evolution is plausible for the Gracilicutes supergroup, given that the respective monophyletic groups contain aerobic and anaerobic bacteria. However, it is important to note that SHC is nearly absent in the FCB group including Bacteroidetes and Fibrobacteres among others. On the other hand, the same assumption of vertical evolution cannot be made for the Terrabacteria supergroup due to the discrete distribution of SHC, with some of them only formed by aerobes like Cyanobacteria or Firmicutes-Bacilli. In fact, SHC is absent in early-branching anaerobic groups of Terrabacteria, such as Clostridia, Negativicutes, Melaniobacteria, and Saganbacteria, which makes it difficult to envisage vertical inheritance of SHC in Terrabacteria. Therefore, the current data do not provide enough resolution to infer the presence of SHC in a common bacterial ancestor, although the results do indicate that the biosynthesis of hopanoid (defined by the SHC enzyme) precedes the diversification of the whole Gracilicutes group and some of the Terrabacteria taxa.

The functional domains of the genes located in the vicinity of the cyclase genes are usually conserved (supplementary fig. S9, Supplementary Material online). This genomic region

includes genes related with hopanoid production in phyla like Actinobacteria or Proteobacteria, with some variation. However, this specific association is not found in other groups like Firmicutes or Cyanobacteria (supplementary fig. S9, Supplementary Material online). Together, these data suggest that SHC has evolved heterogeneously among the bacterial phyla, probably due to the different roles of polycyclic triterpenes in different taxa and niches (Bosak et al. 2008; Belin et al. 2018). This might explain why hopanoids are difficult to attribute to the advent of a specific phylogenetic group, metabolic process, or environmental context.

Bacterial Contribution to the Presence of Sterol Biosynthesis in Eukarya

Due to their evolutionary relevance, the phylogenies of bacterial OSC and SQMO have been analyzed previously (Pearson et al. 2003; Desmond and Gribaldo 2009; Wei et al. 2016; Banta et al. 2017; Gold et al. 2017), although improvements in genomic sampling have led to changes in the evolutionary interpretations of these phylogenies (Wei et al. 2016). Bacterial sterol biosynthesis has historically been restricted to discrete taxa, but it has been extended in recent years, and we reveal the presence of previously unidentified genes involved in sterol biosynthesis in taxa like *Ca. Rokubacteria* and some Actinobacteria orders (Corynebacteriales, Streptomycetales, and Streptosporangiales).

Our topologies of SQMO and OSC reconstructions are globally congruent (fig. 5), suggesting a similar evolutionary history that is supported by the proximity of the *sqmo* and *osc* genes in most bacterial genomes (fig. 5, central panel). The genes present in the vicinity of *osc* are also involved in sterol metabolism, such as some *erg* genes that encode proteins involved downstream in the sterol pathway in eukaryotes (particularly ERG4/ERG24, ERG2, and ERG11), or the Riskie and NAD(P)-binding Rossmann-fold domains that were both recently shown to be involved in C4 sterol demethylation in methylotrophic bacteria (Lee et al. 2018). This is mainly observed in some Proteobacteria, Cyanobacteria, Actinobacteria, and *Ca. Rokubacteria* genomes, illustrating the specialization of the bacterial loci involved in sterol synthesis. Our reconstruction defines four OSC groups (OSC1–4), among which lanosterol synthases (present in groups OSC1 and 4), cycloartenol synthases (OSC2–4), parkeol synthases (OSC1), and arborane synthases (OSC2) have been identified (Bird et al. 1971; Pearson et al. 2003; Wei et al. 2016; Banta et al. 2017). OSC1 forms a very stable basal group of lanosterol/parkeol synthases, with conserved residues at the active site positions (supplementary fig. S11, Supplementary Material online). However, OSC2 contains more diverse sequences, among which we find the arborane synthase from *Eudoraera adriatica* (Supplementary Material online: Banta et al. 2017), representing a less stable phylogenetic group. OSC3 and OSC4 also form stable groups that always branch basally to and within the eukaryotes, respectively.

Two previous hypotheses were postulated to explain the phylogenetic patterns of the bacterial enzymes (Pearson et al.

