Archaeal lineages related to eukaryotes encode functional diterpenoid cyclases

Hanon McShea[1†], Valerie De Anda[2,4†^], Jochen J. Brocks[3], Paula V. Welander[1*] and Brett J. Baker[2,4*]

[1]Department of Earth System Science, Stanford University, Stanford, CA USA
[2]Department of Marine Science, University of Texas Austin, Marine Science Institute, Port Aransas, TX USA
[3]Research School of Earth Sciences, Australian National University, Canberra, ACT, Australia
[4]Department of Integrative Biology, University of Texas Austin, TX USA
[†]Contributed equally
[*]Corresponding authors: welander@stanford.edu and acidophile@gmail.com
^Current affiliation: Fort Lauderdale Research and Education Center, University of Florida, Fort Lauderdale, FL, USA

Keywords: Asgardarchaeota, sterols, diterpenoid cyclases, biosynthesis, eukaryogenesis

Summary.

The first eukaryotic cell originated through the union of an archaeon (Asgardarchaeota) and a bacterium (Alphaproteobacteria)[1–4]. Little is known about the molecular basis of eukaryogenesis, but it is likely that lipids played a key role in this process. Modern eukaryotic membranes contain polycyclic triterpenoids (mainly sterols) that are essential for a variety of cellular functions, but these lipids have not been identified in archaea. The lipid composition of Asgardarchaeota including the newly described class Hodarchaeales, which share a common ancestor with eukaryotes, is unknown. Here, we investigated the potential for Asgards to produce cyclic terpenoids. Phylogenomics coupled with structural prediction revealed that Asgards lack the capacity to make sterols, yet organisms from the clades Hodarchaeales and Kariarcheales encode for divergent homologs of sterol cyclases, predicted to produce diterpenoids. We tested the functionality of these enzymes *in vitro*, showing that they cyclize geranylgeranyl pyrophosphate to form bicyclic halimadienyl pyrophosphate. Halimadienyl lipids have previously been shown to mediate intracellular persistence of *Mycobacterium tuberculosis* in host endosomes[5–7], and may function similarly Asgardarchaeota. This is the first evidence of experimentally validated diterpenoid cyclases in archaea, providing new insights into the biochemistry of these microbes pivotal in the evolution of complex cellular life.

Main. Lipid composition varies between eukaryotic, bacterial, and archaeal membranes, a taxonomic pattern often referred to as the lipid divide[8]. Bacteria and eukaryotes share a common bulk membrane architecture, composed of fatty acyl chains ester-bound to glycerol-3-phosphate (G3P), whereas archaeal membranes primarily consist of isoprenoid alkyl chains ether-linked to glycerol-1-phosphate (G1P). This poses a challenge for the origin of modern eukaryotic membranes, given that current models propose that eukaryotic cells emerged through the engulfment of a bacterial endosymbiont[4] by an archaeal host, likely related to Hodarcheales, phylum Asgardarchaeota or "Asgard" for brevity[1,2,9–11]. However, the lipid composition of Hodarchaeales, as well as the role of lipids in eukaryogenesis, is unknown. Engulfment in eukaryotes, such as phagocytosis, are partly mediated by membrane lipids, with specialized lipids like cholesterol playing a critical role[12]. Cholesterol is one example of a

cyclic triterpenoid, a diverse class of polycyclic lipids that are produced by eukaryotes and bacteria but not yet found in archaea[13].

Eukaryotes primarily produce tetracyclic sterols that regulate fundamental membrane properties such as fluidity, permeability, homeostasis, and dynamics[14,15], and the formation of liquid-ordered membrane microdomains[16], a hallmark of eukaryotic life. Some bacteria also synthesize sterols[17,18], and many synthesize the structurally related pentacyclic hopanoids which also modulate membrane fluidity and permeability in response to physiological stress conditions such as pH and temperature fluctuations[19,20]. Sterols are found in almost all eukaryotes and some bacteria[17,21], while hopanoids are relatively widespread in bacteria[22,23]. The lack of cyclic triterpenoids in archaea raises questions about the origin of these lipids in eukaryotes. Taking advantage of our recent expansion of the Asgard genomic catalogue, which includes hundreds of novel lineages[24], we evaluated the genomic capacity for sterol and other polycyclic terpenoid biosynthesis across ˜5K eukaryotic and archaeal genomes (Supplementary Table 1).

Asgardarchaeota lack sterol biosynthesis. Comparative phylogenetic studies indicate that the last eukaryotic common ancestor (LECA) had a sterol synthesis pathway, with subsequent evolution of this pathway involving gene duplications and losses[25]. These sterol synthesis enzymes are found in diverse bacteria [23,26–29], but not in alphaproteobacterial relatives of the mitochondrion[17]. Thus, there are three plausible hypotheses for the origin of sterols in eukaryotes: sterols were either present in the Asgard ancestor of eukaryotes, arose *de novo* on the long eukaryotic stem lineage[26], or were contributed by a third, likely bacterial[28], organism, either during or after eukaryogenesis. To address the first hypothesis, we investigated the possibility that Asgards are an ancestral source of sterols for eukaryotes. We searched for sterol biosynthesis proteins using hidden Markov models (HMMs), both publicly available[30] and custom models (see Methods, Extended Data Fig. 1, and Supplementary Table 2). This search included (1) isoprenoid lipid synthesis by polyprenyl synthase (PPS; PF00348 and IPR008949), (2) squalene synthesis catalyzed by squalene/phytoene synthase (SQS or HpnCD; PF00494), (3) squalene epoxidation catalyzed by either squalene epoxidase (SMO or SQE; PF08491) or alternative squalene epoxidase (altSMO; PF04116), and (4) terpenoid cyclization by oxidosqualene cyclase (OSC; PF13249 and PF13243). We also gathered homologous biosynthesis genes from public databases (see Methods). Given the functional diversity of these enzymes, we performed phylogenetic analyses to determine whether Asgard proteins are more similar to those involved in sterol biosynthesis or functionally divergent homologs (Extended Data Figs. 2-5, and Supplementary Text).

Isoprenoid synthesis. The first step in sterol biosynthesis is isoprenoid chain elongation, performed by polyprenyl synthetase (PPS). PPSs belong to a large protein family that produces isoprenoid chains of various lengths, and also includes metal-dependent Type I terpenoid cyclases[31]. Phylogenetic analysis revealed that Asgards have an enormous diversity of PPS homologs (total 1508/4519 analyzed), 69% of which belong to uncharacterized groups within the large PPS family (Extended Data Fig. 4, groups A-Y). Group G is composed of 13 Hermodarchaeia, 32 Hodarchaeales, 5 Kariarchaeaceae, and 16 Thorarchaeia, and is monophyletic to octaprenyl synthases, enzymes involved in the biosynthesis of C_{40} (40-carbon) isoprenoids. Group E (including 44 Thorarchaeia, 5 Helarchaeales, 1 Njordarchaeales, 4 Baldrarchaeia and 6 Asgardarchaeia) is closely related to eukaryotic squalene synthases. Group H, containing 6 Hodarchaeales and an Hermodarchaeia sequence, form a monophyletic group with eukaryotic Type I terpenoid cyclases, including fungi (Trichoderma), spikemoss (Selaginella) and acellular slime molds that produce volatile terpenes[32] and diterpene

glycosides. Asgard clade D (composed of 4 Lokiarchaeales and 1 Thorarchaeia) groups with the subfamily that produces C_{20} isoprenoid chains (geranylgeranyl pyrophosphate synthase, or GGPS) (Extended Data Fig. 4). The presence of GGPS homologs is consistent with glycerol dialkyl glycerol tetraether (GDGT) lipid biosynthesis previously reported[9], yet the unexpected distribution of PPS indicates undescribed biosynthetic roles in Asgards.

