Supplementary Information for

Genetics re-establish the utility of 2-methylhopanes as cyanobacterial biomarkers before 750 million years ago

Yosuke Hoshino*, Benjamin J. Nettersheim*, David A. Gold, Christian Hallmann, Galina Vinnichenko, Lennart M. van Maldegem, Caleb Bishop, Jochen J. Brocks, Eric A. Gaucher

*Corresponding author. Email: yhoshino@gfz-potsdam.de; bnettersheim@marum.de

This PDF file includes:

Supplementary Texts 1–4
Supplementary Information references
Supplementary Figs. 1–8
Supplementary Tables 1–7 captions
Supplementary Table 7 references
Supplementary Data 1–2 captions

Other Supplementary Information for this manuscript include the following:

Supplementary Tables 1–7 (xlsx files)
Supplementary Data 1–2 (txt files)

Supplementary Text 1: Distribution of SC and HpnP in bacteria

Although squalene cyclase (SC) and HpnP are distributed in 31 and 12 bacterial phyla, respectively, the majority of HpnP-containing phyla contain only a handful of species that possess HpnP and these atypical species likely acquired HpnP via horizontal gene transfer (HGT). Hence, HpnP is essentially concentrated in three phyla: Rokubacteria, Alphaproteobacteria and Cyanobacteria (Fig. 1A). Yet also in these three phyla, the gene is not universally distributed. Currently, no complete genome is available for Rokubacteria, but 117 out of 140 species that have draft genome data were found to possess SC (84%), and HpnP was present in 50 species among those SC-containing species (42%) (Supplementary Table 1). Six rokubacterial species were found to possess only HpnP (Supplementary Table 1), but it is not clear if the apparent lack of SC is due to the incomplete genome sequencing status or alternatively HpnP in those species have a function that is irrelevant to hopanoid biosynthesis. In Alphaproteobacteria, SC is widespread (but not ubiquitous) in the phylum, while HpnP is mostly constrained to a single order (Hyphomicrobiales; former Rhizobiales) (Fig. 1B), in particular to three out of 17 families in the order, as was previously observed (Beijerinckiaceae, Bradyrhizobiaceae and Methylobacteriaceae; Supplementary Tables 2 and 3)[1]. The majority of species in these three families contain both SC and HpnP. Hyphomicrobiales represents a late-branching taxon in the phylum[2] and thus a late origin of HpnP in Alphaproteobacteria via HGT is inferred. Among 328 species within Hyphomicrobiales for which complete genome data is available, only 93 species (28%) were found to contain SC; the majority of those SC-containing species (71 species; 76%) harbor both SC and HpnP (Supplementary Table 2). Although the inclusion of draft genome data increases the number

of species that contain SC and HpnP, the distribution of HpnP is still mostly limited to the three families described above (Supplementary Table 2).

In Cyanobacteria, HpnP is found in most taxonomic lineages (Fig. 1B), although the distribution of SC and HpnP is not universal even in this phylum. Among 171 species that have complete genome data, 70 species (41%) contain SC (tables S4 and S5). Approximately half of SC-containing Cyanobacteria (32 species; 46%) were found to additionally possess HpnP (Supplementary Tables 4–5). The distribution of SC and HpnP varies between different clades. For instance, the genus Nostoc contains 19 species that have complete genome data. While SC is found in all of 19 species, HpnP is found in 13 species (68%). In contrast, in the genus Leptolyngbya, three out of six species that have complete genome data have both SC and HpnP (50%), while the other three lack both genes. If only complete genomes are considered, only 19 out of 38 cyanobacterial families (50%) harbor SC-containing species and nine families (24%) harbor species that have both SC and HpnP. However, if draft genomes are included, these values rise to 30 and 27 families (79% and 71%), respectively. Hence, the occurrence of SC and in particular HpnP seems to be underestimated, if based only on complete genomes, due to the sporadic distribution of these two genes. The completion of more cyanobacterial genome sequencing would likely further increase the number of HpnP-containing species in Cyanobacteria.

Supplementary Text 2: Triterpenoid-related proteins in myxobacteria

Instead of SC, several species of myxobacteria that have a HpnP homolog have oxidosqualene cyclase (OSC), which is evolutionarily and functionally related to SC and is involved in steroid biosynthesis[3]. However, the distribution of HpnP homologs is punctate in myxobacteria and has

no correlations to the distribution of OSC in the lineage. Additionally, no 2-methylated steroids have been observed in myxobacteria thus far. Hence, it is not likely that the HpnP homolog in myxobacteria is involved in steroid biosynthesis and the function of myxobacterial HpnP homologs remains unknown.

Supplementary Text 3: Syngeneity assessment of Paleoproterozoic 2-methylhopanes from the McArthur Basin, Northern Australia (GR7 and LV09001 cores)

In our current study, the syngeneity of 2-methylhopanes was examined for two Paleoproterozoic drill cores; GR7 and LV09001. Drill core GR7 has long been utilized for paleoecological reconstructions of the Barney Creek Formation (BCF)[4]. A more recently drilled core LV09001 was obtained from the southern McArthur Basin (east of the Emu Fault). In contrast to GR7, where only the upper section of the BCF is thermally well-preserved, LV09001 is thermally well-preserved throughout the entire section of the BCF as well as underlying Teena Dolostone and overlying Reward Dolostone & Lynott Formation. Therefore, LV09001 allows us to extend 2-MHI analysis to a wider stratigraphy and a longer geological time scale. The upper section of the BCF from the GR7 sample exhibits the highest TOC (up to 8%) and the lowest thermal maturity (T_{max} ca. 435-445 °C, HI ca. 500-800 mg hydrocarbons/g organic carbon). Thermal maturity of the LV09001 samples ranges from immature to early mature (T_{max} ca. 420–439 °C).