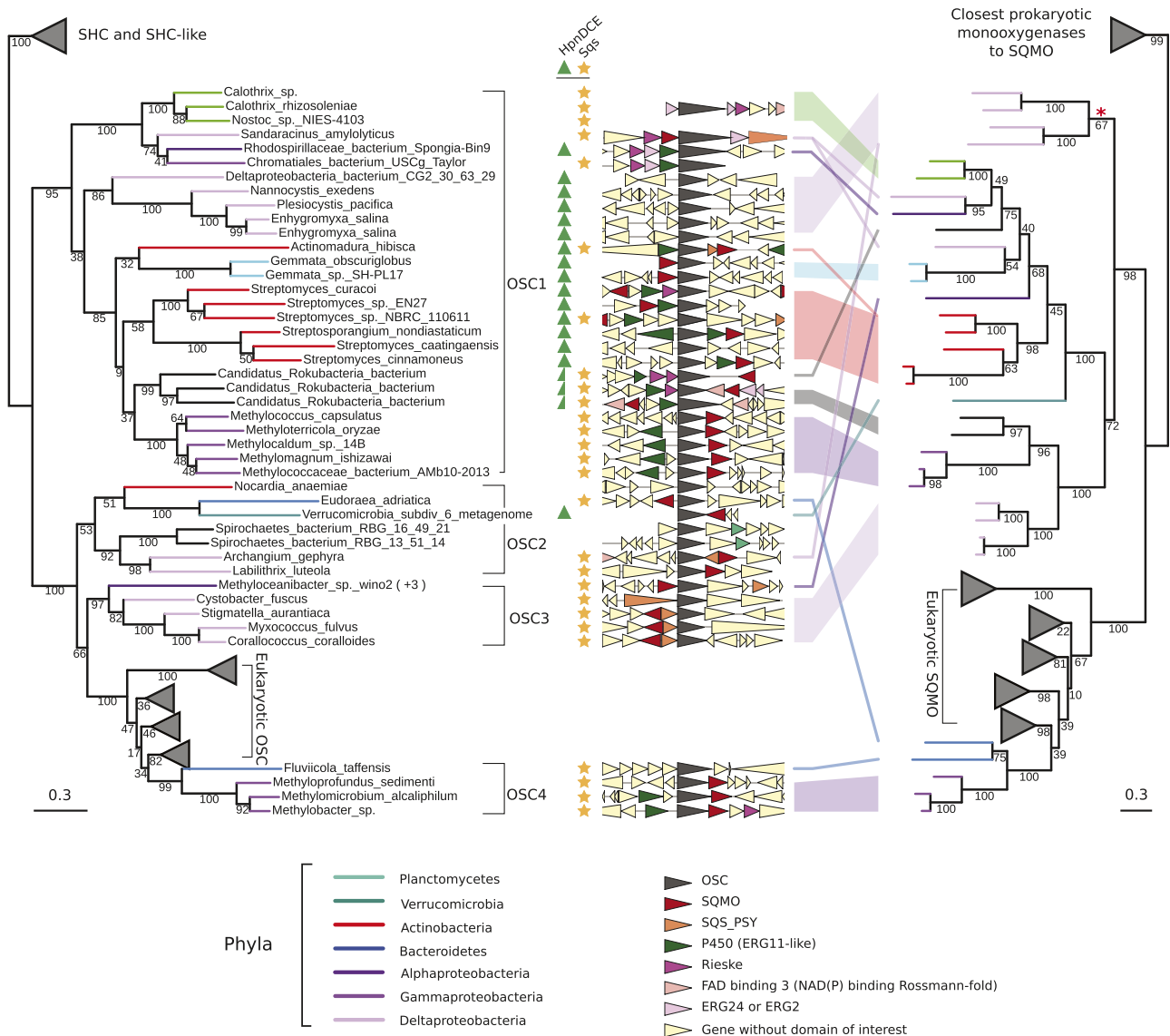


Fig. 5. Phylogenies and genomic context of OSCs and SQMOs. Left: ML reconstruction of OSCs rooted at SHC and with undefined-enzymes as the outgroup. Center-left: Green triangles or yellow stars indicate the presence of precursor synthetic genes in the genome, HpnCDE or Sqs, respectively. Center-right: Synteny of OSC in an ~10-kb area, with the Pfam domains of interest colored according to the legend. Right: ML reconstruction of the SQMO enzymes rooted to their closest monoxygenases as an outgroup. The lines connect the corresponding genes from the same organism. The red asterisk indicates the SQMOs of *Plesiocystis pacifica* and *Enhygromyxa salina*, and other related sequences that are located in the genome contiguous to an undefined-enzyme homologous to SHC (SHC-like; see [supplementary fig. S10, Supplementary Material online](#)). The branches are colored by phyla. The final alignments contained 726 and 471 amino acid positions for triterpene cyclase and SQMO, respectively. The trees were calculated with RAxML using the LG model (automatic model selection).

2003; [Desmond and Grimaldo 2009](#)): 1) that this is an artificial grouping due to the faster evolutionary rates in bacteria, provoking long-branch attraction (LBA) artifacts or 2) that these sequences were transferred to bacteria from ancestral eukaryotes before eukaryotic diversification. These interpretations intrinsically assume that sterol biosynthesis originated in the Eukarya. A possible influence of LBA on our observations is unlikely given the approach we have followed, using different models, methods, replicates, and large sampling to minimize the effect of LBA. However, LBA cannot be fully ruled out because of the higher evolutionary rates in bacteria. Concerning the second interpretation, we inferred several evolutionary histories for different bacterial taxa, showcasing

the complexity of sterol evolution in bacteria ([supplementary text, Supplementary Material online](#)). For example, the role of Proteobacteria in the evolution of sterol seems to be relevant because they are present in all our OSC groups, suggesting that *osc* and *sqmo* genes have been transferred to Proteobacteria more than once ([fig. 5](#)). Myxobacteria (OSC1–3) are possibly the most intriguing case when considering the role of sterols in eukaryogenesis as their genomes also encode other important eukaryotic signatures ([Moreira and López-García 1998](#)). In particular, the Myxobacteria OSC3 group that is closest to the eukaryotic stem also contains two facultative methylophilic alpha-proteobacteria ([fig. 5](#)). This result in part supports the syntrophic hypothesis

whereby Myxobacteria and alpha-proteobacteria methanotrophs could play important roles in the emergence of eukaryotes, in coexistence with archaea (Moreira and López-García 1998; López-García and Moreira 1999). Therefore, irrespective of the syntrophic hypothesis and considering sterol biosynthesis alone, it is likely that the ancestors of Myxobacteria were related in some way with the ancestral eukaryotes, although the direction of gene transfer cannot easily be established. The OSC3 (and OSC2) sequences are those most likely to be related to ancestral eukaryotes, as reflected by their phylogenetic position, with the probable exception of *Eudoraera adriatica* that has arborane synthase activity. Alternatively, our extended taxonomic sampling provides evidence for the presence of OSC1 sequences in a growing number of bacterial phyla, whereas OSC2–4 groups do not appear to grow at the same pace. The position of the OSC1 group does not support a LGT event from eukaryotes. Moreover, additional phylogenetic analysis of the Erg11 enzyme rejects a eukaryotic origin for the C14-demethylation of lanosterol mediated by this cytochrome, in particular for methylotrophic gamma-proteobacteria and actinobacteria species (supplementary fig. S12 and text, Supplementary Material online).

To interpret the origin of sterol biosynthesis genes in the context of eukaryogenesis, we incorporated our observations into the accepted view of the tree of life (the 2 domain topology), in which Bacteria and Archaea have independent origins, and the Eukarya emerged from within the Asgard clade (Eme et al. 2017). The absence of polycyclic triterpene cyclases in Archaea argues against the origin of sterol biosynthesis in Archaea or the last universal common ancestor. Therefore, the origin of sterol biosynthesis pathway can be traced back to Bacteria or early extinct eukaryotic lineages. In this scenario, it is difficult to conceive a eukaryotic origin for OSC, as the absence of triterpene cyclase in Archaea would imply the de novo appearance of OSC during eukaryogenesis. Moreover, our data demonstrate that polycyclic triterpene production was basal in bacterial evolution, possibly providing a more appropriate background for the evolution and divergence of triterpene cyclases. Alternatively, SQMOs belong to a large protein family whose closest homologs are found in prokaryotes, rejecting the origin of SQMO in eukaryotic protein families. In addition, the oxygen requirement of this enzyme implies that sterol biosynthesis occurred in aerobic environments (Summons et al. 2006). These observations reinforce the notion that the presence of sterol genes in eukaryotes is probably due to a bacterial contribution to early aerobic eukaryotic lineages (not archaeon). In the context of alternative scenarios of eukaryogenesis that invoke a symbiotic interaction between archaeal and bacterial organisms (López-García et al. 2017), our findings indicate that sterol synthetic genes would be provided by the bacterial partner.