Squalene synthesis. The next step in sterol production after farnesyl pyrophosphate (FPP) synthesis is the fusion of two FPP (C_{15}) molecules to form the C_{30} molecule squalene. Asgardarchaeota do not appear to code for proteins related to squalene synthase (SQS). Instead, Asgard sequences fall within the phytoene synthases (PSY), which fuse two molecules of C_{20} geranylgeranyl pyrophosphate (GGPP) into C_{40} phytoene, a precursor for carotenoid biosynthesis (Extended Data Fig. 5). Carotenoids are produced by many organisms including archaea[33], and function as antioxidants, membrane stabilizers, and photoreceptors in photosynthetic organisms[34].

Squalene epoxidation. The epoxidation of squalene to form (*S*)-2,3-oxidosqualene is a reaction catalyzed by one of two enzymes. The first, squalene monooxygenase (SMO), belongs to a large FAD-binding superfamily and is found in most eukaryotes and sterol-producing bacteria[35] (Extended Data Fig. 6). The second, AltSMO, belongs to a fatty acid hydroxylase superfamily and was first discovered in diatoms and found to be distributed in a variety of eukaryotes that lack SMO[36]. We identified distant homologs of both types of enzymes in Asgards. Asgard SMO homologs are most closely related to the archaeal GDGT biosynthesis protein geranylgeranyl reductase (Extended Data Fig. 6) while Asgard altSMO homologs are most closely related to beta-carotene hydroxylases from Rhodobacterales (Extended Data Fig. 7). Beta-carotene hydroxylases perform the last step in zeaxanthin biosynthesis, further supporting the possibility of carotenogenesis in Asgards.

Terpenoid cyclization. Cyclization is perhaps the most dramatic and well-studied biochemical step in terpenoid biosynthesis. This reaction occurs via a cascade of bond rearrangements resulting in a polycyclic product. For sterols, it is performed by oxidosqualene cyclase, a member of the Type II terpenoid cyclase family, all of which perform metal-independent cyclization reactions on terpenoid substrates of various lengths. Asgard cyclase homologs are clearly not sterol or hopanoid cyclases. Rather, they fall within the diterpenoid (C_{20}) clade of the Type II cyclase family (Fig. 1**a**). We initially identified four full-length terpenoid cyclases from Asgards, all of which had the active site motif (DxD) necessary to initiate the cyclization reaction[37] (Extended Data Fig. 8). A subsequent search with a custom protein HMM specific to diterpenoid cyclases revealed six additional Asgard cyclases.

Through comparative genomics of expanded archaeal[24] and eukaryotic diversity[38], we have ruled out an archaeal origin of sterols and polycyclic terpenoids. We additionally found no evidence that eukaryotes inherited Asgard genes that would evolve into sterol biosynthesis genes on the eukaryotic stem lineage. Therefore, it is likely that such genes have an origin in bacteria. This highlights the likely significance of genetic contributions from organisms outside the primary eukaryogenetic symbiosis[39] to the evolution of crown eukaryotes. Also, the lack of genes for biosynthesis of polycyclic terpenoids across the entire archaeal domain supports the conclusion that these compounds originate from bacteria, perhaps to address biophysical challenges associated with fatty acid-based membranes. It is possible that polycyclic terpenoids are incompatible with archaeal isoprenoid membranes, due to constraints on either biosynthetic regulation or membrane biophysics. Alternatively, polycyclic terpenoids may confer little advantage in an isoprenoid-based membrane, as opposed to a membrane composed of fatty acid-based phospholipids.

Unique diterpenoid cyclases from archaea. Phylogenetic analysis (see Methods) revealed three major clades of diterpenoid cyclases (Fig. 1**a**). When this tree is rooted using all non-main group diterpenoid cyclases as an outgroup, the most basal clade (Group I) is polyphyletic and composed primarily of sequences from Chloroflexi and Actinomycetota. Group I contains characterized proteins from *Bradyrhizobium japonicum*, *Streptomyces*, and *Kitasatospora* which synthesize copalyl diphosphate, a precursor to either gibberellins[40] or terpenoid antibiotics[41,42], as well as a *Kitasatospora griseola* protein which synthesizes the precursor to terpentecin, another diterpene antibiotic[43,44]. Group II is composed of fungal diterpenoid cyclases, which synthesize copalyl pyrophosphate[45]. Group III contains bacterial, plant[46], and archaeal sequences, including the Asgard homologs and sequences from Theionarchaeia, Thermoplasmata, and Nitrososphaerota. Within this clade, bacterial sequences are most basal and are paraphyletic to the archaeal group, which itself is paraphyletic to the plant homologs (Fig. 1**b**). Despite the long stem of the plant subclade, its position is stable and removing the plant subclade does not affect the topology of the clade, suggesting that this branching order is not an artifact. The archaeal subgroup also contains sequences from Theionarchaeota and Thermoplasmata, as well as the ten Asgard cyclase sequences (Fig. 1**c**), which are nested among the Thermoplasmata homologs. This pattern could be the result of either vertical or horizontal inheritance. Outside the archaeal subclade, paraphyletic bacterial groups are dominated by Chloroflexi, particularly sequences from metagenome-assembled genomes (MAGs) from sponge symbionts. At the very base of the group is a cluster of Actinomycetota sequences, including the lone characterized enzyme in this group, tuberculosinyl cyclase Rv3377c from *Mycobacterium tuberculosis*.

Asgard cyclases make halimadienyl lipids. To determine the function of the putative Asgard cyclases, we performed *in vitro* experiments with the four Asgard proteins found in our initial search, as well as the closely related diterpenoid cyclase Rv3377c from *Mycobacterium tuberculosis* H37Rv, previously shown to synthesize halimadienyl pyrophosphate[47]. These Asgard cyclases include two homologs from a Hodarchaeales genome, one from Kariarchaeales, and one from Lokiarchaeales (Extended Data Table 1). We incubated purified protein (Extended Data Fig. 9) with geranylgeranyl pyrophosphate (GGPP; **1**) and analyzed reaction products by gas chromatography-mass spectrometry (GC-MS). We found that two of the Asgard cyclases convert GGPP (**1'**) to a cyclized product (Fig. 2). Enzymatic dephosphorylation of the product results in halimadienol (**2**), which we identified by comparison to published GC-MS spectra[47–50]. We could only detect **1** or **2** after incubation with phosphatase, suggesting that the cyclization product is halimadienyl pyrophosphate (**2'**). As previously reported for Rv3377c[51], Asgard cyclases were only functional when expressed in a strain overexpressing the *E. coli* folding chaperones GroEL/ES and trigger factor. The halimadienyl molecules reported here appear to be the exception to the rule that cyclic terpenoids, though widespread in bacteria and eukaryotes, are not found in archaea.

A DxD motif is essential for catalysis by Asgard cyclases. Structural modeling of Asgard cyclases using Alphafold2[52] revealed that the catalytic DxD(D) motif aligns with those of crystallized diterpenoid cyclases (Fig. 3). In this motif, the second aspartic acid residue acts as a general acid, protonating the substrate and initiating a bond-rearranging carbocation cascade, supported by the first aspartic acid[53]. The third aspartic acid in the motif, is not strictly necessary for protonation, as shown by the functionality of Hodarchaeales S146_1 and the *M. tuberculosis* homolog Rv3377c[54] and meroterpenoid cyclase MstE[55].

