

# Identification of a methylase required for 2-methylhopanoid production and implications for the interpretation of sedimentary hopanes

Paula V. Welander<sup>a,1</sup>, Maureen L. Coleman<sup>a,1</sup>, Alex L. Sessions<sup>b</sup>, Roger E. Summons<sup>c</sup>, and Dianne K. Newman<sup>a,c,d,2</sup>

<sup>a</sup>Department of Biology, Massachusetts Institute of Technology, 77 Massachusetts Avenue, 68-380, Cambridge, MA 02139; <sup>b</sup>Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, MC100-23, 1200 East California Boulevard, Pasadena, CA 91125; <sup>c</sup>Department of Earth, Atmospheric and Planetary Science, Massachusetts Institute of Technology, 77 Massachusetts Avenue, E25-633, Cambridge, MA 02139; and <sup>d</sup>Howard Hughes Medical Institute, 77 Massachusetts Avenue, 68-171, Cambridge, MA 02139

Edited by John M. Hayes, Woods Hole Oceanographic Institution, Berkeley, CA, and approved January 27, 2010 (received for review November 10, 2009)

The rise of atmospheric oxygen has driven environmental change and biological evolution throughout much of Earth's history and was enabled by the evolution of oxygenic photosynthesis in the cyanobacteria. Dating this metabolic innovation using inorganic proxies from sedimentary rocks has been difficult and one important approach has been to study the distributions of fossil lipids, such as steranes and 2-methylhopanes, as biomarkers for this process. 2-methylhopanes arise from degradation of 2-methylbacteriohopanepolyols (2-MeBHPs), lipids thought to be synthesized primarily by cyanobacteria. The discovery that 2-MeBHPs are produced by an anoxygenic phototroph, however, challenged both their taxonomic link with cyanobacteria and their functional link with oxygenic photosynthesis. Here, we identify a radical SAM methylase encoded by the *hpnP* gene that is required for methylation at the C-2 position in hopanoids. This gene is found in several, but not all, cyanobacteria and also in  $\alpha$ -proteobacteria and acidobacteria. Thus, one cannot extrapolate from the presence of 2-methylhopanes alone, in modern environments or ancient sedimentary rocks, to a particular taxonomic group or metabolism. To understand the origin of this gene, we reconstructed the evolutionary history of HpnP. HpnP proteins from cyanobacteria, *Methylobacterium* species, and other  $\alpha$ -proteobacteria form distinct phylogenetic clusters, but the branching order of these clades could not be confidently resolved. Hence, it is unclear whether HpnP, and 2-methylhopanoids, originated first in the cyanobacteria. In summary, existing evidence does not support the use of 2-methylhopanes as biomarkers for oxygenic photosynthesis.

biomarkers | cyanobacteria | hopanoids | phylogeny | radical SAM

Our understanding of the evolution of life on Earth is based primarily on the fossil record of multicellular organisms. Yet the majority of life's history has been dominated by microorganisms whose biochemical inventions have significantly altered the Earth's environment. Thus, evolution of life and the surface environment of the Earth are interdependent (1). A challenge in determining when specific microbes first arose in Earth history is that they do not leave diagnostic morphological fossils that give insight into their metabolism (2). Because of this, an alternative strategy for studying microbial communities in the context of the Earth's distant past is to use organic compounds as "molecular fossils" or biomarkers. This method traces organic compounds preserved in ancient sedimentary rocks and links them to specific organisms and metabolisms by analogy with their production in modern organisms. Biomarkers have been useful in determining the composition and function of microbial communities in various modern environments (3–5) while biomarker signatures in ancient rocks have helped to reconstruct important events in deep time, such as the first appearance of major groups of organisms (6–8), the loss of biodiversity (9), and the evolution of cyanobacteria and oxygenic photosynthesis (10, 11).

Sedimentary hopanes are biomarkers whose distribution has been studied extensively. These hydrocarbon molecules are recognized as the diagenetic products of bacteriohopanepolyols, functionalized pentacyclic triterpenoids that are more generally referred to as hopanoids. Hopanoids are produced by a variety of bacteria and are thought to play a role in membrane integrity and stress tolerance (12–15). Because the hopanoid hydrocarbon skeleton is resistant to biodegradation, and is readily incorporated and preserved in sediments, these molecules have long been considered "molecular fossils" for ancient bacteria (16). Hopanoids methylated at the C-2 position are detected in both modern environments and, as 2-methylhopane hydrocarbons, in ancient marine and nonmarine sedimentary rocks. They are produced by a variety of cultured cyanobacteria and, consequently, cyanobacteria are proposed to be the most significant producers of 2-methylhopanoids in the modern environment (17). By extrapolation, these molecules have been used as evidence for the antiquity of oxygenic photosynthesis and for the waxing and waning of cyanobacterial primary productivity and nitrogen fixation in the geological past (10, 18, 19). However, the robustness of this interpretation was weakened by the discovery that an anoxygenic phototroph, *Rhodospseudomonas palustris* TIE-1, produces 2-methylhopanoids when grown under strictly anaerobic conditions (20). Because a strain of the same clade had previously been reported not to make 2-methylhopanoids (21), an important question emerged: was this a chance and unrepresentative finding, or is biosynthesis of these molecules widespread?