The scarcity of OSC in Alpha-proteobacteria indicates that this bacterial contribution is unlikely to be related to the mitochondria. The possible functional role of sterol in eukaryogenesis could be related to the origin of the phagocytosis, which is thought to have required bacterial contributions

(Burns et al. 2018). However, these assumptions are still to be confirmed, both the origin of phagocytosis and its requirement for sterols (Takishita et al. 2017), and the possible need for phagocytes to originate mitochondria (Hampl et al. 2019).

The possibility that the origin of the genes involved in sterol biosynthesis is to be found within the Bacteria domain is supported by the aforementioned presence of C4 sterol demethylation in Methyloccocales that can produce sterol, probably also found in other bacteria from Delta-proteobacteria, Cyanobacteria, or Actinobacteria (Lee et al. 2018). This C4 demethylation is unrelated to the corresponding eukaryotic mechanism, showcasing that a pathway for sterol modification exists that is independent of the eukaryotic one (Lee et al. 2018). Importantly, we have recently shown that sterol biosynthesis is essential in the Planctomycetes bacteria *Gemmata obscuriglobus* (an OSC1 member: Rivas-Marín et al. 2019), which implies a breach in the uniqueness of the requirement for sterol in eukaryotes.

General View of Polycyclic Triterpene Evolution and Their Precursors

The Emergence of Cyclases Together with Their Precursors

Finally, we integrated the results obtained for the different enzymes considered in this work in an attempt to provide a unified view of the emergence and evolution of squalene and polycyclic triterpenes. The existence of two possible alternatives of squalene synthesis, HpnCDE or Sqs, raises the question of ancestry. The HpnCDE pathway seems to have originated before the diversification of Gracilicutes and some phyla from Terrabacteria, whereas the same possibility cannot be inferred for Sqs. On the other hand, by combining the phylogeny of HpnD and Sqs, we infer that *hpnCDE* genes were already in the genomes and Sqs appear to be more prone to LGT, probably because integrating the reactions in a single enzyme is in principal more efficient in energetic terms. These observations, in combination with the fact that the *hpnCDE* genes are both phylogenetically and mechanistically closer to carotenoid biosynthesis than Sqs (following the law of increasing complexity: Ourisson and Nakatani 1994), suggest that the *hpnCDE* genes are more ancestral.

By mapping the presence of squalene synthetic enzymes onto the SHC phylogeny, we noted some peculiarities in the distribution of the HpnCDE and Sqs pathways (supplementary fig. S7, Supplementary Material online). There are cases like the Planctomycetes and Actinobacteria phyla where only the HpnCDE enzymes are present, whereas other phyla have both squalene biosynthesis pathways. In other cases, such as Cyanobacteria, only Sqs and not HpnCDE are present, suggesting that Sqs provides an evolutionary advantage. By contrast, Trans IPPS HH enzymes were not found in most of the Firmicutes (some with complete genomes), in the Anamnox from the Planctomycetes phylum or in Delta-proteobacteria like Desulfurobacterales and Desulfuromonadales (with complete genomes: supplementary fig. S7, Supplementary Material online). Thus, precursor biosynthesis in these organisms might be independent of the Trans IPPS HH family and new pathways for precursor biosynthesis might still to be

found. This possibility is illustrated by the Firmicutes, which produce the sesquiterpenoid tetraprenyl- β -curcumene through the cooperation of three enzymes: Hept1, Hept2, and YtpB (Sato et al. 2011a). This precursor is then cyclized by SqhC to produce sporulenol rather than hopanoids, a compound only produced during sporulation (Bosak et al. 2008). Protein searches showed that the Hept2 enzyme (a polyprenyl transferase) is common to most organisms, whereas Hept1 and YtpB have a very restricted distribution in the Firmicute phylum, specifically in those bearing SqhC (supplementary fig. S7, Supplementary Material online). Thus, the biosynthesis of tetraprenyl- β -curcumene probably emerged within the Firmicute ancestor. Another plausible explanation for the absence of Trans IPPS HH in the Anammox, Desulfurobacterales and Desulfuromonadales is that they could obtain squalene from the environment, particularly as these bacteria coexist within syntrophic microbial communities (Morris et al. 2013; Zhu et al. 2019). Therefore, and with the exception of these three specific cases (and possibly few others), HpnCDE or Sqs appear to be present in most hopanoid producers.

The biosynthetic pathways of squalene that emerged together with each triterpene cyclase have not been considered previously. However, these overlooked data are important because it is informative about the metabolic links between polycyclic triterpene biosynthesis and other pathways. We found that the HpnCDE and SHC reconstructions share more similarities (in terms of topology and gene distribution within individual phylum) than those of Sqs and SHC (supplementary figs. 3 and 7, Supplementary Material online). These similarities are mainly illustrated by the presence of the HpnCDE and SHC proteins in the common ancestor of Proteobacteria and Actinobacteria-Chloroflexi-*Ca.* AD3-*Ca.* WPS-2, respectively (figs. 3A and 4). Hence, the HpnCDE pathway is most likely the primitive squalene pathway for hopanoid production. Similar simultaneous evolution of Sqs and OSC is unlikely to have happened as the bacterial OSC1 group suggests a possible bacterial origin of the genes involved in sterol synthesis, and some organisms from this group produce squalene via HpnCDE (fig. 5). On the other hand, it is important to note that the squalene biosynthesis enzymes (HpnCDE or Sqs) are not always associated with triterpene cyclase (fig. 3A and B), suggesting that the squalene biosynthesis pathways did not necessarily emerge to provide precursors of polycyclic triterpenes.

The Emergence of Sqs Could Have Uncoupled Polycyclic Triterpene and Carotenoid Precursor Synthesis

There is a wide variety of carotenoid molecules whose biosynthesis in some organisms is still unknown. The substrates of carotenoid and squalene biosynthesis pathways are not identical but they are chemically similar. In addition, genetic engineering has shown that some carotenoids can be synthesized using Sqs (Furubayashi et al. 2014a). Likewise, the function and specificity of Trans IPPS HH enzymes may be sensitive to a few amino acid changes, illustrating certain promiscuity among these enzymes (Furubayashi et al.