All Asgard cyclases possess a conserved aminated residue at position H341 (corresponding to sequence position in Rv3377c; PDB ID 6VPT), which forms a hydrogen

bond to the general acid and positions it for proton transfer[56]. These Asgard cyclases also feature a bulky aromatic residue at position W380 to protect the general acid when substrate is not bound. However, neither protein from Hodarchaeales S146 has a bulky aromatic group at position W285, which is implicated in the same protective function[54,57]. Three of the Asgard cyclases retain the Y409 residue (corresponding to sequence position in PtmT2; PDB ID 5BP8) critical in all Type II terpenoid cyclases for regenerating the general acid via water coordination[58]. Interestingly, the Lokiarchaeota MAG CR_4 cyclase, which was not functional in our assay, has a phenylalanine instead of a tyrosine at this position and therefore lacks the hydroxyl group required to hydrogen bond with water. All Asgard proteins have the positively charged lysine K402 (corresponding to sequences position in 5BP8) essential for coordinating the substrate pyrophosphate. However, only the nonfunctional Lokiarchaeal protein has lysine K193, a residue previously identified as crucial for pyrophosphate binding in a different diterpenoid cyclase from a streptomycete[56]. Both cyclases from Hodarchaeales S146 possess the final pyrophosphate-binding residue (R/K350), which is absent in cyclases from the other Asgards. Mg^{2+} coordination facilitated by several acidic residues is thought to contribute to pyrophosphate binding[56]. Hodarchaeales S146-1 has two of these three acidic residues, Lokiarchaeota CR_4 has one, and the other two Asgard enzymes have none. These amino acid substitutions do not fully account for the lack of activity from Hodarchaeales S146-2 in our assay.

Putative halimadiene function in Asgards. Our study revealed diterpenoids cyclases in Asgard genomes including Hodarchaeales and identified halimadienyl pyrophosphate as their lipid product. This represents the first evidence of a terpenoid cyclase that produces polycyclic lipids in archaea. This advances our understanding of lipid metabolism in the Asgards, however the role of halimadiene lipids in these archaea is unclear. Most literature reports of halimadiene lipids have been in flowering plants and liverworts where they may have defensive[59] or allelopathic effects[60]. In animals, halimadienes have been found sparsely in marine invertebrates, including an Antarctic nudibranch for which they are thought to function in defense[61], and *Raspaila* and *Agelas* demosponges, which contain halimadienyl purines[62,63]. Since these sponges and their symbionts remain unsequenced, it is unknown whether the sponge alkaloid is made by microbial symbionts or by the animal itself. However, the halimadienyl cyclase clade in our phylogenetic analyses contains a large number of homologs from bacterial sponge symbionts belonging to the phylum Chloroflexota[64]. Thus, a bacterial source for sponge halimadienyl alkaloids is a possibility.

Mycobacterium tuberculosis is the only prokaryote previously known to synthesize halimanes, secreting halimadienyl-adenosine conjugates constitutively and in significant quantities (over 1% of total lipids)[6]. During the intracellular phase of its life cycle, *M. tuberculosis* is engulfed by host macrophages into an acidifying phagosome that matures into a lysosome. To block this maturation, *M. tuberculosis* secretes 1-halimadienyl-adenosine into the endosomal compartment, where it acts as a Lewis base to neutralize endosomal pH, complementing the effect of vATPase exclusion[65]. This functionality depends on both the lipid moiety, for membrane translocation, and the nucleoside moiety, for basicity[6]. The antacid effect also inactivates endosome hydrolases, leading to a buildup of triglycerides and cholesteryl esters which *M. tuberculosis* uses as source of energy and carbon[7]. This mechanism, which requires 1-halimadienyl-adenosine, facilitates the intracellular persistence of *M. tuberculosis* for decades. Interestingly, halimadienyl lipids are only found in pathogenic species of *Mycobacterium* and production of halimadienyl-adenosine in otherwise non-pathogenic species results in virulence in the mouse lung[66].

In addition, mammalian endosomes infected with *M. tuberculosis* resemble sponge endosomes bearing microbial symbionts. In *M. tuberculosis*, halimadienyl alkaloids are necessary to produce the distended and lipid-rich endosomal compartment characteristic of infection[6,7]. Intracellular symbionts of sponges are also surrounded by expanded lipid-rich host endosomes[67], and have also been shown to persist in host amoebocytes by halting lysosome acidification, although by different mechanisms[68,69]. From the host perspective, the homology of deuterostome macrophages and sponge amoebocytes has been postulated since the nineteenth century[70]. It is possible that sponge symbionts and *M. tuberculosis* share a host cell type and a method of intracellular persistence, albeit with different consequences for the host organisms.

It is possible that Asgards utilize halimadienyl conjugates like *M. tuberculosis*. In *M. tuberculosis*, halimadienyl pyrophosphate is converted to 1-halimadienyl-adenosine through nucleoside addition, thought to be catalyzed by putative adenosyl transferase Rv3378c[71,72]. We identified a homolog of this putative adenosyl transferase in the Hodarchaeales S146 genome, which we have shown also has a functional halimadienyl cyclase (Fig. 2**b**). If this Asgard produces 1-halimadienyl-adenosine, it is possible that halimadienyl nucleosides could affect the maintenance of intracellular symbionts acquired through the endomembrane trafficking system[2,73] by controlling phagolysosome maturation from the host side. While Asgards are not currently known to be endosymbionts or hosts of endosymbionts, they have been shown to form symbiotic interactions with other species[9], and seem to have been involved in the complex endosymbiosis that produced the first eukaryotic cell. Perhaps halimadienyl conjugates play a role in these symbioses.

One of the biggest gaps in our understanding of the origin of the eukaryotic cell is the lipid membrane composition of the Asgards involved, and the role of lipids in symbiotic interactions. Here we searched for polycyclic triterpenoid biosynthesis genes across a genomic catalogue of eukaryotes and archaea, including Asgardarchaeota. This revealed several diterpenoids cyclases in Asgards appear to be ancestral to those in plants. Experimental characterization of these proteins from Hodarchaeales and Kariarchaeaceae confirmed that they cyclize an isoprenoid substrate to produce halimadienyl pyrophosphate. Halimadienyl lipids have only been studied in an intracellular pathogen, where they facilitate persistence inside the host cell. This first description of terpenoid cyclases in archaea advances our understanding of the lipid biochemistry of these fascinating uncultured organisms, and hints at a potential mechanism for mediation of endosymbiotic interactions with other cells in nature.

Data availability. The data supporting the findings in this study are available in the main text, extended data, and supplementary information.

Acknowledgments. We thank Alysha Lee, Andy Garcia, Charles Hu, Maryam Khademian, Minh Tu, Tharika Liyanage, and Maggie Horst for helpful discussions, as well as Daniel Fernandez, Olivia Pattelli, and the Stanford Macromolecular Structure Knowledge Center (MSKC) for advice, enthusiasm, and facilities for protein purification. We also thank Emily Hyde for the technical support on the taxonomy assignments of the archaeal genomic catalogue using the Genome Taxonomy Database Toolkit (GTDB-Tk). H.M. and P.V.W. were supported by NSF Grant EAR-1752564. H.M. was supported by the National Science Foundation (NSF) Graduate Research Fellowship, the Stanford Enhancing Diversity in Graduate Education (EDGE) Fellowship, the Stanford School of Sustainability McGee and Levorsen Graduate Research Grant, and the Stanford MSKC Training Program in Biophysical and Structural Analysis of Biological Macromolecules. Portions of the experiments were performed in the Stanford Geomicrobiology Shared Laboratories Core Facility (RRID:SCR_025000).

Computational analyses were performed on the Sherlock high-performance computing cluster administered by the Stanford Research Computing Center. This work was also supported by the Moore-Simons Project on the Origin of the Eukaryotic Cell, Simons and Moore Foundation 73592LPI to B.J.B. (https://doi.org/10.46714/735925LPI).