Drill core GR7

Drill core GR7 has different thermal maturities between the upper and the lower sections. Only the upper ~200 m section of the core appears thermally suitable for trace biomarker analyses, but

samples from both sections were analyzed for comparison. Most samples were analyzed by the bulk analysis, while the 45.4 m sample was analyzed by the slice-extraction experiment and the 685.8 m & 869.6 m samples were analyzed by interior-exterior experiments (Supplementary Table 7). The total hopane concentration substantially varies with depth and broadly displays an anticorrelation to the 2-MHI. Hopane-rich samples generally have low 2-MHI values (mostly <3%), except for the 218.1 m sample, while hopane-poor samples generally have higher 2-MHI values (up to 16.7%). The spatial distribution of biomarkers within rock specimens (*i.e.*, hopanes and steranes) for the 683.5 m and 869.9 m samples that yield the highest 2-MHI values of 16.7 and 13.2%, respectively, suggest that biomarkers are largely restricted to the sample exterior, while the sample interior is devoid of detectable biomarkers. These observations suggest that the elevated 2-MHI values for hopane-poor bulk samples are likely to reflect contamination overprint that was probably introduced during drilling, cutting and storage. 2-Methylhopanes in the contaminants generally exhibit high 2-MHI values. In contrast, hopane-rich samples, which are less susceptible to surficial contamination and thus may retain syngenetic (methyl)hopane signatures, exhibit low 2-MHI values. In the present study, GR7 samples that were only bulk analyzed were not included in our dataset to minimize uncertainty.

Drill core LV09001

The molecular inventory of core samples from hole LV09001 allows for the first time to investigate the methylhopane distribution in the lower (deeper) section of the BCF as well as underlying Teena Dolostone and overlying Reward Dolostone and Lynott Formation. All samples were separated into interior and exterior portions. The level of surficial contamination in the core was assessed by interior/exterior comparisons on seven samples. Steranes are absent in all interior samples and

restricted only to some exteriors. Hopanes are abundant in both the exterior and the interior of samples for all formations. The abundance of hopanes is nearly identical between the exterior and the interior of all analyzed samples and for this drill hole there is no sign of substantial surficial contamination. These observations indicate that the contamination level of LV09001 drill core is very low, unlike (parts of) the GR7 core, and its impact on our biomarker analyses is negligible with nearly all of the detected hopanes being inferred to be syngenetic. 2-MHI values are generally low throughout the core (Supplementary Table 6), regardless of total hopane abundance. The low 2-MHI values for non-contaminated LV09001 samples are consistent with those for hopane-rich GR7 samples that are geographically separated from the LV09001 samples within the McArthur Basin. In the present study, all of the LV09001 samples are included in our dataset.

Possibility of diagenetic 2-methylhopane production

Apart from biological 2-methylhopanoid production, we addressed the possibility that diagenetically-mediated reactions may lead to the methylation of hopanoids, thereby mimicking biological signatures. For example, 2- and 3-methylsteroids that have been observed in geological records for decades[5] were recently shown to readily form through an abiogenic alkylation process affecting double bonds and/or functional groups in non-alkylated steroid precursors[6]. In the present study, pyrolysis experiments using diplopterol (a major bacterial hopanoid) as well as cholesterol and its saturated equivalent cholestanol (control compounds) were conducted to test the possibility of an abiogenic origin for 2- and 3-methylhopanoids. Whereas cholesterol and cholestanol, which possess a functionalized A-ring, underwent the expected C-2 and C-3 methylations, diplopterol did not generate 2- or 3-methylhopanes (Supplementary Figs. 7–8). Our results suggest that

bacterial hopanoids, which are not known to possess a functional group in the A-ring, are unlikely to be subjected to abiogenic C-2 and C-3 methylations.

Supplementary Text 4: Vitamin B_{12} (cobalamin) dependency in algae

The vitamin B_{12} dependency of plants and algae is complex. For instance, land plants do not require vitamin B_{12} because they can produce methionine using the vitamin B_{12}-independent form of methionine synthase (MetE), while algae have MetE and/or the vitamin B_{12}-dependent form (MetH), depending on species[7]. Red algae retain MetE and additionally MetH in some species, but the majority of green algae have only MetH and thus are vitamin B_{12}-dependent. A selective advantage of having MetH instead of MetE is a substantially higher efficiency of methionine biosynthesis by MetH (x50 times)[8,9]. MetE and MetH do not share a sequence similarity and are inferred to have evolved independently[10]. However, both enzymes are taxonomically widespread in Archaeplastida (land plants + algae) and phylogenetic analyses suggest that the common ancestors of red algae, green algae and the entire Archaeplastida clade had both forms of enzymes, except only for the common ancestor of land plants that likely had only MetE[7]. Therefore, the MetE or MetH gene was lost in individual algal lineages during algal evolution.

Vitamin B_{12}-dependent green algae rely on the uptake of this nutrient produced by symbiotic bacteria or archaea since no eukaryotes can biosynthesize vitamin B_{12} *de novo*[11]. There are two known pathways to biosynthesize the corrin ring of vitamin B_{12} – oxygen-independent and oxygen-dependent pathways[12]. The oxygen-independent pathway is mainly present in anaerobic lineages such as Clostridia (Firmicutes) and Desulfobacteria (Deltaproteobacteria), but is also found in

some aerobic lineages such as Cyanobacteria and Thaumarchaeota[12]. The oxygen-independent pathway may have been important for Cyanobacteria to endure oxygen stress at the onset of oxygenic photosynthesis[13]. In contrast, the oxygen-dependent pathway is mostly confined to some late-branching taxa of Alpha- and Gammaproteobacteria[12]. However, the major contributors of vitamin B_{12} to algae in modern environments are those bacteria that utilize the oxygen-dependent pathway, while vitamin B_{12} production in deeper waters is largely contributed by Thaumarchaeota[13]. Cyanobacteria produce pseudocobalamin, instead of vitamin B_{12}, because Cyanobacteria utilize a different ligand that attaches to the corrin ring. Thus, Cyanobacteria cannot directly contribute to the vitamin B_{12} pool, although a small fraction of algae are known to possess enzymes to convert pseudocobalamin to vitamin B_{12}[8].

The wide occurrence of vitamin B_{12} dependency in green algae reflects an ancestral origin of the algal symbiosis with vitamin B_{12}-producing microbes and the accompanied MetE gene loss. However, the expansion of vitamin B_{12}-producing and 2-methylhopanoid-producing Alphaproteobacteria in the Ediacaran remains speculative because a symbiotic relationship between marine algae and HpnP-containing species has not been directly observed in modern environments thus far. For instance, Hyphomicrobiales is more frequently associated with terrestrial or freshwater environments than with oceans (e.g. lichens)[14]. However, metagenomic surveys predict the presence of both Hyphomicrobiales and alphaproteobacterial HpnP in marine settings[15,16]. Also, some Hyphomicrobiales species are in fact affiliated with marine algae[17,18]. Studies that focus on the taxonomic diversity of the modern marine algal-bacterial relationship would be beneficial to shed more light on the co-evolution of vitamin B_{12}-auxotrophic algae and bacterial donors[19].