2014b). Therefore, the existence of metabolic cross-talks between the squalene and carotenoid biosynthesis pathways is a possibility. In fact, the involvement of three enzymes in the squalene synthetic pathway, suggests that the HpnCDE pathway may be metabolically more versatile due to the possible redirection of the intermediary compounds, for example, leading to carotenoid production. Therefore, if these HpnCDE proteins constitute a versatile functional pathway, the appearance of Sqs might be linked to the individualization of the squalene pathway.

One taxonomic group that exemplifies the possible versatility of the HpnCDE pathway is the Planctomycetes phylum, whose genomes only contains the HpnCDE enzymes (not crtB/M or Sqs). The colonies of most of members of this group have red, pink, or orange pigmentation and such pigmentation is suggested to be due to their unusual carotenoids, recently detected, but the biosynthesis of which has not been studied (Kallscheuer et al. 2019). Planctomycetal genomes have a variety of polyprenyl-transferases that may act on isoprenoid derivatives. In addition, the genomes of Planctomycetes do not encode CrtB or CrtM homologs, yet they encode HpnCDE enzymes and the amino oxidases related to carotenoids, such as CrtN or CrtP (supplementary fig. S13A and B, Supplementary Material online), illustrating the unconventional pathway of carotenoid biosynthesis in this phylum. We addressed this conundrum experimentally. Independent interruption of the aminoxidase *crtN* and *hpnE* genes in the planctomycete *Planctopirus limnophila* led to white colonies (supplementary fig. 13C, Supplementary Material online). We performed carotenoids extraction on these strains (supplementary fig. 13D, Supplementary Material online), followed by spectrophotometric analysis. The typical carotenoid UV/vis spectrum is observed in the wild-type samples but absent in the *crtN* and *hpnE* mutant ones (supplementary fig. 13E, Supplementary Material online). Hence, the pigmentation in Planctomycetes may well be due to the presence of carotenoid-like molecules and perhaps, HpnCDE enzymes are involved in their synthesis. This possibility is reinforced by the enhanced red pigmentation of the Planctomycete *Gemmata obscuriglobus* when sterol production is interrupted (Rivas-Marin et al. 2019). Similarly, interrupting hopanoid production in *P. limnophila* (at its SHC) leads to a more intense pink coloration of the colonies (Rivas-Marin E, personal communication). These results suggest that the accumulation of squalene (or its precursors) could be redirected toward carotenoid synthesis. Together, these experimental data suggest a metabolic link between carotenoid biosynthesis and the HpnCDE products, although additional experiments are necessary to test this assumption.

Conversely, bacteria like Cyanobacteria have Sqs and CrtB but not the HpnCDE pathway. Together, we suggest that some organisms containing only the HpnCDE pathway could use the same machinery for the biosynthesis of hopanoid and carotenoid precursors, although this does not mean that Sqs do not provide precursors for carotenoids, as some evidence might suggest (Furubayashi et al. 2014a). However, the absence of amino oxidases in the genomic neighborhood of Sqs

(supplementary fig. S2, Supplementary Material online) suggests that these enzymes are more specific to squalene production. Thus, we hypothesize that the emergence of Sqs could have provided a metabolic advantage that separates squalene biosynthesis from carotenoid precursor biosynthesis, thereby maintaining adequate and controlled physiological levels of polycyclic triterpenes and carotenoids.

Conclusions

In this study, we aimed to decipher the evolutionary relationship between squalene and carotenoid biosynthesis. Our results indicate that the HpnCDE pathway is a close derivative of the carotenoid pathway. Given the distribution of the HpnCDE and Sqs enzymes in the phylogenetic reconstructions, the origin of the HpnCDE pathway probably precedes that of Sqs. The HpnCDE enzymes may fulfill different roles in squalene production, demonstrating certain functional versatility of this pathway, whereas the functional origin of Sqs is possibly related to the individualization of squalene production. In both cases, squalene biosynthesis is presumed to have a bacterial origin. Downstream of the polycyclic triterpene pathway, triterpene cyclases are only found in Bacteria and Eukarya. Phylogenetic reconstructions suggest that bacteria contributed to polycyclic triterpene biosynthesis during eukaryogenesis adding another component to the eukaryotic membrane that has its origin associated with the bacterial contribution to early eukaryotic lineages. For example, the eukaryotic isoprenoid biosynthesis (mevalonate) pathway and the biosynthesis of G3P in Eukaryotes, and some members of the Archaea, are thought to have a bacterial origin (Lombard et al. 2012; Villanueva et al. 2017; Hoshino and Gaucher 2018).

Various lines of research can be proposed to test our conclusions and it would be interesting to address the possible alternative roles of the products of the HpnCDE enzymes. It will also be of interest to investigate if OSC or SQMO is essential in more bacteria from the OSC1 group, especially in actinobacterial species. Such dependence may have biomedical implications given that they are close relatives to *Mycobacteria* spp. in which sterol is essential for their infectious life cycle. Also, the absence of polycyclic triterpenes in Archaea seems noteworthy. Testing whether Archaea can incorporate hopanoids or sterols into their membrane would be interesting in order to address the compatibility of polycyclic triterpenes with the biochemical nature of their membranes.

Materials and Methods

Data Set Construction and Preparation for Alignment

We investigated different protein families involved in carotenoid and polycyclic triterpene biosynthesis: Trans IPPS HH (GenBank query sequences: Sqs, CAA71097.1; CrtB, WP_011338116.1; HpnD, CAA04733.1), the closest amine oxidases to HpnE (CAA04734.1); triterpene cyclases (CAA04735.1), the closest oxygenases to SQMO (CAA97201.1) and the closest demethylases to Erg11 (KZV10728.1); and RpoB as a phylogenetic marker

(YP_499096.2). We designed a workflow divided into two phases according to the nature of the data sets to be analyzed. In the first phase, we set out to build representative data sets of each protein family. We conducted a genome context analyses, and we built HMM profiles of each family for further analyses using a representative data set of organisms combining GEBA genomes (Mukherjee et al. 2017) and the first representative genome of each PAM proteome from UniProt. The rationale was to maximize both the taxonomic and genomic diversity of prokaryotes. As such, once we obtained a curated representation of prokaryotes, we extended the data set by adding 36 representative eukaryotes. Given the poor representation of bacteria containing sterol genes, we also included the bacterial genomes encoding OSCs that were available in the December 2017 release of the NCBI repository.