Author contributions. H.M.: conceptualization, methodology, investigation, formal analysis, writing – original draft; V.D.A.: conceptualization, methodology, investigation, formal analysis, writing – review & editing; J.J.B.: conceptualization, writing – reviewing & editing; P.V.W.: conceptualization, supervision, writing – reviewing & editing; B.J.B.: conceptualization, supervision, writing – reviewing & editing.

Competing interest declaration. The authors declare no conflict of interest.

Figures.

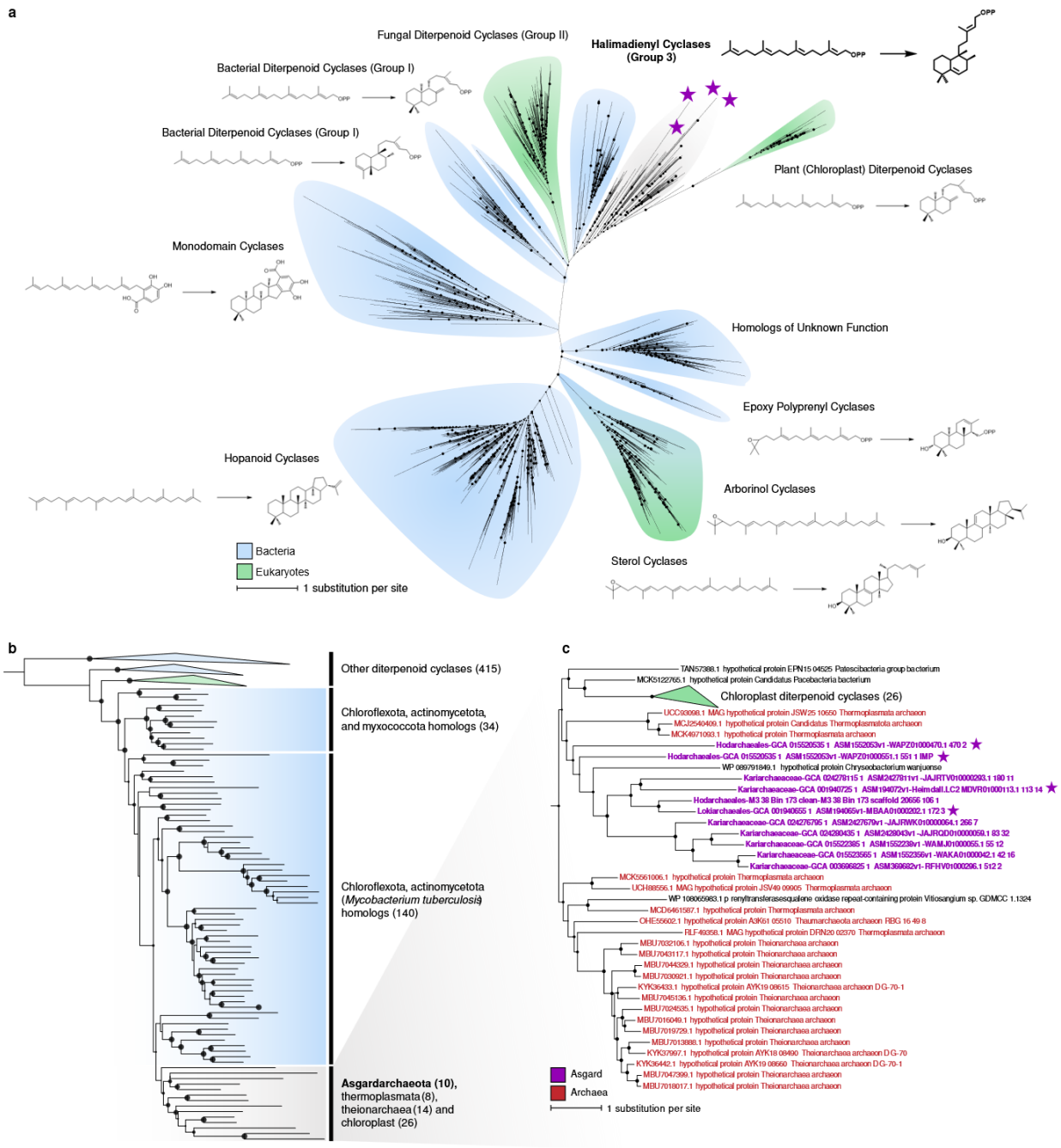

Figure 1: Evolutionary history of Asgard cyclase homologs. (A) Position of Asgard diterpenoid cyclases in the Type II terpenoid cyclase maximum likelihood phylogeny. Phylogeny was estimated under the structure-based EX_EHO model of protein evolution with 5 free rate categories. Reactions performed by characterized enzymes are shown at the edges. Clades containing prokaryotic homologs are shaded in blue, while those containing eukaryotic homologs are green. (B) Diterpenoid cyclase clade, with halimadienyl cyclase subclade expanded. (C) Detail of Asgard-containing subclade. Proteins tested in this study are marked with stars.

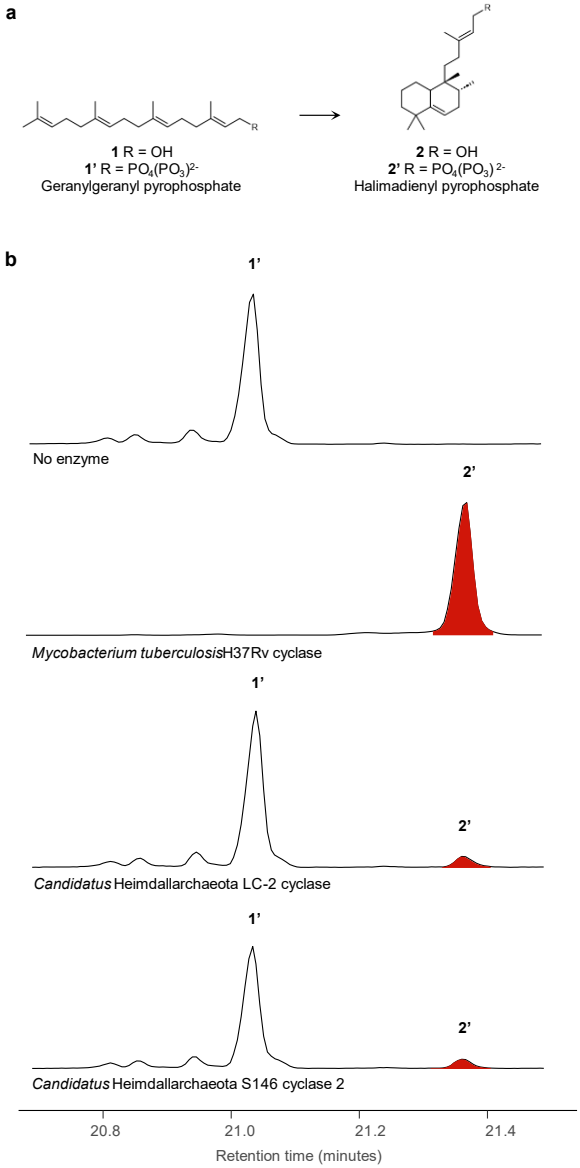

Figure 2: Asgard cyclases catalyze the cyclization of geranylgeranyl pyrophosphate to form halimadienyl lipids. (A) Cyclization of geranylgeranyl pyrophosphate (GGPP) to halimadienyl pyrophosphate. (B) Extracted ion chromatograms (*m/z* 119, 189, 191, 290, 362) of *in vitro* reactions performed with GGPP and cyclases as labeled. Reaction products were dephosphorylated, extracted with hexanes, and trimethylsilated before gas chromatography-mass spectrometry.