Supplementary Information references

1 Ricci, J. N., Michel, A. J. & Newman, D. K. Phylogenetic analysis of HpnP reveals the origin of 2-methylhopanoid production in Alphaproteobacteria. *Geobiology* **13**, 267-277 (2015).

2 Muñoz-Gómez, S. A. et al. An updated phylogeny of the Alphaproteobacteria reveals that the parasitic Rickettsiales and Holosporales have independent origins. *eLife* **8**, e42535 (2019).

3 Hoshino, Y. & Gaucher, E. A. Evolution of bacterial steroid biosynthesis and its impact on eukaryogenesis. *Proc. Natl. Acad. Sci. USA* **118**, e2101276118 (2021).

4 Brocks, J. J. et al. Biomarker evidence for green and purple sulphur bacteria in a stratified Palaeoproterozoic sea. *Nature* **437**, 866-870 (2005).

5 Summons, R. E. & Capon, R. J. Fossil steranes with unprecedented methylation in ring-A. *Geochim. Cosmochim. Acta* **52**, 2733-2736 (1988).

6 van Maldegem, L. M. et al. Geological alteration of Precambrian steroids mimics early animal signatures. *Nat. Ecol. Evol.* **5**, 169–173 (2021).

7 Helliwell, K. E., Wheeler, G. L., Leptos, K. C., Goldstein, R. E. & Smith, A. G. Insights into the Evolution of Vitamin B_{12} Auxotrophy from Sequenced Algal Genomes. *Mol. Biol. Evol.* **28**, 2921-2933 (2011).

8 Helliwell, Katherine E. et al. Cyanobacteria and Eukaryotic Algae Use Different Chemical Variants of Vitamin B_{12}. *Curr. Biol.* **26**, 999-1008 (2016).

9 Mordukhova Elena, A. & Pan, J.-G. Evolved Cobalamin-Independent Methionine Synthase (MetE) Improves the Acetate and Thermal Tolerance of *Escherichia coli*. *Appl. Environ. Microbiol.* **79**, 7905-7915 (2013).

10 Deobald, D., Hanna, R., Shahryari, S., Layer, G. & Adrian, L. Identification and characterization of a bacterial core methionine synthase. *Sci. Rep.* **10**, 2100 (2020).

11 Croft, M. T., Lawrence, A. D., Raux-Deery, E., Warren, M. J. & Smith, A. G. Algae acquire vitamin B12 through a symbiotic relationship with bacteria. *Nature* **438**, 90-93 (2005).

12 Shelton, A. N. et al. Uneven distribution of cobamide biosynthesis and dependence in bacteria predicted by comparative genomics. *ISME J.* **13**, 789-804 (2018).

13 Heal, K. R. et al. Two distinct pools of B_{12} analogs reveal community interdependencies in the ocean. *Proc. Natl. Acad. Sci. USA* **114**, 364-369 (2017).

14 Erlacher, A. et al. Hyphomicrobiales as functional and endosymbiontic members in the lichen symbiosis of *Lobaria pulmonaria* L. *Front. Microbiol.* **6** (2015).

15 Yilmaz, P., Yarza, P., Rapp, J. Z. & Glöckner, F. O. Expanding the World of Marine Bacterial and Archaeal Clades. *Front. Microbiol.* **6** (2016).

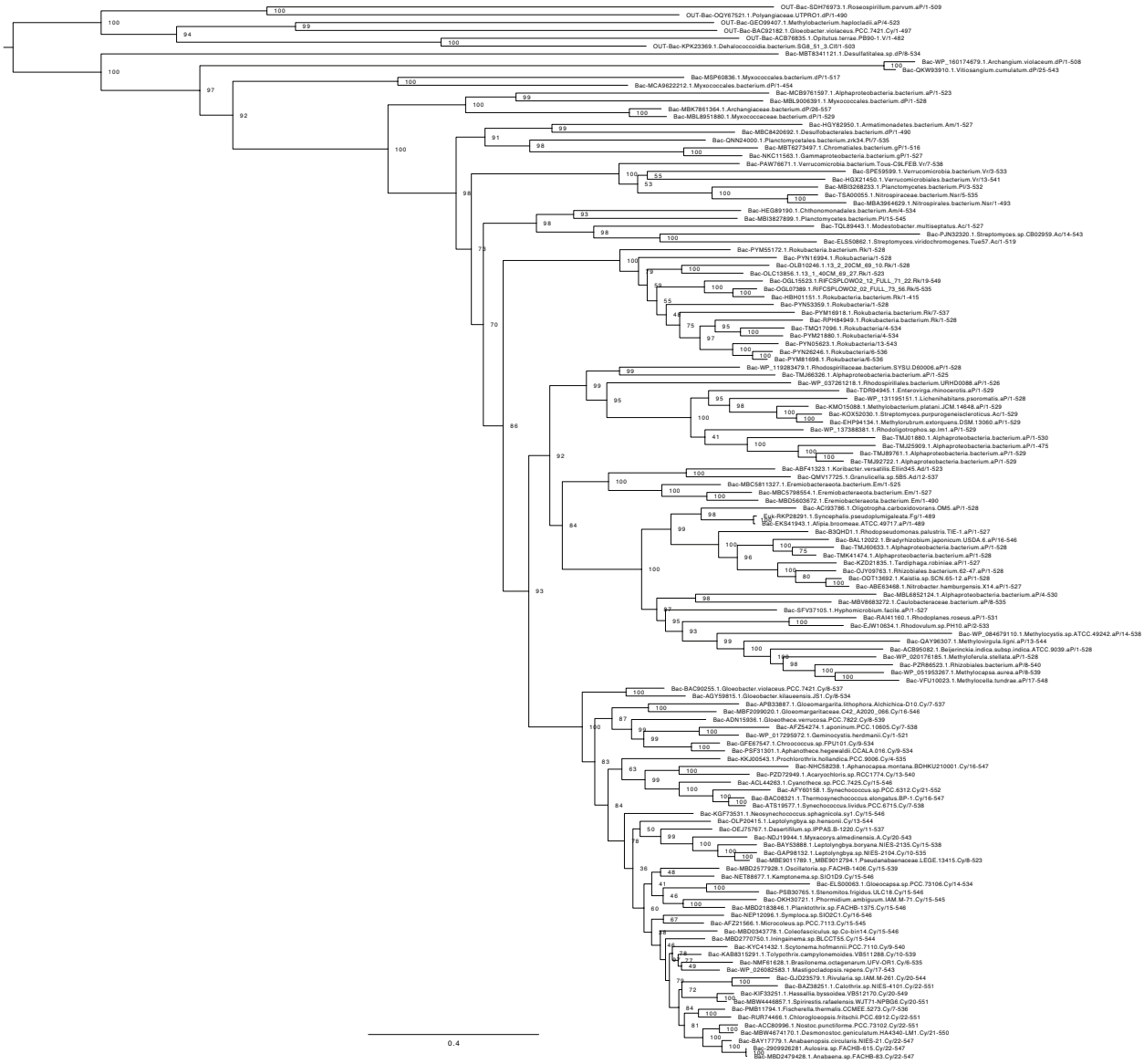
16 Ricci, J. N. et al. Diverse capacity for 2-methylhopanoid production correlates with a specific ecological niche. *ISME J.* **8**, 675-684 (2014).