Currently, the taxonomic distribution of 2-methylhopanoids in modern bacteria is based primarily on lipid analyses of cultured organisms. Because the majority of bacterial diversity has not been cultured (22) and because the production of 2-methylhopanoids varies under different growth conditions (15, 20, 23), it is likely that our assessment of the taxonomic distribution of these molecules is incomplete. As a result, there is inadequate data to unambiguously interpret sedimentary 2-methylhopanes as biomarkers for cyanobacteria or other taxa. To address these issues and gain further insight into a poorly resolved biosynthetic pathway, we utilized a genetic and phylogenetic approach to identify the protein responsible for methylation of hopanoids at C-2. The use of a genetic marker offers three concrete advantages. First, it overcomes many of the constraints imposed by culturing and

Author contributions: P.V.W., M.L.C., and D.K.N. designed research; P.V.W. and M.L.C. performed research; A.L.S. and R.E.S. contributed new reagents/analytic tools; P.V.W., M.L.C., A.L.S., R.E.S., and D.K.N. analyzed data; and P.V.W., M.L.C., and D.K.N. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

<sup>1</sup>P.V.W. and M.L.C. contributed equally to this work.

<sup>2</sup>To whom correspondence should be addressed. E-mail: dkn@mit.edu.

This article contains supporting information online at [www.pnas.org/cgi/content/full/0912949107/DCSupplemental](http://www.pnas.org/cgi/content/full/0912949107/DCSupplemental).

condition-dependent expression when trying to identify other bacteria with the capacity to make 2-methylhopanoids. Second, it allows environmental samples to be screened for the presence of microorganisms capable of 2-methylhopanoid biosynthesis. Third, it permits phylogenetic analysis that can be used to infer the evolutionary history of the enzyme catalyzing the diagnostic step in their biosynthesis.

## Results

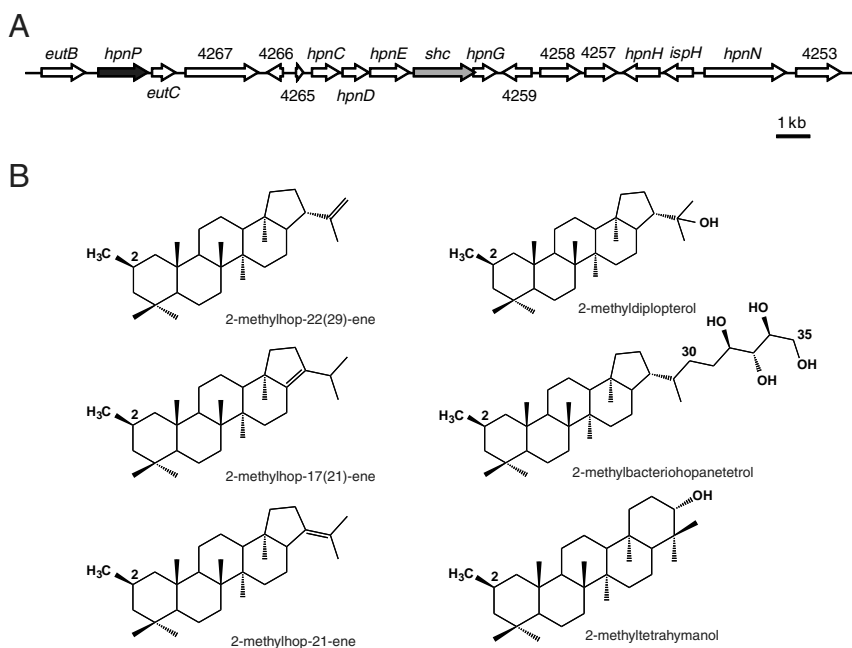
To determine which gene is required for C-2 methylation, we first identified putative hopanoid biosynthetic genes in the *R. palustris* TIE-1 genome. Currently, the only definitively known hopanoid biosynthetic step is the initial cyclization of squalene to the basic hopene structure diploptene by the squalene-hopene cyclase (*shc*). A previous study in *Zymomonas mobilis* and *Bradyrhizobium japonicum* identified several putative hopanoid biosynthesis genes surrounding the *shc* locus (24). We examined the regions upstream and downstream of the *shc* locus in TIE-1, and found a similar putative hopanoid biosynthetic gene cluster (Fig 1A). Approximately nine kilobases (kb) upstream of the squalene-hopene cyclase gene we observed an open reading frame (ORF 4269) that was annotated as a putative B-12 binding radical SAM protein (Fig. 1A). Because *R. palustris* cells have been shown to produce 2-methylhopanoids labeled at the C-2 position when fed labeled methionine (20, 25), it seemed reasonable that this methylation might occur through an S-adenosylmethionine (SAM) dependent mechanism. The presence of a radical SAM motif in ORF 4269, thus, made it an attractive candidate for the methylase.