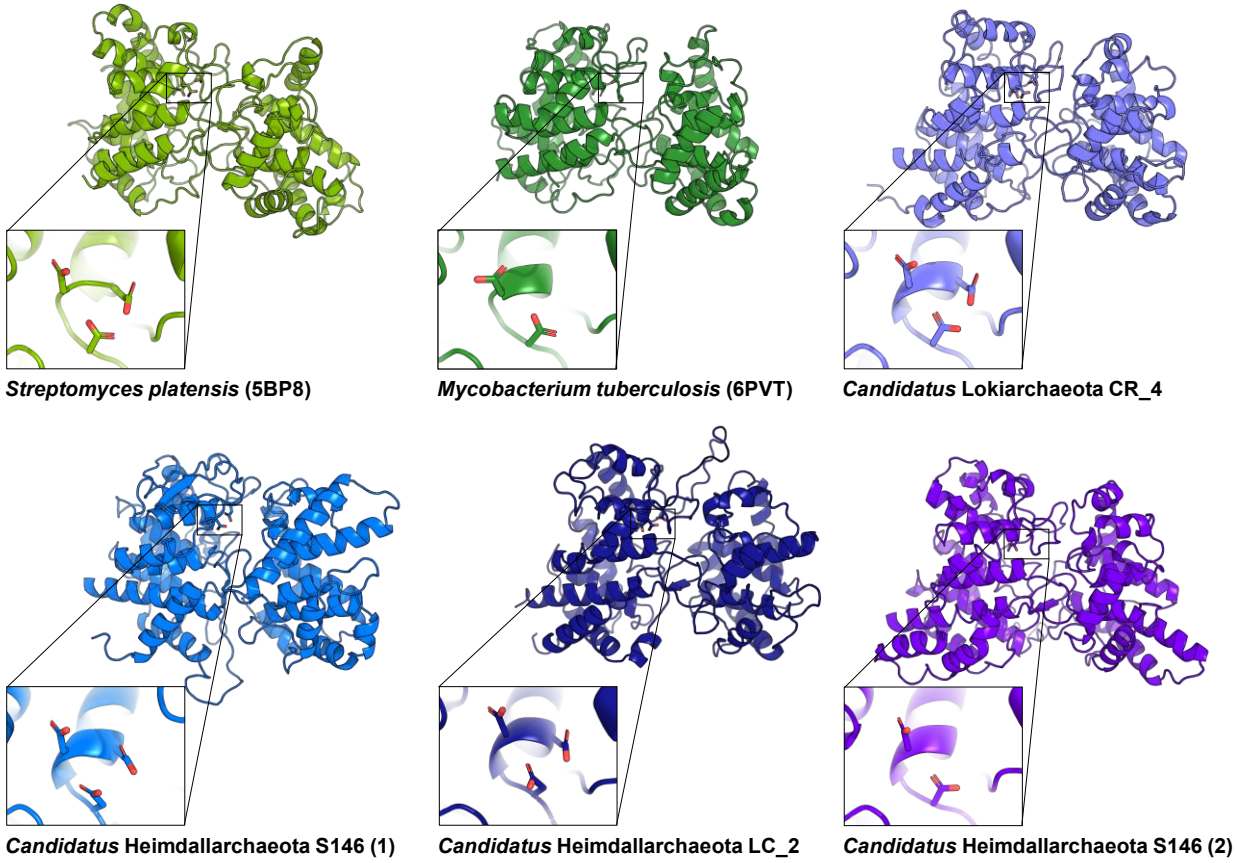

Figure 3: Diterpenoid cyclases from Asgardarchaeota have the general acid motif DxDD typical of Type II cyclases. 5BP8 and 6PVT are previously published crystals of related bacterial cyclases, others are Asgard cyclases, predicted using the Colabfold implementation of Alphafold2[74,75].

References.

1. Eme, L. *et al.* Inference and reconstruction of the heimdallarchaeial ancestry of eukaryotes. *Nature* **618**, 992–999 (2023).
2. Zaremba-Niedzwiedzka, K. *et al.* Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* **541**, 353–358 (2017).
3. Sagan, L. On the origin of mitosing cells. *Journal of Theoretical Biology* **14**, 225-IN6 (1967).
4. Martijn, J., Vosseberg, J., Guy, L., Offre, P. & Ettema, T. J. G. Deep mitochondrial origin outside the sampled alphaproteobacteria. *Nature* **557**, 101–105 (2018).
5. Young, D. C. *et al.* In Vivo Biosynthesis of Terpene Nucleosides Provides Unique Chemical Markers of Mycobacterium tuberculosis Infection. *Chemistry & Biology* **22**, 516–526 (2015).
6. Buter, J. *et al.* Mycobacterium tuberculosis releases an antacid that remodels phagosomes. *Nat Chem Biol* **15**, 889–899 (2019).
7. Bedard, M. *et al.* A terpene nucleoside from M. tuberculosis induces lysosomal lipid storage in foamy macrophages. *J Clin Invest* **133**, (2023).
8. Koga, Y. Early evolution of membrane lipids: How did the lipid divide occur? *Journal of Molecular Evolution* **72**, 274–282 (2011).
9. Imachi, H. *et al.* Isolation of an archaeon at the prokaryote–eukaryote interface. *Nature* **577**, 519–525 (2020).
10. López-García, P. & Moreira, D. The Syntrophy hypothesis for the origin of eukaryotes revisited. *Nat Microbiol* **5**, 655–667 (2020).

11. Spang, A. *et al.* Proposal of the reverse flow model for the origin of the eukaryotic cell based on comparative analyses of Asgard archaeal metabolism. *Nature Microbiology* (2019) doi:10.1038/s41564-019-0406-9.
12. Saharan, O. & Kamat, S. S. Mapping lipid pathways during phagocytosis. *Biochemical Society Transactions* **51**, 1279–1287 (2023).
13. Salvador-Castell, M., Tourte, M. & Oger, P. M. In Search for the Membrane Regulators of Archaea. *International Journal of Molecular Sciences* **20**, 4434 (2019).
14. Mouritsen, O. G. & Zuckermann, M. J. What's so special about cholesterol? *Lipids* **39**, 1101–1113 (2004).
15. Bloch, K. E. Sterol structure and membrane function. *Critical Reviews in Biochemistry and Molecular Biology* **14**, 47–92 (1983).
16. Xu, X. & London, E. The Effect of Sterol Structure on Membrane Lipid Domains Reveals How Cholesterol Can Induce Lipid Domain Formation. *Biochemistry* **39**, 843–849 (2000).
17. Wei, J. H., Yin, X. & Welander, P. V. Sterol synthesis in diverse bacteria. *Frontiers in Microbiology* **7**, 1–19 (2016).
18. Lee, A. K., Wei, J. H. & Welander, P. V. De novo cholesterol biosynthesis in bacteria. *Nat Commun* **14**, 2904 (2023).
19. Sáenz, J. P. *et al.* Hopanoids as functional analogues of cholesterol in bacterial membranes. *Proceedings of the National Academy of Sciences* **112**, 11971–11976 (2015).
20. Welander, P. V. *et al.* Hopanoids Play a Role in Membrane Integrity and pH Homeostasis in Rhodopseudomonas palustris TIE-1 ᴖ †. **191**, 6145–6156 (2009).
21. Takishita, K. *et al.* Microbial Eukaryotes that Lack Sterols. *Journal of Eukaryotic Microbiology* **64**, 897–900 (2017).
22. Belin, B. J. *et al.* Hopanoid lipids: from membranes to plant-bacteria interactions. *Nature Reviews Microbiology* **16**, 304–315 (2018).
23. Santana-Molina, C., Rivas-Marin, E., Rojas, A. M. & Devos, D. P. Origin and Evolution of Polycyclic Triterpene Synthesis. *Molecular Biology and Evolution* **37**, 1925–1941 (2020).
24. Appler, K. E. *et al.* Oxygen metabolism in descendants of the archaeal-eukaryotic ancestor. 2024.07.04.601786 Preprint at https://doi.org/10.1101/2024.07.04.601786 (2024).
25. Desmond, E. & Gribaldo, S. Phylogenomics of Sterol Synthesis: Insights into the Origin, Evolution, and Diversity of a Key Eukaryotic Feature. *Genome Biology and Evolution* **1**, 364–381 (2010).
26. Brocks, J. J. *et al.* Lost world of complex life and the late rise of the eukaryotic crown. *Nature* **618**, 767–773 (2023).
27. Gold, D. A., Caron, A., Fournier, G. P. & Summons, R. E. Paleoproterozoic sterol biosynthesis and the rise of oxygen. *Nature* **543**, 420–423 (2017).
28. Hoshino, Y. & Gaucher, E. A. Evolution of bacterial steroid biosynthesis and its impact on eukaryogenesis. *Proceedings of the National Academy of Sciences* **118**, e2101276118 (2021).
29. Hoshino, Y. & Gaucher, E. A. Impact of steroid biosynthesis on the aerobic adaptation of eukaryotes. *Geobiology* **22**, e12612 (2024).
30. Mistry, J. *et al.* Pfam: The protein families database in 2021. *Nucleic Acids Res* **49**, D412–D419 (2021).
31. Christianson, D. W. Structural and Chemical Biology of Terpenoid Cyclases. *Chem. Rev.* **117**, 11570–11648 (2017).
32. Chen, X., Köllner, T. G., Xiong, W., Wei, G. & Chen, F. Emission and biosynthesis of volatile terpenoids from the plasmodial slime mold Physarum polycephalum. *Beilstein J Org Chem* **15**, 2872–2880 (2019).
33. Kull, D. R. & Pfander, H. Isolation and Structure Elucidation of Carotenoid Glycosides from the Thermoacidophilic Archaea Sulfolobus shibatae. *J. Nat. Prod.* **60**, 371–374 (1997).
34. Britton, G. Structure and properties of carotenoids in relation to function. *The FASEB Journal* **9**, 1551–1558 (1995).
35. Huijbers, M. M. E., Montersino, S., Westphal, A. H., Tischler, D. & van Berkel, W. J. H. Flavin dependent monooxygenases. *Archives of Biochemistry and Biophysics* **544**, 2–17 (2014).