17 Barott, K. L. et al. Microbial diversity associated with four functional groups of benthic reef algae and the reef-building coral *Montastraea annularis*. *Environ. Microbiol*. **13**, 1192-1204 (2011).

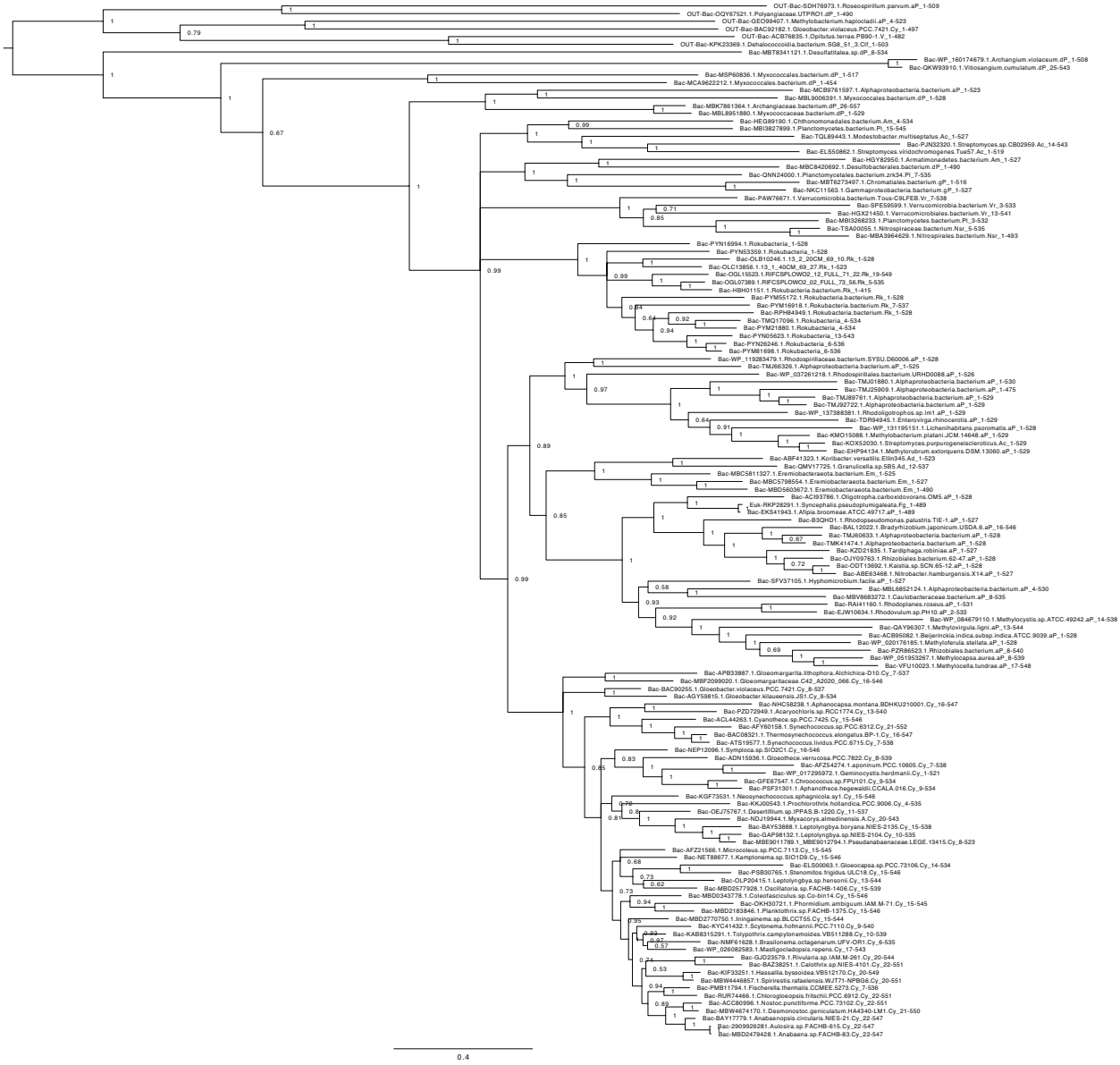
18 Wu, Z., Yang, X., Lin, S., Lee, W. H. & Lam, P. K. S. A Rhizobium bacterium and its population dynamics under different culture conditions of its associated toxic dinoflagellate *Gambierdiscus balechii*. *Mar. Life Sci. Technol.* **3**, 542-551 (2021).

19 Ramanan, R., Kim, B.-H., Cho, D.-H., Oh, H.-M. & Kim, H.-S. Algae–bacteria interactions: Evolution, ecology and emerging applications. *Biotechnol. Adv.* **34**, 14-29 (2016).

20 Wang, S., Meade, A., Lam, H.-M. & Luo, H. Evolutionary Timeline and Genomic Plasticity Underlying the Lifestyle Diversity in Hyphomicrobiales. *mSystems* **5**, e00438-00420 (2020).