ORF 4269 was deleted in frame by using homologous recombination mediated gene deletion, and total lipid extracts were assayed by gas chromatography-mass spectrometry (GC-MS) for the production of methylated hopanoids. *R. palustris* TIE-1 has been shown to produce at least six distinct C-2 methylated triterpenoids (Fig. 1B) and the resulting deletion strain did not produce any of these molecules (Fig. 2 and Fig. S1). The production of C-2 methylated hopanoids was restored by providing a copy of the *R. palustris* methylase gene under the native promoter on a self-replicating plasmid (Fig. 2). Therefore, ORF 4269 is the

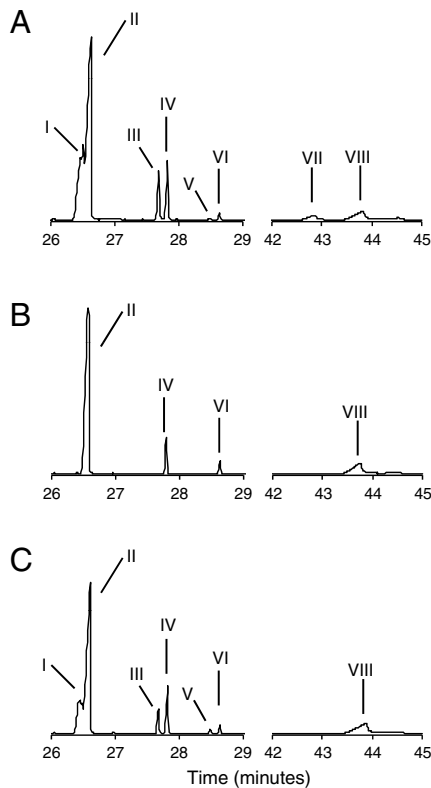
only gene required for methylation at the C-2 position and we propose the locus be renamed *hpnP*, following the nomenclature established in *Z. mobilis* and *B. japonicum* (24).

Of the 197 sequenced bacterial genomes (completed and draft) that contain the *shc* gene, and presumably produce hopanoids, only 30 contain an ortholog of the *hpnP* methylase gene (Fig. 3 and Table S1). Bacteria that contain the methylase are found in three different bacterial groups: the acidobacteria, the cyanobacteria, and one major subclade of the rhizobiales within the  $\alpha$ -proteobacteria. Of these thirty bacteria, only ten have been tested for hopanoid production and all ten of these strains produce 2-methylhopanoids (17, 20, 26, 27). There are seven *Methylobacterium* and two *Cyanothece* strains that have the methylase gene but these specific strains have not been tested for hopanoid production. However, 2-methylhopanoid production has been shown in other *Methylobacterium* and *Cyanothece* strains, making it likely that these bacteria also produce 2-methylhopanoids (17, 28). Furthermore, there are five cyanobacteria, four  $\alpha$ -proteobacteria, two  $\gamma$ -proteobacteria, two  $\delta$ -proteobacteria, two  $\beta$ -proteobacteria, two actinobacteria, and one firmicute that have been tested for hopanoid production and whose genomes have been sequenced. None of these bacteria contain an ortholog of the methylase gene and although they produce hopanoids, they do not produce 2-methylhopanoids (Table S2). Based on the available data, there is, thus far, a perfect correspondence between the presence of the methylase gene and the production of 2-methylhopanoids. We, therefore, assume that homologous HpnP proteins share a common function.

As can be seen in Table S1, the organisms that contain the methylase are found in a variety of environments, including freshwater, soil, wastewater, stem nodules, and root nodules. Consistent with the rarity of 2-methylhopanoids in cultured marine cyanobacteria (17), none of the sequenced cyanobacteria that contain the methylase gene were isolated from marine environments. Because the majority of sedimentary 2-methylhopanes are thought to derive from ancient shallow marine environments dominated by phototrophs (18, 19), we anticipated that our bioinformatics analysis might reveal the presence of the HpnP



**Fig. 1.** Identification of a putative hopanoid biosynthetic gene cluster in *R. palustris* TIE-1. (A) The *shc* locus (Gray Arrow), which is required for hopanoid biosynthesis, is surrounded by several putative hopanoid biosynthesis genes. Upstream of *shc*, a hypothetical B-12 binding radical SAM protein was identified (*hpnP*, Black Arrow) as a C-2 methylase candidate. (B) The structure of the six C-2 methylated triterpenoids produced by *R. palustris* TIE-1: three 2-methylhopenes (hop-22(29)-ene, hop-17(21)-ene, and hop-21-ene), 2-methyldiplopteroI, 2-methyltetrahymanol, and 2-methylbacteriohopanetetrol.