36. Pollier, J. *et al.* A widespread alternative squalene epoxidase participates in eukaryote steroid biosynthesis. *Nature Microbiology* **4**, 226–233 (2019).
37. Sato, T. & Hoshino, T. Functional Analysis of the DXDDTA Motif in Squalene-Hopene Cyclase by Site-directed Mutagenesis Experiments: Initiation Site of the Polycyclization Reaction and Stabilization Site of the Carbocation Intermediate of the Initially Cyclized A-Ring. *Bioscience, Biotechnology, and Biochemistry* **63**, 2189–2198 (1999).
38. Richter, D. J. *et al.* EukProt: A database of genome-scale predicted proteins across the diversity of eukaryotes. *Peer Community Journal* **2**, (2022).
39. Williams, T. A. *et al.* Phylogenetic reconciliation: making the most of genomes to understand microbial ecology and evolution. *The ISME Journal* **18**, wrae129 (2024).
40. Morrone, D. *et al.* Gibberellin biosynthesis in bacteria: Separate ent-copalyl diphosphate and ent-kaurene synthases in Bradyrhizobium japonicum. *FEBS Letters* **583**, 475–480 (2009).
41. Jayasuriya, H. *et al.* Isolation and Structure of Platencin: A FabH and FabF Dual Inhibitor with Potent Broad-Spectrum Antibiotic Activity. *Angewandte Chemie International Edition* **46**, 4684–4688 (2007).
42. Smanski, M. J. *et al.* Dedicated ent-kaurene and ent-atiserene synthases for platensimycin and platencin biosynthesis. *Proceedings of the National Academy of Sciences* **108**, 13498–13503 (2011).
43. Dairi, T. *et al.* Eubacterial Diterpene Cyclase Genes Essential for Production of the Isoprenoid Antibiotic Terpentecin. *Journal of Bacteriology* **183**, 6085–6094 (2001).
44. Tamamura, T. *et al.* ISOLATION AND CHARACTERIZATION OF TERPENTECIN, A NEW ANTITUMOR ANTIBIOTIC. *J. Antibiot.* **38**, 1664–1669 (1985).
45. Quin, M. B., Flynn, C. M. & Schmidt-Dannert, C. Traversing the fungal terpenome. *Natural Product Reports* **31**, 1449–1473 (2014).
46. Jia, Q. *et al.* Origin and early evolution of the plant terpene synthase family. *Proceedings of the National Academy of Sciences* **119**, e2100361119 (2022).
47. Maugel, N., Mann, F. M., Hillwig, M. L., Peters, R. J. & Snider, B. B. Synthesis of (±)-Nosyberkol (Isotuberculosinol, Revised Structure of Edaxadiene) and (±)-Tuberculosinol. *Org. Lett.* **12**, 2626–2629 (2010).
48. Nakano, C., Okamura, T., Sato, T., Dairi, T. & Hoshino, T. Mycobacterium tuberculosis H37Rv3377c encodes the diterpene cyclase for producing the halimane skeleton. *Chem. Commun.* 1016–1018 (2005) doi:10.1039/B415346D.
49. Prach, L., Kirby, J., Keasling, J. D. & Alber, T. Diterpene production in Mycobacterium tuberculosis. *The FEBS Journal* **277**, 3588–3595 (2010).
50. Mann, F. M. *et al.* Characterization and Inhibition of a Class II Diterpene Cyclase from Mycobacterium tuberculosis: IMPLICATIONS FOR TUBERCULOSIS *. *Journal of Biological Chemistry* **284**, 23574–23579 (2009).
51. Nakano, C. & Hoshino, T. Characterization of the Rv3377c Gene Product, a Type-B Diterpene Cyclase, from the Mycobacterium tuberculosis H37 Genome. *ChemBioChem* **10**, 2060–2071 (2009).
52. Bordin, N. *et al.* AlphaFold2 reveals commonalities and novelties in protein structure space for 21 model organisms. *Commun Biol* **6**, 1–12 (2023).
53. Hoshino, T. & Sato, T. Squalene – hopene cyclase: catalytic mechanism and substrate recognition. *Chemical Communications* **0**, 291–301 (2002).
54. Zhang, Y. *et al.* Crystal Structure and Mechanistic Molecular Modeling Studies of Mycobacterium tuberculosis Diterpene Cyclase Rv3377c. *Biochemistry* **59**, 4507–4515 (2020).
55. Moosmann, P. *et al.* Cyanobacterial ent-Sterol-Like Natural Products from a Deviated Ubiquinone Pathway. *Angewandte Chemie - International Edition* **56**, 4987–4990 (2017).
56. Rudolf, J. D. *et al.* Structure of the ent-Copalyl Diphosphate Synthase PtmT2 from Streptomyces platensis CB00739, a Bacterial Type II Diterpene Synthase. *J. Am. Chem. Soc.* **138**, 10905–10915 (2016).
57. Köksal, M., Hu, H., Coates, R. M., Peters, R. J. & Christianson, D. W. Structure and mechanism of the diterpene cyclase ent-copalyl diphosphate synthase. *Nat Chem Biol* **7**, 431–433 (2011).