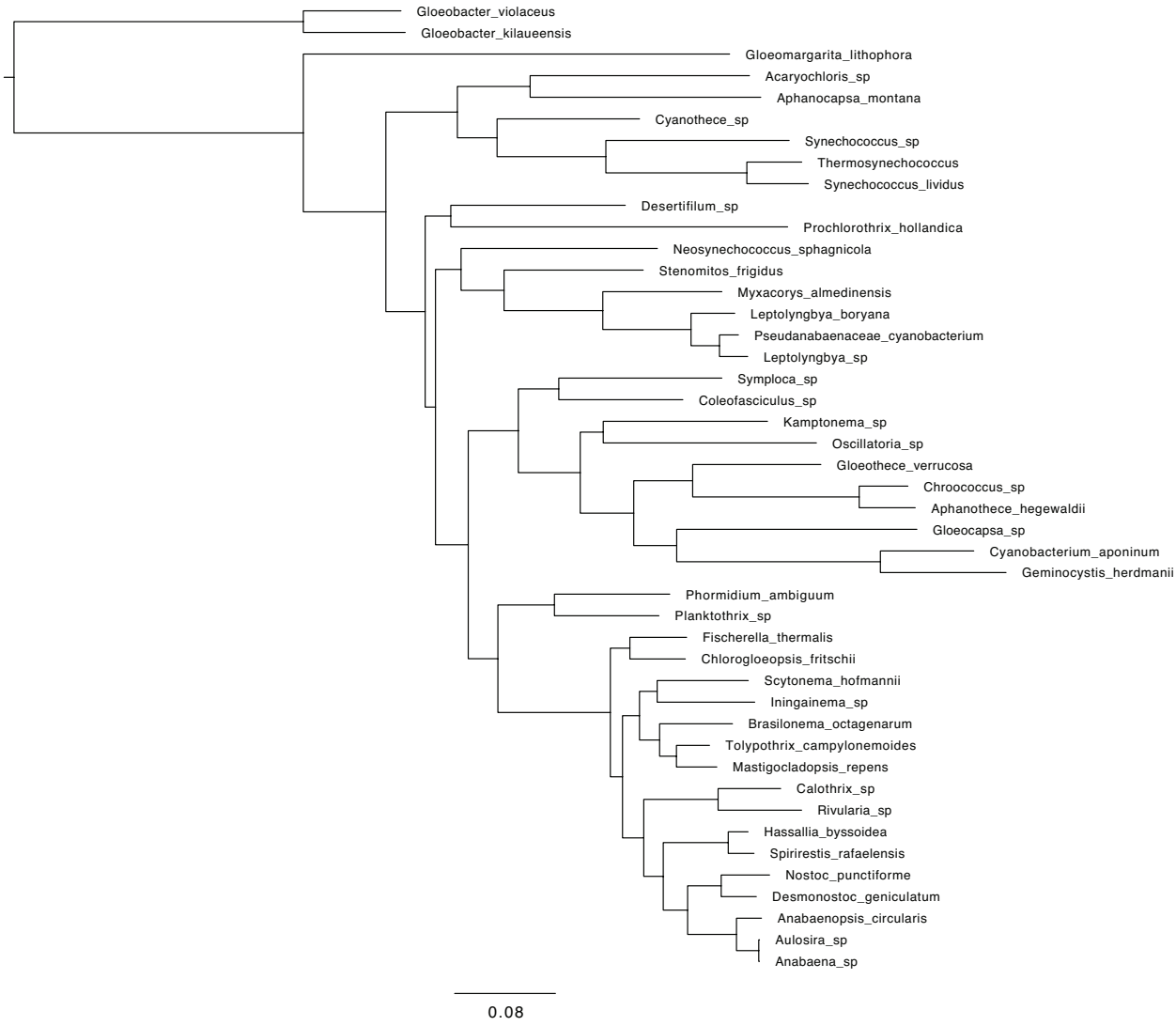


Supplementary Fig. 1. Maximum likelihood phylogenetic tree of HpnP. The figure is identical to Fig. 2 in the main text. The scale bar represents 0.4 amino acid replacements per site per unit evolutionary time. See Supplementary Data S1 for the sequence alignment.

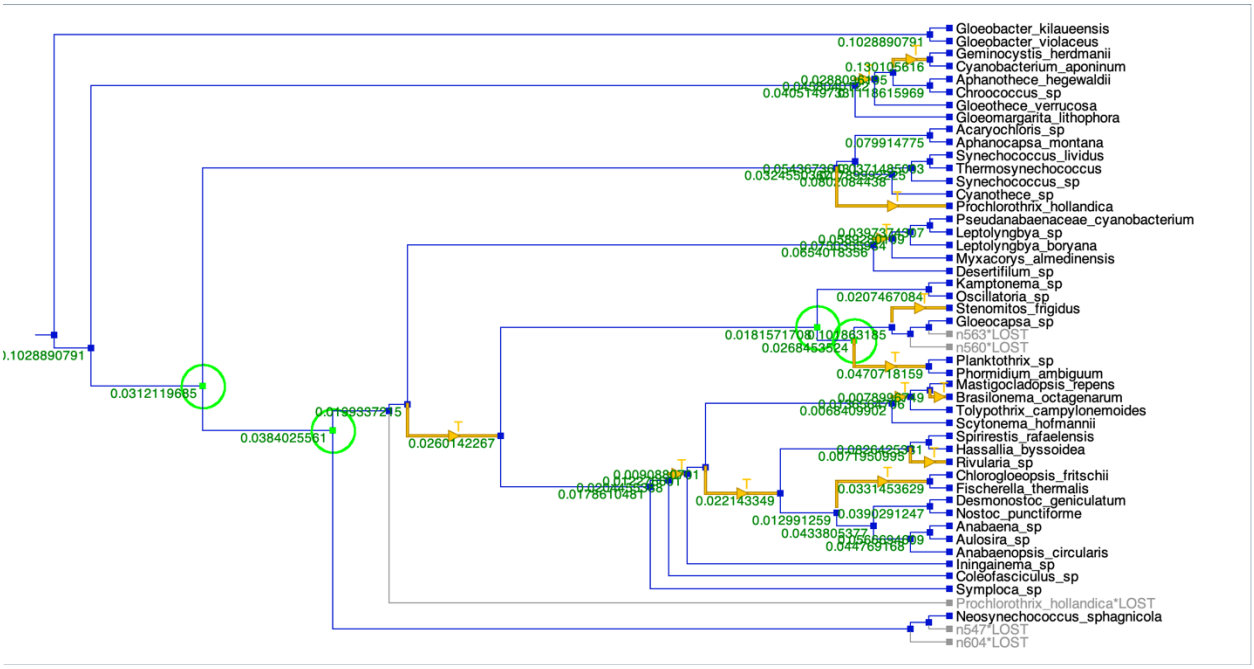


Supplementary Fig. 2. Bayesian phylogenetic tree of HpnP. The scale bar represents 0.4 amino acid replacements per site per unit evolutionary time. See Supplementary Data S1 for the sequence alignment.