**Fig. 2.** GC-MS total ion chromatograms of acetylated total lipid extracts of *R. palustris* cultures. The  $\Delta hpnP$  mutant does not produce any C-2 methylated hopanoids. (A) *R. palustris* TIE-1. (B) *R. palustris*  $\Delta hpnP$ . (C) *R. palustris*  $\Delta hpnP$  complemented with the *hpnP* gene on a self-replicating plasmid. Numbered compounds: I, 2-methylhopenes; II, hopenes; III, 2-methyldiplopterol; IV, diplopterol; V, 2-methyltetrahymanol; VI, tetrahymanol; VII, 2-methylbacteriohopanetetrol; and VIII, bacteriohopanetetrol. Acetylated total lipid extracts were analyzed by high temperature GC-MS and compounds were identified by comparison of retention times and mass spectra to authentic compounds and published mass spectra (Table S6). Peak II is a co-elution of two hopene structures, hop-22(29)-ene and hop-21-ene. No methylated hopanoids eluted between 29–42 min (full chromatogram shown in Fig. S1). The complemented  $\Delta hpnP$  strain did produce 2-methylbacteriohopanetetrol (compound VII) as verified by analysis of the 205 Da mass chromatogram and the mass spectrum; however, the peak is too broad to be seen at the resolution of this figure of the GC-MS total ion current.

methylase in a yet untested marine cyanobacterium. However, the  $\alpha$ -proteobacterium *Nitrobacter* sp. Nb-311A was the only marine organism in both the genomic and the marine metagenomic databases to contain a homolog of the HpnP methylase.

To understand the origin of this protein, we reconstructed its phylogeny by maximum likelihood and tested the sensitivity of the topology to the input alignment and choice of outgroup taxa. The LG substitution model (29) chosen by ProfTest (30) was used for all trials, and 10 random starting trees were used to prevent the tree search from getting trapped in a local optimum. Because the quality of the alignment is a major determinant of tree accuracy (31, 32), we used three alignment programs (CLUSTALW, Muscle, and T-Coffee) (33–35). In addition, we used three filtering strategies (relaxed, stringent, and no filtering) to remove uncertain alignment positions (Fig. S2). Although filtering may increase the signal to noise ratio of the data, it also discards potentially informative sites, and it may not be beneficial for short alignments (36, 37). To root the tree, we used all members of a sister clade of radical SAM proteins as the outgroup (Fig. S3), an approach suggested as optimal (38). For comparison, we added a second outgroup clade in some trials (Fig. S3).

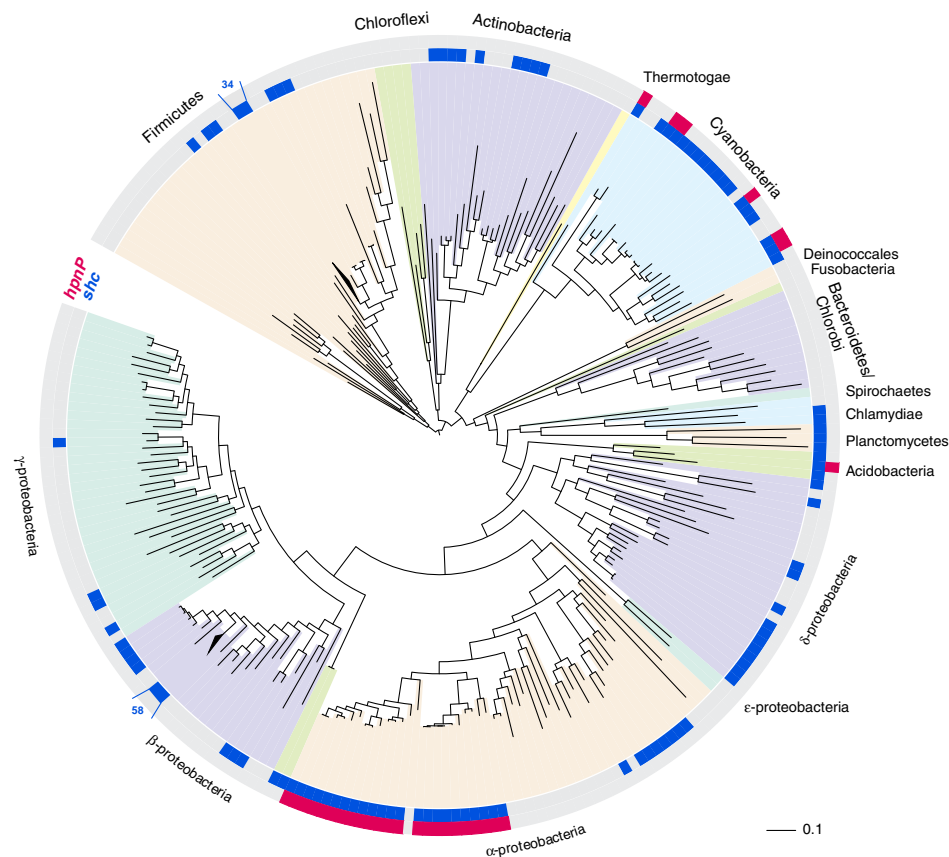
The unrooted ingroup topology shows distinct clusters for the HpnP protein from *Methylobacterium* species, the *Bradyrhizobium/Nitrobacter/Rhodospseudomonas* group, and the cyanobacteria (Fig. S4). The branching order of these groups, however, varied significantly between trials. Three different rooted topologies emerged (Fig. 4 and Table S3). Most commonly, the  $\alpha$ -proteobacterial HpnP sequences were monophyletic and nested within lineages of cyanobacterial HpnP sequences, and from this topology it is most parsimonious to infer that the ancestral HpnP resided in a cyanobacterium (Fig. 4A). In some cases, the cyanobacterial and  $\alpha$ -proteobacterial sequences formed sister clades (Fig. 4B), or the cyanobacterial sequences were nested within the  $\alpha$ -proteobacterial sequences (Fig. 4C). This ambiguity suggests that there is insufficient phylogenetic signal to confidently resolve the branching order of the major clades. We tested whether the best topology recovered for a given alignment was significantly better than the two competing topologies using the Approximately Unbiased test (39). This test failed to reject either of the competing hypotheses (Table S4) and, thus, even for a given alignment, we cannot confidently choose the best tree.