Protein searches were made through pHMMER (Potter et al. 2018) using different cutoff thresholds. For distant homologs, thresholds of 1e-3 (Trans IPPS HH, -HpnC, HpnD, and Sqs-) and 1e-5 (SHC, HpnE, and SQMO) were selected, also accounting for alignment length. In order to identify close homologs, we selected a stringent threshold of 1e-50 for Erg11 and 1e-80 for RpoB (filtering out those not bearing SHC). For all Trans IPPS HH homologs, we extracted the SQS_PSY domain, which covers almost the entire sequence. We confirmed the presence of the C-terminal region in Cyanobacteria and Firmicutes by extracting full sequences, and realigning them using MAFFT (Katoh and Standley 2013). The rest of the protein families were directly aligned using the MAFFT-linsi mode. Positions in the alignment with more than 95% of gaps were removed through Trimal (Capella-Gutiérrez et al. 2009). Finally, we removed from the alignment the partial sequences that cover <90% of the mean length of the alignment for Trans IPPS HH and 60% of the alignment for the rest of the proteins. Each set of sequences reduced the redundant sequences up to 99%, except the Trans IPPS HHs for which up to 50% of redundancy was settled in order to obtain reliable phylogenetic reconstructions. The final alignments obtained were designated as representative alignments.

The second phase aimed to increase our sampling by searching the nonredundant NCBI database. However, the taxonomy of prokaryotic genomes was standardized according to GTDB, a resource that improves the classification of uncultured bacteria, and provides a sound basis for ecological and evolutionary studies (Parks et al. 2018). We perform pHMMER searches applying a cutoff threshold of 1e-5 for the entire protein data set. For those genomes that are not translated to proteins, we performed TblastN (Gertz et al. 2006) searches applying a cutoff threshold of 1e-2 for the entire protein data set. For both strategies, the hits exhibiting an alignment coverage >60% were considered for further analyses. To increase the quality of the data, we progressively removed redundant sequences (using CDHIT: Fu et al. 2012) for each prokaryotic phylum or eukaryotic kingdom according to the GTDB definitions. First, we applied an 80% redundancy threshold and if the taxonomic set was more than 100 sequences, a second threshold of 60% was applied. No redundancy was applied to the gene distribution data set. Each set

of sequences (both for phylogeny and for gene distribution) was then aligned to the respective HMM profile obtained from the previous representative alignment using HMMALIGN (Potter et al. 2018), and the partial sequences that aligned <80% of the mean sequence length of each set were removed. However, for the Trans IPPS HH set (~20,000 sequences), we designed an alternative filtering protocol in which a preliminary step included the removal of sequences with more than 90% identity using CDHIT (Fu et al. 2012). We then aligned them to the respective representative HMM profile obtained previously, further removing redundancy at the 50% level. This approach was devised in order to reduce the computational time, to improve visualization and to help reach convergence in the phylogenetic analyses. We then obtained the region of interest to reconstruct the phylogeny, excluding artifacts or errors in gene prediction. From the resulting Trans IPPS HH tree of gene distribution, the HpnD, HpnC, and Sqs subfamilies were selected in order to achieve more accurate reconstructions. The regions of interest of each set (the HMM alignment) were converted into unaligned sequences that were realigned through the MAFFT-linsi mode (Katoh and Standley 2013), removing those positions with more than 95% of gaps through Trimal. For the partition reconstruction of HpnCDE, we selected those genomes that only have one copy of HpnD, HpnC, and HpnE, and concatenated the respective alignments. Note that the HpnD alignment also included main taxonomic groups bearing only HpnD. We referred to this second phase as the NCBI data set.

The genomes that did not exhibit similarity with Trans IPPS HH, such as Syntrophobacterales or Annamox organisms, were additionally checked through TBlastN searches and no hits were found.

Phylogeny

Our representative alignments of the Trans IPPS HH, triterpene cyclase, OSC, and SQMO, and ERG11 homologs, were used to perform Bayesian phylogenetic reconstructions with MrBayes (Ronquist and Huelsenbeck 2003). Evolutionary models were chosen according to Prottest (Darriba et al. 2011) and the Markov chain Monte Carlo analyses were sufficiently long to obtain good values of convergence. For alignments derived from the NCBI data set, we performed maximum-likelihood (ML) reconstructions using RAxML (Stamatakis 2014) and IQ-TREE (Nguyen et al. 2015). Although RAxML and IQ-TREE reconstructions provided congruent topologies for the same protein set, RAxML did not provide satisfying bootstrap values at basal nodes for some data sets and thus, we focused on the IQ-TREE reconstructions. We obtained branch supports with the ultrafast bootstrap (Hoang et al. 2018) implemented in the IQ-TREE software. In this case, the evolutionary models of each set of sequences were automatically selected using ModelFinder (Kalyaanamoorthy et al. 2017) implemented in IQ-Tree and chosen according to BIC criterion. In addition, we also performed reconstructions with alternative evolutionary complex models like LG4X, obtaining general congruent topologies between the trees from the same protein data

set. We inferred the ML tree of the HpnCDE concatenated data set using the edge-unlinked partition model in IQ-TREE (Chernomor et al. 2016). The trees resulting from the representative and the NCBI data sets were generally congruent (only basal and poorly resolved nodes differed in some cases), suggesting that the phylogenies are robust to the reconstruction method (Bayesian vs. ML) and to the differential taxonomic sampling. The trees obtained from the NCBI data set constructed by different evolutionary models were also congruent with each other, providing robustness to our reconstructions. All trees were visualized and annotated using iTOL (Letunic and Bork 2019).

Genome Context Analyses

We defined the genome context as the arrangement of neighboring genes relative to the gene of interest. To analyze the genome contexts of proteins containing the SQS_PSY domain and of cyclases, we extracted the genomic sequence of the genes 10-kb upstream and downstream. From these genomic fragments, we extracted the CDS using PRODIGAL (Hyatt et al. 2010), and we further annotated the coding proteins using the PFAM database (Finn et al. 2011). We ran HMMSCAN (Potter et al. 2018), and we parsed the output to keep the longest coverage and best *e*-value in order to minimize the effect of overlapping domains.