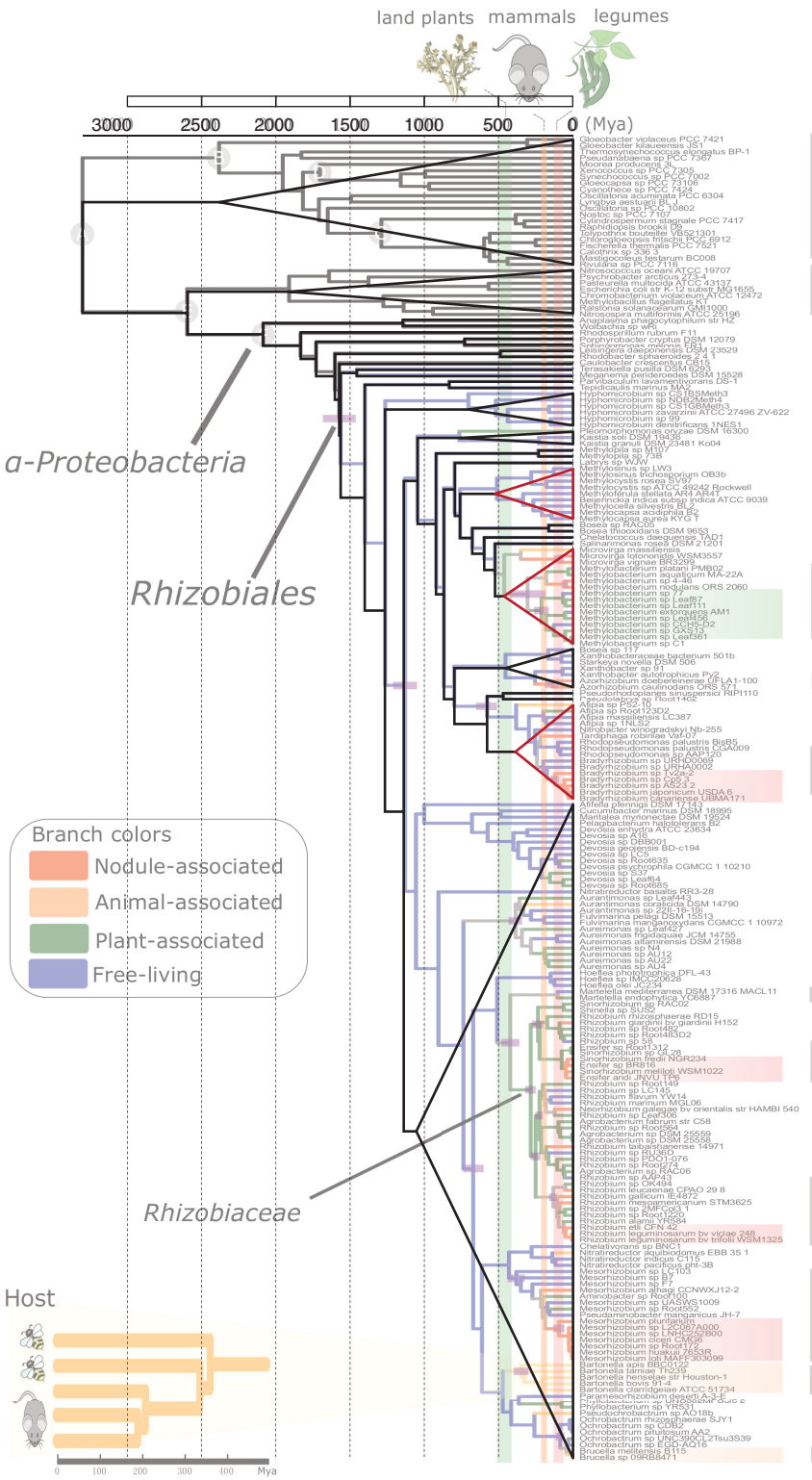
Supplementary Fig. 3. Maximum likelihood tree of class B radical SAM methyltransferases (HpnP homologs). The figure is identical to Fig. 2 insert. The scale bar represents 0.5 amino acid replacements per site per unit evolutionary time. See Supplementary Data S2 for the sequence alignment.



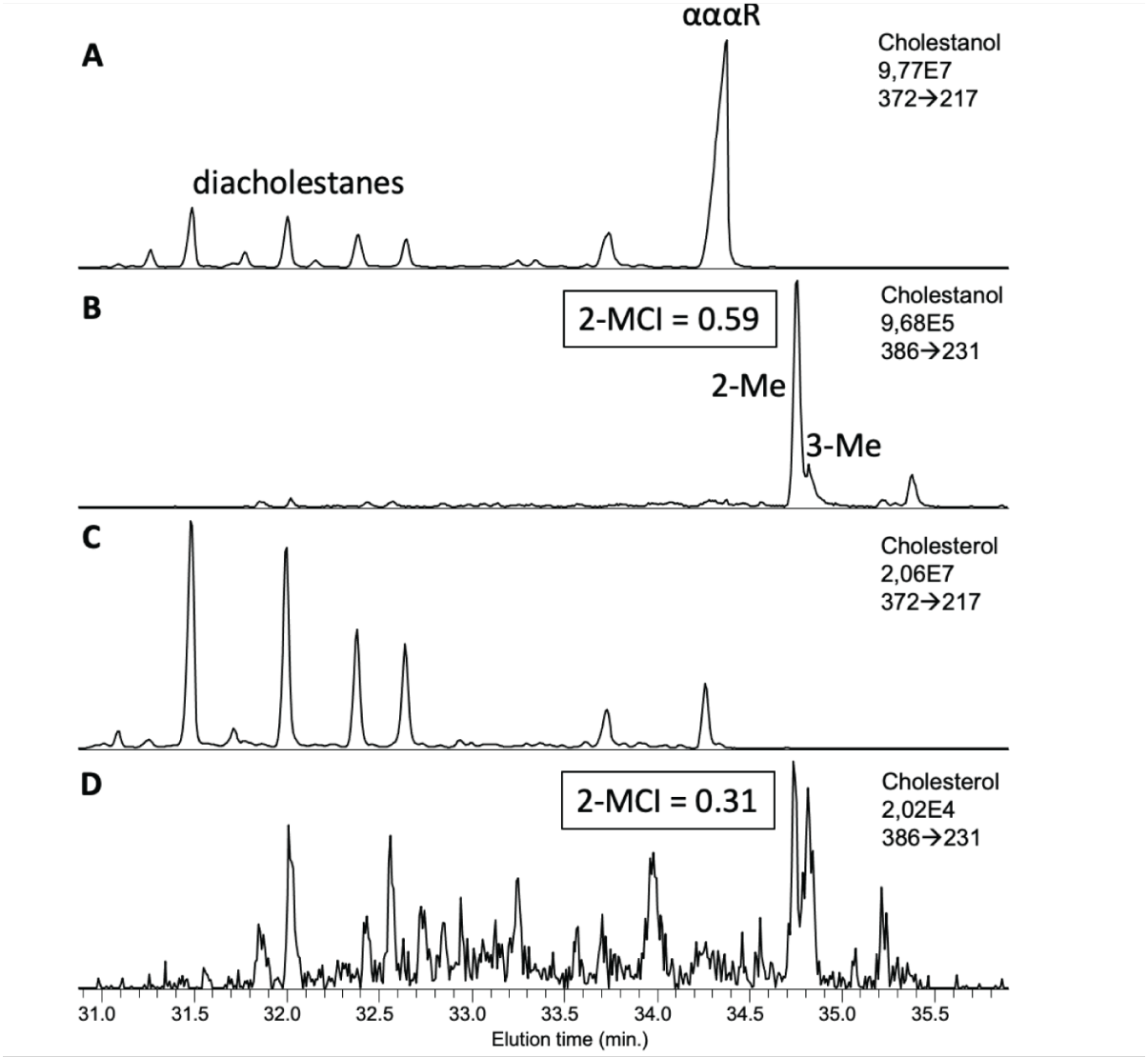
Supplementary Fig. 4. Species tree of 44 HpnP-containing Cyanobacteria. The scale bar represents 0.08 amino acid replacements per site per unit evolutionary time.