Regardless of the precise branching order, our results reveal a complex evolutionary history for HpnP. Its phyletic distribution could arise two ways: either it was present in the ancestor of cyanobacteria and  $\alpha$ -proteobacteria and was repeatedly lost, or it was horizontally transferred between cyanobacteria and  $\alpha$ -proteobacteria. Gene transfer between these phyla has been observed (40), and cyanobacteria most often share genes with the  $\alpha$ -proteobacteria compared to other phyla (41). Within the cyanobacteria, HpnP is sporadically distributed and the HpnP phylogeny is not congruent with the species phylogeny (42). This implies gene transfer and gene loss among cyanobacteria and, thus, it is impossible to generalize about whether all ancient cyanobacteria produced 2-methylhopanoids. The position of the lone acidobacterium in the HpnP tree is also uncertain. Within the  $\alpha$ -proteobacteria, the HpnP and species phylogenies are similar (43), implying vertical descent within this group. From this phylogenetic analysis, it is clear that finding more taxa to fill in the HpnP tree is critical for better resolution of this protein's history.

## Discussion

In this study, we identified a radical SAM methylase encoded by *hpnP* that is required for C-2 methylation of bacterial hopanoids. At present, the *hpnP* gene appears to be a robust predictor of the capacity to produce 2-methylhopanoids; whether it will remain an exclusive predictor remains to be seen as more methylase genes are identified in the future. For now, however, the use of 2-methylhopanoids and sedimentary 2-methylhopanes as indicators for cyanobacteria prompted us to assess the distribution of the *hpnP* gene product in bacterial genomes. We found that the presence of HpnP was not restricted to the cyanobacteria as it was also found in the  $\alpha$ -proteobacteria (specifically the rhizobiales) and one acidobacterium.

The presence of this methylase in the rhizobiales and acidobacteria reinforces the need for caution in using 2-methylhopanoids as indicators of cyanobacteria in modern environments. For example, the detection of 2-methylhopanoids in soil, freshwater, and marine environmental samples has been used as a proxy for cyanobacteria in several ecological studies (44, 45) but the presence of rhizobiales and acidobacteria in these habitats is also well documented (46–48). The occurrence of the C-2 methylase in these two groups also raises the possibility that the acidobacteria or rhizobiales could have been the source, in part or in whole, for ancient sedimentary 2-methylhopanes. We attempted to address this possibility through a phylogenetic analysis of HpnP that had the potential to reveal whether the ancestral methylase was most likely found in cyanobacteria,  $\alpha$ -proteobacteria, or acidobacteria. However, the small number of HpnP



**Fig. 3.** Occurrence of squalene-hopene cyclase (*shc*) and *hpnP* methylase genes in distantly related Bacteria. The tree represents phylogenetic relationships among bacterial species and was constructed by maximum likelihood using five concatenated core proteins (Adk, RpoB, GyrB, RecA, and SecY). **Blue Bars** in inner ring indicate strains with *shc*; **Red Bars** in outer ring indicate strains with *hpnP*. All sequenced isolates containing *shc* and/or *hpnP* are depicted as well as representatives from other bacterial families. The *Bacillus* (34 isolates) and *Burkholderia* (58 isolates) genera have been collapsed.

sequences available and the sensitivity of the tree construction to the input alignment made it difficult to resolve the evolutionary history of this protein. Therefore, our phylogenetic analysis of HpnP does not presently support the use of 2-methylhopanoids as biomarkers for the antiquity of cyanobacteria and leaves open the possibility that other phyla may have been responsible for the deposition of ancient sedimentary 2-methylhopanes. But is this reasonable?

Geologists interpret most ancient 2-methylhopane deposition sites to have been shallow marine carbonate platforms, which in modern times are usually dominated by marine phototrophs such as cyanobacteria (10, 19). Although the majority of the  $\alpha$ -proteobacteria and the one acidobacterium identified in this study are from soil or freshwater environments, given the wide ecological distribution of these groups (48, 49), we cannot exclude the scenario that their ancestors could have inhabited shallow marine environments. Leaving this possibility open seems especially appropriate given that, to date, no marine cyanobacteria have been shown to produce methylated hopanoids nor do we see the methylase gene in any of the sequenced genomes of marine cyanobacteria. Interestingly, the only marine organism we identified in this study was *Nitrobacter* sp. Nb-311A, a nitrite-oxidizing  $\alpha$ -proteobacterium belonging to the rhizobiales. Nitrite-oxidizing bacteria have been found in practically all terrestrial, marine, and fresh water habitats so it is plausible that their ancestors could have been responsible for 2-methylhopane deposition (50). Indeed, in the absence of independent evidence linking any one of these groups to an ancient environment, it seems prudent to remain agnostic regarding the identity of the earliest source of 2-methylhopanes at present. In the future, with a larger sequence database for HpnP, it should be possible to resolve the evolutionary history of the biosynthesis of 2-methylhopanes with confidence.