Identification of Potential Carotenoid Biosynthetic Genes in *P. limnophila*

Fresh electrocompetent cells were prepared from 400 ml of a *P. limnophila* DSM3776 culture at OD₆₀₀ of 0.4 in modified PYGV as previously described (Rivas-Marin, E. 2016). The cells were washed twice with 100 and 50 ml of ice-cold double distilled sterile water and once with 2 ml of ice-cold 10% glycerol. Then, the pellet was resuspended in 400 ml of ice-cold 10% glycerol, and aliquots of 100 μ l were dispensed into 0.1-mm gapped electroporation cuvettes along with 1 μ l of EZ-Tn5 solution and 1 ml of Type-One restriction inhibitor (Epicenter). Electroporation was performed with a Bio-Rad Micropulser (Ec3 pulse, voltage [V] 3.0 kV). Electroporated cells were immediately recovered in 1 ml of cold modified PYGV and incubated at 28°C for 2 h with shaking. The cells were then plated onto modified PYGV plates supplemented with kanamycin 50 μ g ml⁻¹ and were incubated at 28°C until colony formation after 5–7 days. White colonies were segregated onto fresh selection plates. To verify Tn5 insertions and their locations, DNA was isolated using the Wizard Genomic DNA Purification Kit (Promega), and analyzed by semirandom polymerase chain reaction (PCR) (doi:10.1128/JB.183.21.6207-6214.2001). Genomic DNA was used as the template DNA in a 20- μ l PCR mixture containing primer Map Tn5 A fwd (5'-ATCAGGACATAGCGTTGGC) and either primer CEKG 2A (5'-GGCCACGCGTCTGACTAGTACN₁₀AGAG-3'), CEKG 2B (5'-GGCCACGCGTCTGACTAGTACN₁₀ACGCC-3'), or CEKG 2C (5'-GGCCACGCGTCTGACTAGTACN₁₀GATAT-3'); 1 μ l of a 1:5 dilution of this reaction mixture was used as the template DNA for a second PCR performed with primers Map Tn5 B fwd (5'-AAGAGCTTGGCGCGAATG-3') and CEKG 4 (5'-GGCCACGCGT

GACTAGTAC-3'). For the first reaction, the thermocycler conditions were 94 °C for 2 min, followed by six cycles of 94 °C for 30 s, 42 °C for 30 s (with the temperature reduced 1 °C per cycle), and 72 °C for 3 min and then 25 cycles of 94 °C for 30 s, 65 °C for 30 s, and 72 °C for 3 min; for the second reaction, the thermocycler conditions were 30 cycles of 94 °C for 30 s, 65 °C for 30 s, and 72 °C for 3 min. The DNA of purified PCR products (GFX PCR DNA and Gel Band Purification Kit GE Healthcare) was sequenced by using primer Map Tn5 B fwd.

Carotenoid Extraction and Spectrophotometric Analysis

Cell pellets of 200-ml cultures (OD 1) were lyophilized (dry weight 0.07 g) and resuspended in 3 ml of ultra-pure ethanol. Cells were lysed by ultrasonication (amplitude 35%; Branson; Sonifier SFX250) for 80 s (5 s on and 5 s off). The extracts were then clarified by centrifugation (16,000 × g) and stored at 4 °C until analysis. Extraction was performed under dim light to prevent carotenoid degradation. The UV/vis spectra (wavelength range of 300–700 nm) of the different sample were formed in UV-1800 Spectrophotometer (Shimadzu).

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

C.S.-M., E.R.-M., and D.P.D. were funded by the Spanish Ministry of Economy and Competitiveness (grants BFU2013-40866-P and BFU2016-78326-P). A.M.R. was funded by the Spanish Ministry of Economy and Competitiveness (RTI2018-096735-B-100). The authors thank J. Avalos and M.C. Limón, Universidad de Sevilla, Spain, for advice on carotenoid analysis.

References

Banta AB, Wei JH, Gill CCC, Giner J-L, Welander PV. 2017. Synthesis of arborane triterpenols by a bacterial oxidosqualene cyclase. *Proc Natl Acad Sci U S A*. 114(2):245–250.

Battistuzzi FU, Hedges SB. 2009. A major clade of prokaryotes with ancient adaptations to life on land. *Mol Biol Evol*. 26(2):335–343.

Belin BJ, Busset N, Giraud E, Molinaro A, Silipo A, Newman DK. 2018. Hopanoid lipids: from membranes to plant–bacteria interactions. *Nat Rev Microbiol*. 16(5):304–315.

Betts HC, Puttick MN, Clark JW, Williams TA, Donoghue PCJ, Pisani D. 2018. Integrated genomic and fossil evidence illuminates life's early evolution and eukaryote origin. *Nat Ecol Evol*. 2(10):1556–1562.

Bird CW, Lynch JM, Pirt FJ, Reid WW, Brooks CJW, Middleditch BS. 1971. Steroids and squalene in *Methylococcus capsulatus* grown on methane. *Nature* 230(5294):473–474.

Bode HB, Zeggel B, Silakowski B, Wenzel SC, Reichenbach H, Müller R. 2003. Steroid biosynthesis in prokaryotes: identification of myxobacterial steroids and cloning of the first bacterial 2,3(S)-oxidosqualene cyclase from the myxobacterium *Stigmatella aurantiaca*. *Mol Microbiol*. 47(2):471–481.

Bosak T, Losick RM, Pearson A. 2008. A polycyclic terpenoid that alleviates oxidative stress. *Proc Natl Acad Sci U S A*. 105(18):6725–6729.

Burns JA, Pittis AA, Kim E. 2018. Gene-based predictive models of trophic modes suggest Asgard archaea are not phagocytotic. *Nat Ecol Evol*. 2(4):697–704.

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25:1972–1973.

Cavalier-Smith T. 2002. The phagotrophic origin of eukaryotes and phylogenetic classification of Protozoa. *Int J Syst Evol Microbiol*. 52(2):297–354.

Chernomor O, von Haeseler A, Minh BQ. 2016. Terrace aware data structure for phylogenomic inference from supermatrices. *Syst Biol*. 65(6):997–1008.

Damsté JSS, Rijpstra WIC, Dedysh SN, Foesel BU, Villanueva L. 2017. Pheno- and genotyping of hopanoid production in Acidobacteria. *Front Microbiol*. 8:968.

Darriba D, Taboada GL, Doallo R, Posada D. 2011. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics*. 27(8):1164–1165.