58. Sato, T. & Hoshino, T. Catalytic Function of the Residues of Phenylalanine and Tyrosine Conserved in Squalene-Hopene Cyclases. *Bioscience, Biotechnology, and Biochemistry* **65**, 2233–2242 (2001).
59. Rubin, D. *et al.* Halimane Diterpenes in the Alpine Daisy Celmisia viscosa: Absolute Configuration, 2,6-Dideoxyhexopyran-3-ulosides, Conformational Flexibility, and Intraspecific Variation. *J. Nat. Prod.* **85**, 1893–1903 (2022).
60. del Corral, S., Cuffini, S. L., Cardoso, S. G., Bortoluzzid, A. J. & Palacios, S. M. Phytotoxic halimanes isolated from *Baccharis salicifolia* (Ruiz & Pad.) Pers. *Phytochemistry Letters* **5**, 280–283 (2012).
61. Iken, K., Avila, C., Fontana, A. & Gavagnin, M. Chemical ecology and origin of defensive compounds in the Antarctic nudibranch Austrodoris kerguelenensis (Opisthobranchia: Gastropoda). *Marine Biology* **141**, 101–109 (2002).
62. Cullen, E. & Devlin, J. P. Agelasine: A Novel Quaternary 9-Methyladenine from the Sponge Agelasdispar. *Can. J. Chem.* **53**, 1690–1691 (1975).
63. Rudi, A., Aknin, M., Gaydou, E. & Kashman, Y. Asmarines I, J, and K and Nosyberkol: Four New Compounds from the Marine Sponge Raspailia sp. *J. Nat. Prod.* **67**, 1932–1935 (2004).
64. Bayer, K., Jahn, M. T., Slaby, B. M., Moitinho-Silva, L. & Hentschel, U. Marine Sponges as Chloroflexi Hot Spots: Genomic Insights and High-Resolution Visualization of an Abundant and Diverse Symbiotic Clade. *mSystems* **3**, 10.1128/msystems.00150-18 (2018).
65. Sturgill-Koszycki, S. *et al.* Lack of Acidification in Mycobacterium Phagosomes Produced by Exclusion of the Vesicular Proton-ATPase. *Science* **263**, 678–681 (1994).
66. Ghanem, M. *et al.* Heterologous Production of 1-Tuberculosinyladenosine in Mycobacterium kansasii Models Pathoevolution towards the Transcellular Lifestyle of Mycobacterium tuberculosis. *mBio* **11**, 10.1128/mbio.02645-20 (2020).
67. Ereskovsky, A. V., Gonobobleva, E. & Vishnyakov, A. Morphological evidence for vertical transmission of symbiotic bacteria in the viviparous sponge Halisarca dujardini Johnston (Porifera, Demospongiae, Halisarcida). *Marine Biology* **146**, 869–875 (2005).
68. Nguyen, M., Liu, M. & Thomas, T. Ankyrin-repeat proteins from sponge symbionts modulate amoebal phagocytosis. *Molecular Ecology* **23**, 1635–1645 (2014).
69. Reynolds, D. & Thomas, T. Evolution and function of eukaryotic-like proteins from sponge symbionts. *Molecular Ecology* **25**, 5242–5253 (2016).
70. Metchnikoff, E. *Embryologische Studien an Medusen: Ein Beitrag zur Genealogie der Primitiv-organe*. (A. Holder, 1886).
71. Chan, H.-C. *et al.* Structure and inhibition of tuberculosinol synthase and decaprenyl diphosphate synthase from Mycobacterium tuberculosis. *J Am Chem Soc* **136**, 2892–2896 (2014).
72. Layre, E. *et al.* Molecular profiling of Mycobacterium tuberculosis identifies tuberculosinyl nucleoside products of the virulence-associated enzyme Rv3378c. *Proceedings of the National Academy of Sciences* **111**, 2978–2983 (2014).
73. Makarova, K. S. *et al.* Diversity, origin, and evolution of the ESCRT systems. *mBio* **15**, e00335-24 (2024).
74. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
75. Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nat Methods* **19**, 679–682 (2022).

Methods for McShea & De Anda et al., Archaeal lineages related to eukaryotes encode functional diterpenoid cyclases.

Archaeal database. We obtained a total of 3,685 publicly available archaeal genomes from NCBI in September 2022 (Supplementary Table 1). Taxonomic affiliations were assigned using both the archaeal Genome Taxonomy Database GTDB-Tk v2.3.2[1] and the conventional high-ranking superphylum names (Asgard, DPANN, Euryarchaeota, TACK). Specifically, 668 TACK, 1,406 Euryarchaeota, 644 DPANN, and 867 Asgard genomes described in Appler et al. (2024)[2].

Biosynthesis pathways. HMM models were downloaded from Pfam 37.0. Sequences were searched using MEBS v1.2[3]. Searches were carried out using the gathering thresholds provided for each model (option -cut_ga) also implemented in the default options in MEBS.

Terpenoid cyclase database curation. We curated a Type II terpenoid cyclases database to analyze the phylogenetic position and evolutionary history of Asgard cyclases. Briefly, we first used sequences from each known clade of Type II cyclases to query public databases: Joint Genome Institute Integrated Microbial Genomes & Microbiomes database (JGI IMG; https://img.jgi.doe.gov), the National Center for Biological Informatics (NCBI) nonredundant (nr), clustered nonredundant (clustered nr), sequence read archive (sra) and whole-genome shotgun contigs (wgs) databases (https://www.ncbi.nlm.nih.gov/genbank), and the European Molecular Biology Laboratory – European Bioinformatics Institute (EMBL-EBI) MGnify database (https://www.ebi.ac.uk/metagenomics). JGI IMG and NCBI were queried via BLAST [4], with an expect threshold of 0.05 and a word size of 5 for NCBI, and an e-value cutoff of 1e-5 for JGI. MGnify was queried via phmmer [5] with an e-value cutoff of 1e-5. We also downloaded all hits for Pfams PF13243 (squalene-hopene cyclase C-terminal domain) and PF13249 (squalene-hopene cyclase N-terminal domain) from EBI (https://www.ebi.ac.uk/interpro/entry/pfam/#table). These searches returned 44,209 unique sequences, which, after a length filter of 350 (the length of the smallest known Type II cyclase[6]) to 900 (the length of the largest known Type II domain-bearing cyclase[7]) and subsetting by CD-hit (command cd-hit -i in.fasta -o out.fasta -c 0.90 -n 5 -M 6000 -d 0 -T 8) [8], yielded a database of 13,680 sequences. Finally, after initial phylogenies were estimated and the split between diterpenoid and triterpenoid cyclases was apparent, we created a custom hidden Markov models (HMM) using domain-wise (β and γ) alignments of the diterpenoid cyclase group. The diterpenoid β domain model was very similar to the triterpenoid β domain model as measured by overlapping HMMER[9] hits on the whole database. In contrast, the diterpenoid γ domain model had no reciprocal hits with the triterpenoid γ domain model. These diterpenoid cyclase-specific models were used to search our databases again, returning 6 additional diterpenoid cyclase homologs from Asgards.

Phylogenetic estimation and analysis of terpenoid cyclases. Cyclase sequences were aligned using the MAFFT linsi algorithm[10] implemented in Magus[11]. The alignment was trimmed using trimAl[12] with a gap threshold of 0.1 (command trimal -in in.aln -out out.aln -gt 0.1), and further subsetted with a maximum sequence identity of 50% (command trimal -in in.aln -out out.aln -maxidentity 0.5). All phylogenies were estimated with IQ-TREE v2.2.2.6[13–15]. The best-fit model of amino acid sequence evolution, EX_EHO + R5, was determined using multiple ModelFinder[16] runs to completely explore model space, and branch support was calculated with ≤ 10,000 ultrafast

bootstrap approximations[17]. Tree visualization was performed in Dendroscope 3.8.10[18] and iTOL 6.8.2[19].