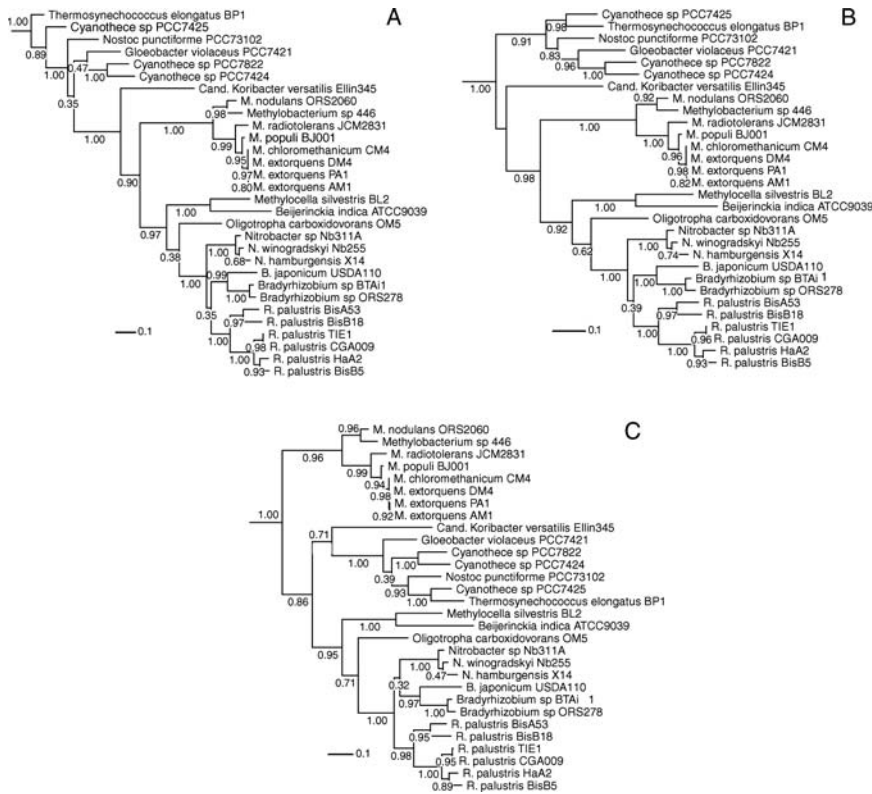
In conclusion, four lines of evidence suggest that 2-methylhopanes cannot be used as biomarkers for oxygenic photosynthesis: *i*) *hpnP* is distributed across many modern bacteria that do not engage in oxygenic photosynthesis, *ii*) cyanobacteria are not obligate oxygenic photoautotrophs (51), *iii*) not all cyanobacteria make 2-methylhopanoids (17), and *iv*) in the cyanobacterium *Nostoc punctiforme*, 2-methylhopanoids localize to the outer membranes of akinetes, a survival structure that is not photosynthetically active (23). Resolving the biological function of 2-methylhopanoids in modern bacteria is, therefore, necessary to better interpret the meaning of 2-methylhopanes in ancient sediments. Regardless of their cellular function, the application of (phylo)genetics to identify and interpret the history of a gene required for the biosynthesis of a specific geostable compound demonstrates how two classes of molecular fossils—genes and biomarkers—can constrain each other and provide a more informed understanding of microbial evolution and metabolism.

## Materials and Methods

**Bacterial Strains and Growth Conditions.** Bacterial strains used in this study are listed in Table S5. *Escherichia coli* strains were grown in lysogeny broth (LB) at 37 °C. *Rhodospseudomonas palustris* strains were grown chemoheterotrophically in unbuffered YP medium (0.3% yeast extract, 0.3% peptone) at 30 °C in the dark while shaking at 250 RPM. For growth on solid medium, LB or YP was solidified with 1.5% agar and supplemented, if necessary, with gentamicin at 20  $\mu\text{g}/\text{mL}$  (*E. coli*) or 800  $\mu\text{g}/\text{mL}$  (*R. palustris*).

## DNA Methods, Plasmid Construction, Transformation, and Strain Construction.

All plasmid constructions are described in Table S5. QIAprep Spin Miniprep Kit (Qiagen) was used for isolation of plasmid DNA from *E. coli*. Genomic DNA from *R. palustris* strains was isolated by using the DNeasy Blood and Tissue Kit (Qiagen). DNA sequences of all cloning intermediates were confirmed by sequencing at the Biopolymers Laboratory in the MIT Center for Cancer Research. *E. coli* strains were transformed by electroporation by using an Electroporator 2510 (Eppendorf) as recommended by the supplier. Plasmids were mobilized from *E. coli* S17-1 into *R. palustris* by conjugation



**Fig. 4.** Phylogenetic relationships among HpnP methylases from three distant bacterial phyla. (A) The best tree found in trial 15 (Table S3), with early branching cyanobacterial HpnP. (B) The best tree in trial nine, with monophyletic cyanobacterial HpnP. (C) The best tree in trial two, with early branching *Methylobacterium* HpnP. Branch support is indicated by nonparametric approximate likelihood ratio test (aLRT) statistics. Genus abbreviations: M = *Methylobacterium*, N = *Nitrobacter*, B = *Bradyrhizobium*, and R = *Rhodospseudomonas*.

on YP agar plates that were incubated overnight at 30 °C (52, 53). Deletion of the *hpnP* locus in TIE-1 and complementation of the deletion mutant were done as previously described (15).