Deschamps P, Zivanovic Y, Moreira D, Rodriguez-Valera F, López-García P. 2014. Pangenome evidence for extensive interdomain horizontal transfer affecting lineage core and shell genes in uncultured planktonic thaumarchaeota and euryarchaeota. *Genome Biol Evol*. 6(7):1549–1563.

Desmond E, Gribaldo S. 2009. Phylogenomics of sterol synthesis: insights into the origin, evolution, and diversity of a key eukaryotic feature. *Genome Biol Evol*. 1:364–381.

Eme L, Spang A, Lombard J, Stairs CW, Ettema T. 2017. Archaea and the origin of eukaryotes. *Nat Rev Microbiol*. 15(12):711–723.

Finn RD, Clements J, Eddy SR. 2011. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res*. 39(Suppl):W29–W37.

Fischer WW, Pearson A. 2007. Hypotheses for the origin and early evolution of triterpenoid cyclases. *Geobiology* 5:19–34.

Frickey T, Kannenberg E. 2009. Phylogenetic analysis of the triterpene cyclase protein family in prokaryotes and eukaryotes suggests bidirectional lateral gene transfer. *Environ Microbiol*. 11(5):1224–1241.

Fu L, Niu B, Zhu Z, Wu S, Li W. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28(23):3150–3152.

Furubayashi M, Li L, Katabami A, Saito K, Umeno D. 2014a. Construction of carotenoid biosynthetic pathways using squalene synthase. *FEBS Lett*. 588(3):436–442.

Furubayashi M, Li L, Katabami A, Saito K, Umeno D. 2014b. Directed evolution of squalene synthase for dehydrosqualene biosynthesis. *FEBS Lett*. 588(18):3375–3381.

Gertz EM, Yu YK, Agarwala R, Schäffer AA, Altschul SF. 2006. Composition-based statistics and translated nucleotide searches: improving the TBLASTN module of BLAST. *BMC Biol*. 4(1):41.

Gold DA, Caron A, Fournier GP, Summons RE. 2017. Paleoproterozoic sterol biosynthesis and the rise of oxygen. *Nature* 543(7645):420–423.

Gouy R, Baurain D, Philippe H. 2015. Rooting the tree of life: the phylogenetic jury is still out. *Philos Trans R Soc B* 370(1678):20140329.

Grudzinski W, Nierzwicki L, Welc R, Reszczyńska E, Luchowski R, Czub J, Gruszecki WI. 2017. Localization and orientation of xanthophylls in a lipid bilayer. *Sci Rep*. 7(1):9619.

Gruszecki WI, Strzałka K. 2005. Carotenoids as modulators of lipid membrane physical properties. *Biochim Biophys Acta* 1740(2):108–115.

Haml V, Čepička I, Eliáš M. 2019. Was the mitochondrion necessary to start eukaryogenesis? *Trends Microbiol*. 27(2):96–104.

Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol*. 35(2):518–522.

Hoshino Y, Gaucher EA. 2018. On the origin of isoprenoid biosynthesis. *Mol Biol Evol*. 35(9):2185–2197.

Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, Butterfield CN, Hemsdorf AW, Amano Y, Ise K, et al. 2016. A new view of the tree of life. *Nat Microbiol*. 1:16048.

Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11(1):119.

Jia Q, Chen X, Köllner TG, Rinkel J, Fu J, Labbé J, Xiong W, Dickschat JS, Gershenzon J, Chen F. 2019. Terpene synthase genes originated from