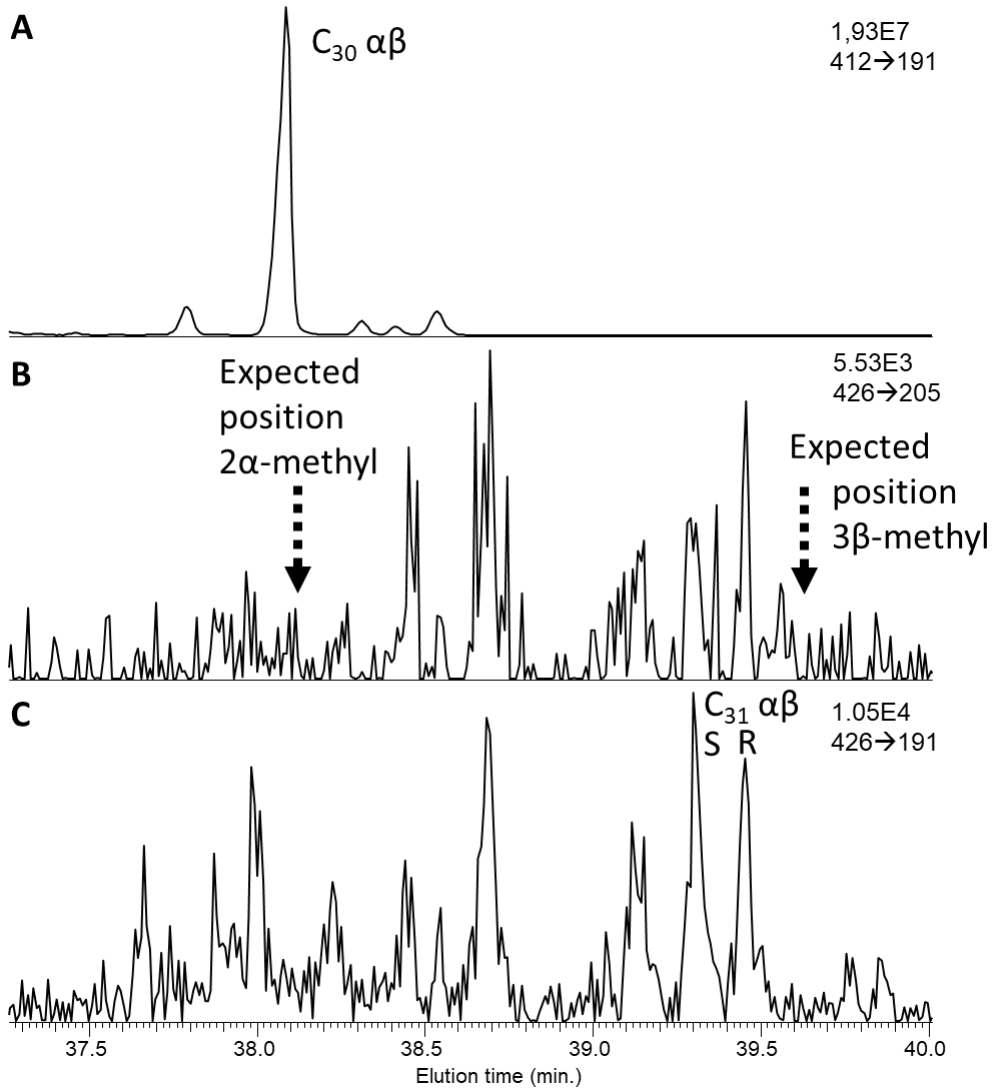
Supplementary Fig. 5. The comparison of cyanobacterial HpnP tree and the species tree by Notung. The label T indicates horizontal gene transfer.



Supplementary Fig. 6. Published dated phylogeny of Cyanobacteria and Alphaproteobacteria[20] and the superimposition of the simplified tree that is used in Fig. 4A.



Supplementary Fig 7. Pyrolysis experiment of cholestanol and cholesterol educts. (A) *m/z* 372 → 217 transition showing cholestanes and (B) *m/z* 386 → 231 showing methylcholestanes in cholestanol pyrolysates. (C) *m/z* 372 → 217 transition showing cholestanes and (D) *m/z* 386 → 231 showing methylcholestanes in cholesterol pyrolysates. The 2-methylcholestane index (2-MCI) is calculated analogously to the 2-MHI (%) for the ααα R isomers as 2-methylcholestane/(cholestane+2-methylcholestane)*100. Compound abbreviations: ααα R, C_{27} 5α(H), 14α(H), 17α(H)-cholestane (20R); 2-Me, 2-methylcholestane; 3-Me, 3-methylcholestane.



Supplementary Fig. 8. Pyrolysis experiment of diplopterol (1 µL injection from ca. 20 µL total extract, resulting in the overloading of C_{30} αβ-hopane). (A) *m/z* 412 → 191 transition displaying a dominance of the regular C_{30} αβ-hopane isomer. (B) *m/z* 426 → 205 transition displaying the absence of C_{31} methylhopanes with two expected elution times for 2α- and 3β-methylhopanes (according to AGSO reference standard). (C) *m/z* 426 → 191 transition displaying the C_{31} αβ-homohopane R and S isomers. Trace peaks in the *m/z* 426 → 205 transition are likely to mostly represent side-chain methylated hopanoids that yield higher intensities in the *m/z* 426 → 191 transition, although individual compounds are not identified, except for C_{31} αβ-homohopanes.