**Gas Chromatography-Mass Spectrometry (GC-MS) Analysis.** For hopanoid analysis, strains were grown to late stationary phase (7 d). Total lipid extracts were prepared by solvent extraction, derivatized as acetates, and analyzed by high temperature GC-MS utilizing a DB-XXL column as previously described (15). Compounds were identified by comparison of retention times and mass spectra to authentic compounds (tetrahymanol from *Trimyema* sp.; diploptero from *Methylococcus capsulatus*; bacteriohopanepentol from *M. capsulatus*) and published mass spectra [hopenes(54); tetrahymanol: (55)]. C-2 methylation was inferred from the shift of fragment *m/z* 191 to 205, *m/z* 369 to 383 and from the relative retention time compared to the desmethyl homolog (~0.15 min earlier). Retention times and molecular ions are listed in Table S6.

**Phylogeny of HpnP.** Homologs of HpnP were identified by BLAST and aligned by using ClustalW, Muscle, or T-Coffee (33–35). Alignments were trimmed by Gblocks (56) using stringent (default parameters) or relaxed settings. For the stringent case, no gaps were allowed in any sequence. For the relaxed case, the minimum number of sequences for conserved and flanking positions were set to their minimum allowed values, maximum number of nonconserved positions was set to 100, minimum length of a block was 4, and gaps were allowed in half the sequences. Fig S2 shows an example alignment generated by Muscle and trimmed by Gblocks. The appropriateness of the two outgroup families was confirmed by a more exhaustive tree of related proteins (Fig. S3) that clearly shows that HpnP and the outgroup families are closely related but distinct clades.

Maximum likelihood trees were constructed by PhyML (29) using the LG + gamma model [chosen by ProtTest (30)], 6 gamma rate categories, 10 random starting trees, SPR + NNI branch swapping, and substitution parameters estimated from the data. Branch support was determined by SH-like approximate LRT statistics (29). The species tree was constructed similarly by using five concatenated protein sequences (RpoB, RecA, Adk, GyrB, and SecY) from 375 species, trimmed using Gblocks to 1679 positions, and rooted in the Firmicutes. The species tree is not intended to be exhaustive, but is used to show the distribution of the *shc* and *hpnP* genes across bacterial phyla.

We used the Approximately Unbiased test (39) implemented in CONSEL (57) to compare tree topologies. The topologies shown in Fig. 4 were used as constraints, and branch lengths and substitution parameters were optimized on these topologies by PhyML. The likelihood of the best unconstrained tree was then compared to the likelihoods of the two alternative trees, giving an AU test p-value for each tree (Table S4). Any tree with a p-value less than 0.05 can be rejected. If multiple trees have p-values above this cutoff, then we cannot confidently choose the true tree from among them.

**ACKNOWLEDGMENTS.** We thank David Doughty, Jacob Waldbauer, and Eric Alm for technical assistance and helpful discussions, and the reviewers for their constructive comments that improved the manuscript. This work was supported by grants from the National Aeronautics and Space Administration (NASA) Exobiology Program (A.L.S., D.K.N., and R.E.S.), the NASA Astrobiology Institute (R.E.S.), a National Science Foundation Postdoctoral Minority Fellowship (P.V.W.), and an Agouron Institute Geobiology Postdoctoral Fellowship (M.L.C.). D.K.N. is an Investigator of The Howard Hughes Medical Institute.