- bacteria through horizontal gene transfer contribute to terpenoid diversity in fungi. *Sci Rep.* 9(1):9223.
- Kallscheuer N, Moreira C, Airs R, Llewellyn CA, Wiegand S, Jogler C, Lage OM. 2019. Pink- and orange-pigmented Planctomycetes produce saxoranthin-type carotenoids including a rare C45 carotenoid. *Environ Microbiol Rep.* 11(6):741–748.
- Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods.* 14(6):587–589.
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 30(4):772–780.
- Klassen JL. 2010. Phylogenetic and evolutionary patterns in microbial carotenoid biosynthesis are revealed by comparative genomics. *PLoS One* 5(6):e11257.
- Kupisz K, Sujak A, Patyra M, Trebacz K, Gruszecki WI. 2008. Can membrane-bound carotenoid pigment zeaxanthin carry out a trans-membrane proton transfer? *Biochim Biophys Acta* 1778(10):2334–2340.
- Lee AK, Banta AB, Wei JH, Kiemle DJ, Feng J, Giner J-L, Welander PV. 2018. C-4 sterol demethylation enzymes distinguish bacterial and eukaryotic sterol synthesis. *Proc Natl Acad Sci U S A.* 115(23):5884–5889.
- Letunic I, Bork P. 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 47(W1):W256–W259.
- Li FW, Brouwer P, Carretero-Paulet L, Cheng S, Vries J. D, Delaux PM, Eily A, Koppers N, Kuo LY, Li Z, et al. 2018. Fern genomes elucidate land plant evolution and cyanobacterial symbioses. *Nat Plants* 4(7):460–472.
- Lombard J, López-García P, Moreira D. 2012. The early evolution of lipid membranes and the three domains of life. *Nat Rev Microbiol.* 10(7):507–515.
- López-García P, Eme L, Moreira D. 2017. Symbiosis in eukaryotic evolution. *J Theor Biol.* 7(434):20–33.
- López-García P, Moreira D. 1999. Metabolic symbiosis at the origin of eukaryotes. *Trends Biochem Sci.* 24:88–93.
- López-García P, Zivanovic Y, Deschamps P, Moreira D. 2015. Bacterial gene import and mesophilic adaptation in archaea. *Nat Rev Microbiol.* 13(7):447–456.
- Moreira D, López-García P. 1998. Symbiosis between methanogenic Archaea and δ -Proteobacteria as the origin of eukaryotes: the syntrophic hypothesis. *J Mol Evol.* 47(5):517–530.
- Morris BEL, Henneberger R, Huber H, Moissl-Eichinger C. 2013. Microbial syntrophy: interaction for the common good. *FEMS Microbiol Rev.* 37(3):384–406.
- Mukherjee S, Seshadri R, Varghese NJ, Eloe-Fadrosh EA, Meier-Kolthoff JP, Göker M, Coates RC, Hadjithomas M, Pavlopoulos GA, Paez-Espino D, et al. 2017. 1,003 reference genomes of bacterial and archaeal isolates expand coverage of the tree of life. *Nat Biotechnol.* 35(7):676–683.
- Nakashima T, Inoue T, Oka A, Nishino T, Osumi T, Hata S. 1995. Cloning, expression, and characterization of cDNAs encoding *Arabidopsis thaliana* squalene synthase. *Proc Natl Acad Sci U S A.* 92(6):2328–2332.
- Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 32(1):268–274.
- Ourisson G, Nakatani Y. 1994. The terpenoid theory of the origin of cellular life: the evolution of terpenoids to cholesterol. *Chem Biol.* 1(1):11–23.
- Ourisson G, Rohmer M, Poralla K. 1987. Prokaryotic hopanoids and other polyterpenoid sterol surrogates. *Annu Rev Microbiol.* 41(1):301–333.
- Pan JJ, Solbiati JO, Ramamoorthy G, Hillerich BS, Seidel RD, Cronan JE, Almo SC, Poulter CD. 2015. Biosynthesis of squalene from farnesyl diphosphate in bacteria: three steps catalyzed by three enzymes. *ACS Cent Sci.* 1(2):77–82.
- Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil PA, Hugenholtz P. 2018. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol.* 36(10):996–1004.
- Pearson A, Budin M, Brocks JJ. 2003. Phylogenetic and biochemical evidence for sterol synthesis in the bacterium *Gemmata obscuriglobus*. *Proc Natl Acad Sci U S A.* 100(26):15352–15357.
- Peretó J, López-García P, Moreira D. 2004. Ancestral lipid biosynthesis and early membrane evolution. *Trends Biochem Sci.* 29(9):469–477.
- Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD. 2018. HMMER web server: 2018 update. *Nucleic Acids Res.* 46(W1):W200–W204.
- Rivas-Marin E, Stettner S, Gottshall EY, Santana-Molina C, Helling M, Basile F, Ward NL, Devos DP. 2019. Essentiality of sterol synthesis genes in the planctomycete bacterium *Gemmata obscuriglobus*. *Nat Commun.* 10:1–6.
- Rodríguez-Concepción M, Avalos J, Bonet ML, Boronat A, Gomez-Gomez L, Hornero-Mendez D, Limon MC, Meléndez-Martínez AJ, Olmedilla-Alonso B, Palou A, et al. 2018. A global perspective on carotenoids: metabolism, biotechnology, and benefits for nutrition and health. *Prog Lipid Res.* 70:62–93.
- Rohmer M, Bouvier P, Ourisson G. 1979. Molecular evolution of biomembranes: structural equivalents and phylogenetic precursors of sterols. *Proc Natl Acad Sci U S A.* 76(2):847–851.
- Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19(12):1572–1574.
- Sáenz JP, Grosser D, Bradley AS, Lagny TJ, Lavrynenko O, Broda M, Simons K. 2015. Hopanoids as functional analogues of cholesterol in bacterial membranes. *Proc Natl Acad Sci U S A.* 112(38):11971–11976.
- Sáenz JP, Sezgin E, Schwille P, Simons K. 2012. Functional convergence of hopanoids and sterols in membrane ordering. *Proc Natl Acad Sci U S A.* 109:14236–14240.
- Sandmann G. 2002. Molecular evolution of carotenoid biosynthesis from bacteria to plants. *Physiol Plant.* 116(4):431–440.
- Sato T, Yoshida S, Hoshino H, Tanno M, Nakajima M, Hoshino T. 2011a. Sesquiterpenes (C₁₅ terpenes) biosynthesized via the cyclization of a linear C₁₅ isoprenoid by a tetraprenyl- β -curcumene synthase and a tetraprenyl- β -curcumene cyclase: identification of a new terpene cyclase. *J Am Chem Soc.* 133(25):9734–9737.
- Sato T, Hoshino H, Yoshida S, Nakajima M, Hoshino T. 2011b. Bifunctional triterpene/sesquiterpene cyclase: tetraprenyl- β -curcumene cyclase is also squalene cyclase in *Bacillus megaterium*. *J Am Chem Soc.* 133(44):17540–17543.
- Sezgin E, Levental I, Mayor S, Eggeling C. 2017. The mystery of membrane organization: composition, regulation and roles of lipid rafts. *Nat Rev Mol Cell Biol.* 18(6):361–374.
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9):1312–1313.
- Summons RE, Bradley AS, Jahnke LL, Waldbauer JR. 2006. Steroids, triterpenoids and molecular oxygen. *Philos Trans R Soc B* 361(1470):951–968.
- Takishita K, Chikaraishi Y, Leger MM, Kim E, Yabuki A, Ohkouchi N, Roger AJ. 2012. Lateral transfer of tetrahymanol-synthesizing genes has allowed multiple diverse eukaryote lineages to independently adapt to environments without oxygen. *Biol Direct* 7(1):5.
- Takishita K, Chikaraishi Y, Tanifuji G, Ohkouchi N, Hashimoto T, Fujikura K, Roger AJ. 2017. Microbial eukaryotes that lack sterols. *J Eukaryot Microbiol.* 64(6):897–900.
- Villanueva L, Rijpstra WIC, Schouten S, Damsté J. 2014. Genetic biomarkers of the sterol-biosynthetic pathway in microalgae. *Environ Microbiol Rep.* 6(1):35–44.
- Villanueva L, Schouten S, Damsté J. 2017. Phylogenomic analysis of lipid biosynthetic genes of Archaea shed light on the “lipid divide”. *Environ Microbiol.* 19(1):54–69.
- Wei JH, Yin X, Welander PV. 2016. Sterol synthesis in diverse bacteria. *Front Microbiol.* 7:990.

- Welander PV. 2019. Deciphering the evolutionary history of microbial cyclic triterpenoids. *Free Radic Biol Med.* 140:270–278.
- Zaremba-Niedzwiedzka K, Caceres EF, Saw JH, Bäckström D, Juzokaite L, Vancaester E, Seitz KW, Anantharaman K, Starnawski P, Kjeldsen KU, et al. 2017. Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* 541(7637):353–358.
- Zhu G, Wang S, Wang C, Zhou L, Zhao S, Li Y, Li F, Jetten MSM, Lu Y, Schwark L. 2019. Resuscitation of anammox bacteria after >10,000 years of dormancy. *ISME J.* 13(4):1098–1109.