Supplementary table captions

Supplementary Table 1 (separate xlsx file). Accession numbers of SC and HpnP protein sequences for Rokubacteria. This list contains all available genomes in the phylum, regardless of their assembly level because no complete genome is available for this phylum.

Supplementary Table 2 (separate xlsx file). Distribution of SC and HpnP in Hyphomicrobiales in Alphaproteobacteria. The occurrence of SC and HpnP is summarized based on complete genome data and on draft genome data, respectively. The total number of available draft genomes is not counted and instead only the presence/absence of SC or HpnP is suggested in the table.

Supplementary Table 3 (separate xlsx file). Accession numbers of SC and HpnP protein sequences for Hyphomicrobiales. Only families that harbor HpnP-containing species are included in the list. Species that have a complete genome are all included regardless of their containing the SC or HpnP gene. Species that have only a draft genome data are included only if they contain both SC and HpnP genes and also if none of the species that have a complete genome data in the same genus have the HpnP gene. Abbreviation: OSC, oxidosqualene cyclase; CobG, precorrin-3B synthase; CobF, precorrin-6A synthase.

Supplementary Table 4 (separate xlsx file). Distribution of SC and HpnP in Cyanobacteria. The occurrence of SC and HpnP is summarized based on complete genome data and on draft genome data, respectively. The total number of available draft genomes is not counted and instead only the presence/absence of SC or HpnP is suggested in the table.

Supplementary Table 5 (separate xlsx file). Accession numbers of SC and HpnP protein sequences for Cyanobacteria. Species that have a complete genome are all included regardless of their containing the SC or HpnP gene. Species that have only a draft genome data are included only if they contain both SC and HpnP genes and also if none of the species that have a complete genome data in the same genus have the HpnP gene.

Supplementary Table 6 (separate xlsx file). Total hopane abundance and the 2-methylhopane index (2-MHI) in the drill core GR7 (Barney Creek Formation). Bold-face indicates samples that have hopanes only in a low abundance but display a high 2-MHI value. Bulk-analyzed samples are not included in our dataset (Fig. 4B). n.d. = not detected.

Supplementary Table 7 (separate xlsx file). Sample list for the 2-MHI record (Fig. 4B). Samples newly analyzed in this study are shown in brown, while samples from the literature are shown in black. References for previous studies are shown in Supplementary Material.

Supplementary Table 7 references

1 Williams, M. A Lipid Biomarker Investigation of Organic Matter Sources and Methane Cycling in Alaskan Thaw Lake Sediments, *Doctoral dissertation*, University of California Riverside, (2012).

2 Kuypers, M. M. M., van Breugel, Y., Schouten, S., Erba, E. & Damsté, J. S. S. N_2-fixing Cyanobacteria supplied nutrient N for Cretaceous oceanic anoxic events. *Geology* **32**, 853-856 (2004).

3 Wang, G., Wang, T. G., Simoneit, B. R. T., Zhang, L. & Zhang, X. Sulfur rich petroleum derived from lacustrine carbonate source rocks in Bohai Bay Basin, East China. *Org. Geochem.* **41**, 340-354 (2010).

4 Schaefer, B. et al. Microbial life in the nascent Chicxulub crater. *Geology* **48**, 328-332 (2020).

5 Castro, J. M. et al. Complex and protracted environmental and ecological perturbations during OAE 1a - Evidence from an expanded pelagic section from south Spain (Western Tethys). *Glob. Planet. Change* **183**, 103030 (2019).

6 Stüeken, E. E. et al. Effects of pH on redox proxies in a Jurassic rift lake: Implications for interpreting environmental records in deep time. *Geochim. Cosmochim. Acta* **252**, 240-267 (2019).

7 Saito, R. et al. Secular changes in environmental stresses and eukaryotes during the Early Triassic to the early Middle Triassic. *Palaeogeogr. Palaeocl*. **451**, 35-45 (2016).

8 Xia, L. et al. Coupling of paleoenvironment and biogeochemistry of deep-time alkaline lakes: A lipid biomarker perspective. *Earth-Sci. Rev*. **213**, 103499 (2021).

9 Słowakiewicz, M. et al. Shale-gas potential of the Mid-Carboniferous Bowland-Hodder Unit in the Cleveland Basin (Yorkshire), Central Britain. *J. Pet. Geol.* **38**, 59-75 (2015).

10 Lengger, S. K., Melendez, I. M., Summons, R. E. & Grice, K. Mudstones and embedded concretions show differences in lithology-related, but not source-related biomarker distributions. *Org. Geochem*. **113**, 67-74 (2017).

11 Spaak, G. et al. Environmental conditions and microbial community structure during the Great Ordovician Biodiversification Event; a multi-disciplinary study from the Canning Basin, Western Australia. *Glob. Planet. Change* **159**, 93-112 (2017).

12 Lee, C. et al. Carbon isotopes and lipid biomarkers from organic-rich facies of the Shuram Formation, Sultanate of Oman. *Geobiology* **11**, 406-419 (2013).

13 Parfenova, T. M. First data on methylhopanes in Lower Cambrian organic matter of the Siberian platform. *Dokl. Earth Sci.* **475**, 775-779 (2017).

14 Grosjean, E., Love, G. D., Kelly, A. E., Taylor, P. N. & Summons, R. E. Geochemical evidence for an Early Cambrian origin of the 'Q' oils and some condensates from north Oman. *Org. Geochem*. **45**, 77-90 (2012).

15 Stolper, D. A. et al. Paleoecology and paleoceanography of the Athel silicilyte, Ediacaran–Cambrian boundary, Sultanate of Oman. *Geobiology* **15**, 401-426 (2017).

16 Grosjean, E., Love, G. D., Stalvies, C., Fike, D. A. & Summons, R. E. Origin of petroleum in the Neoproterozoic–Cambrian South Oman Salt Basin. *Org. Geochem.* **40**, 87-110 (2009).

Supplementary data captions

Supplementary Data 1 (separate txt file). Multiple sequence alignment for the HpnP phylogeny.

Supplementary Data 2 (separate txt file). Multiple sequence alignment for the phylogeny of HpnP and HpnP-like proteins.