Heterologous gene expression and protein purification. Four Asgard cyclase gene sequences, along with positive control Rv3377c from *M. tuberculosis*, were codon-optimized for expression in *Escherichia coli* and synthesized by Twist Bioscience (South San Francisco, CA) in a pET-28a(+) plasmid. Plasmids sequences were confirmed by Plasmidsaurus (Eugene, OR) and plasmids were transformed by heat shock into competent NiCo21(DE3) cells (New England Biolabs) for expression, along with pACYC for overexpression of the GroEL/ES chaperone system and Trigger Factor (Addgene #83923). Media was supplemented with kanamycin (15 µg/mL) and chloramphenicol (20 µg/mL) to maintain selection for both plasmids. Plasmids and expression strains are described in detail in Extended Data Tables 2 and 3, respectively.

Expression strains were cultured in biological triplicate in 1 L terrific broth (TB) in a 2 L flask. Cultures were grown at 37 °C while shaking at 225 rpm to an OD_{600} of ~0.6, cooled on ice, induced with 500 µM isopropyl β-D-1-thiogalactopyranoside (IPTG), then grown an additional ~20 h at 16 °C shaking at 225 rpm. Cells were harvested by centrifugation at 10,000 × g. Pellets were stored at −20 °C or immediately resuspended in 30 mL lysis buffer (50 mM Tris, pH 7.5, 2 mM MgCl2, 0.5 mM tris(2-carboxyethyl)phosphine, 10% glycerol, 0.1% Triton X-100, 0.3 M NaCl, XX mM lysozyme, 0.3 mL Xpert Protease Inhibitor Cocktail; modified from [20]) for sonication. Disruption by sonication was achieved with a 6.4 mm microtip at 50% amplitude on a 30 s on, 30 s off cycle for 5 minutes of on time. Lysates were cleared by centrifugation at 15,000 × g.

Protein was purified by immobilized metal affinity chromatography with a 5 mL HisTrap FF Nickel column (Cytiva Life Sciences) on an Akta Pure (General Electric) fast protein liquid chromatograph (FPLC). After equilibration with loading buffer (50mM Tris, pH 7.5; 0.1 M NaCl), the sample was loaded at 0.5 mL/min, and elution buffer (50mM Tris, pH 7.5; 0.1 M NaCl, 0.5 M imidazole) concentration was increased to 100% at 4%/min. Protein was further purified by size-exclusion chromatography, and purity was assessed via SDS-PAGE (Extended Data Fig. 7).

***In vitro* enzymatic activity assay and workup.** Enzymatic reactions were prepared by combining 20 µg enzyme and 50 µM geranylgeranyl pyrophosphate (GGPP) (Sigma-Aldrich) in 2.5 mL of buffer (50 mM Tris, pH 7.5, 0.1 mM MgCl2, 0.1% Tween-80, after [21]). The reactions were incubated at 30 °C for 18 h. Reaction products were dephosphorylated with the addition of 50 µL Quick CIP (and 250 µL rCutsmart buffer, both from New England Biolabs) and incubation at 37 °C for 20 m. The reactions were then extracted three times with an equal volume of hexanes, with centrifugation at 2,800 × g before each extraction to facilitate separation of polar and nonpolar phases. Extracts were dried under a gentle stream of N_2 gas and subsequently derivatized to trimethylsilyl ethers by the addition of 50 µL pyridine and 50 µL *N,O*-Bis(trimethylsilyl) trifluoroacetamide for 1 h at 70 °C prior to GC-MS analysis.

Product determination by GC-MS. Lipid extracts were separated on an Agilent 7890B Series gas chromatograph (GC) with helium as the carrier gas at a constant flow of 1.1 ml/min. The GC program ran as follows: 50 °C for 3 minutes, ramp 14 °C /min to 300 °C, hold 3 min, ramp 10 °C /min to 330 °C, and hold for 5 min (modified from [22]). Separation was achieved using two tandem DB17-HT columns (30 m × 0.25 mm i.d. × 0.15 µm film thickness). 2 µL of sample were injected into a Gerstel-programmable temperature vaporization (PTV) injector operated in splitless mode

at 250 °C. The GC was coupled to a 5977A Series mass selective detector (MSD) with the source at 320 °C and operated in electron ionization (EI) mode scanning from 90 to 600 Da in 0.3 s. Compounds were identified by comparing mass spectra to previously published spectra.

Methods References

1. Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk v2: memory friendly classification with the genome taxonomy database. *Bioinformatics* **38**, 5315–5316 (2022).
2. Appler, K. E. *et al.* Oxygen metabolism in descendants of the archaeal-eukaryotic ancestor. 2024.07.04.601786 Preprint at https://doi.org/10.1101/2024.07.04.601786 (2024).
3. De Anda, V. *et al.* MEBS, a software platform to evaluate large (meta)genomic collections according to their metabolic machinery: unraveling the sulfur cycle. *GigaScience* **6**, gix096 (2017).
4. Camacho, C. & Madden, T. BLAST+ Release Notes. in *BLAST® Help [Internet]* (National Center for Biotechnology Information (US), 2023).
5. Prakash, A., Jeffryes, M., Bateman, A. & Finn, R. D. The HMMER Web Server for Protein Sequence Similarity Search. *Current Protocols in Bioinformatics* **60**, 3.15.1-3.15.23 (2017).
6. Moosmann, P. *et al.* A monodomain class II terpene cyclase assembles complex isoprenoid scaffolds. *Nat. Chem.* **12**, 968–972 (2020).
7. Mitsuhashi, T., Okada, M. & Abe, I. Identification of Chimeric αβγ Diterpene Synthases Possessing both Type II Terpene Cyclase and Prenyltransferase Activities. *ChemBioChem* **18**, 2104–2109 (2017).
8. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).
9. Eddy, S. R. Accelerated Profile HMM Searches. *PLOS Computational Biology* **7**, e1002195 (2011).
10. Rozewicki, J., Li, S., Amada, K. M., Standley, D. M. & Katoh, K. MAFFT-DASH: integrated protein sequence and structural alignment. *Nucleic Acids Research* **47**, W5–W10 (2019).
11. Smirnov, V. & Warnow, T. MAGUS: Multiple sequence Alignment using Graph clUStering. *Bioinformatics* **37**, 1666–1672 (2021).
12. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
13. Bui, M. Q. *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution* **37**, 1530–1534 (2020).
14. Dang, C. C. *et al.* nQMaker: Estimating Time Nonreversible Amino Acid Substitution Models. *Systematic Biology* syac007 (2022) doi:10.1093/sysbio/syac007.
15. Nguyen, L. T., Schmidt, H. A., Von Haeseler, A. & Minh, B. Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* **32**, 268–274 (2015).
16. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermiin, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* **14**, 587–589 (2017).

17. Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution* **35**, 518–522 (2018).
18. Huson, D. H. & Scornavacca, C. Dendroscope 3: An Interactive Tool for Rooted Phylogenetic Trees and Networks. *Systematic Biology* **61**, 1061–1067 (2012).
19. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Research* **49**, W293–W296 (2021).
20. Zhang, Y. *et al.* Crystal Structure and Mechanistic Molecular Modeling Studies of Mycobacterium tuberculosis Diterpene Cyclase Rv3377c. *Biochemistry* **59**, 4507–4515 (2020).
21. Nakano, C. & Hoshino, T. Characterization of the Rv3377c Gene Product, a Type-B Diterpene Cyclase, from the Mycobacterium tuberculosis H37 Genome. *ChemBioChem* **10**, 2060–2071 (2009).
22. Zhou, K. *et al.* Functional characterization of wheat ent-kaurene(-like) synthases indicates continuing evolution of labdane-related diterpenoid metabolism in the cereals. *Phytochemistry* **84**, 47–55 (2012).