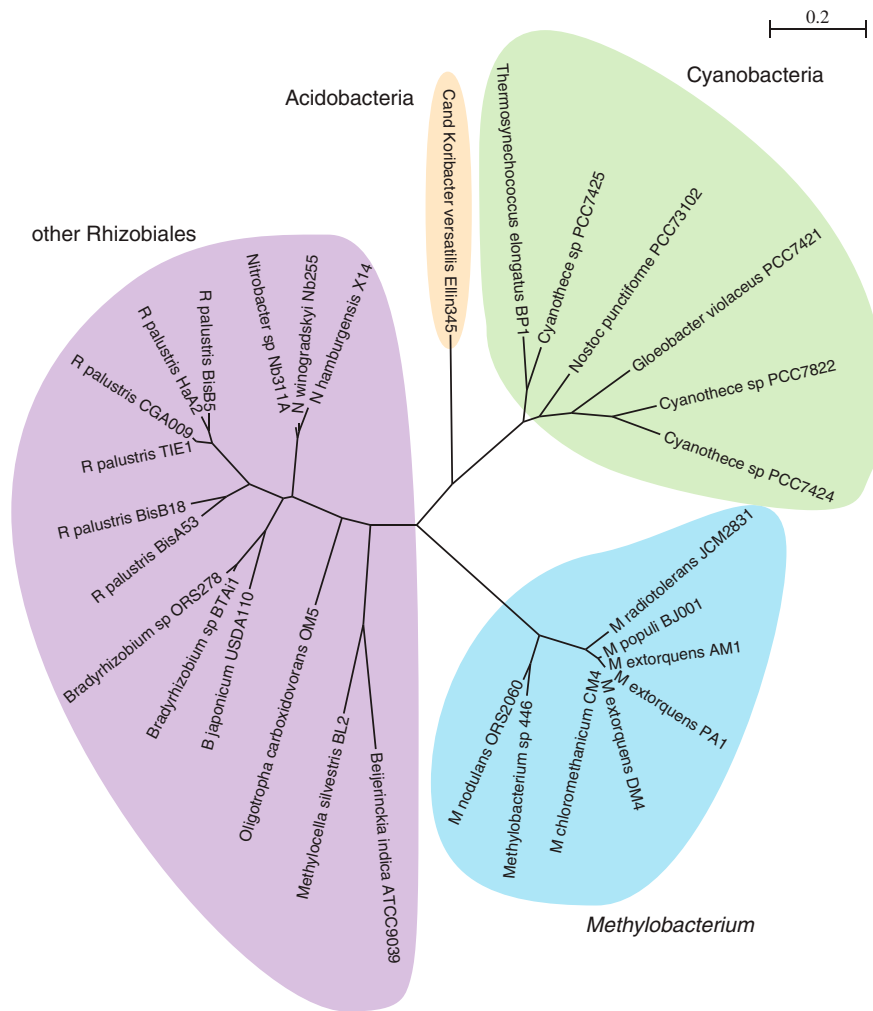
- Knoll AH (2003) The geological consequences of evolution. *Geobiology* 1:3–14.
- Cavalier-Smith T, Brasier M, Embley TM (2006) Introduction: How and when did microbes change the world?. *Philos T Roy Soc B* 361:845–850.
- Elvert M, Niemann H (2008) Occurrence of unusual steroids and hopanoids derived from aerobic methanotrophs at an active marine mud volcano. *Org Geochem* 39:167–177.
- Talbot HM, Watson DF, Pearson EJ, Farrington P (2003) Diverse biohopanoid compositions of non-marine sediments. *Org Geochem* 34:1353–1371.
- Hinrichs KU, Hayes JM, Sylva SP, Brewer PG, DeLong EF (1999) Methane-consuming archaeobacteria in marine sediments. *Nature* 398:802–805.
- Brocks JJ, et al. (2005) Biomarker evidence for green and purple sulfur bacteria in a stratified Palaeoproterozoic sea. *Nature* 437:866–870.
- Moldowan JM, Talyzina NM (1998) Biogeochemical evidence for dinoflagellate ancestors in the Early Cambrian. *Science* 281:1168–1170.
- Brocks JJ, Logan GA, Buick R, Summons RE (1999) Archean molecular fossils and the early rise of eukaryotes. *Science* 285:1033–1036.
- Grice K, et al. (2005) Photic zone euxinia during the Permian-Triassic superanoxic event. *Science* 307:706–709.
- Brocks JJ, Buick R, Summons RE, Logan GA (2003) A reconstruction of Archean biological diversity based on molecular fossils from the 2.78 to 2.45 billion-year-old Mount Bruce Supergroup, Hamersley Basin, Western Australia. *Geochim Cosmochim Acta* 67:4321–4335.
- Summons RE, Jahnke LL, Hope JM, Logan GA (1999) 2-Methylhopanoids as biomarkers for cyanobacterial oxygenic photosynthesis. *Nature* 400:554–557.

12. Bosak T, Losick RM, Pearson A (2008) A polycyclic terpenoid that alleviates oxidative stress. *Proc Natl Acad Sci USA* 105:6725–6729.
13. Poralla K, Hartner T, Kannenberg E (1984) Effect of temperature and pH on the hopanoid content of *Bacillus acidocaldarius*. *FEMS Microbiol Lett* 23:253–256.
14. Poralla K, Muth G, Hartner T (2000) Hopanoids are formed during transition from substrate to aerial hyphae in *Streptomyces coelicolor* A3(2). *FEMS Microbiol Lett* 189:93–95.
15. Welander PV, et al. (2009) Hopanoids play a role in membrane integrity and pH homeostasis in *Rhodospseudomonas palustris* TIE-1. *J Bacteriol* 191:6145–6156.
16. Ourisson G, Albrecht P (1992) Hopanoids. 1. Geohopanoids—the most abundant natural products on Earth. *Acc Chem Res* 25:398–402.
17. Talbot HM, et al. (2008) Cyanobacterial bacteriohopanepolyol signatures from cultures and natural environmental settings. *Org Geochem* 39:232–263.
18. Eigenbrode JL, Freeman KH, Summons RE (2008) Methylhopane biomarker hydrocarbons in Hamersley Province sediments provide evidence for Neoproterozoic aerobicity. *Earth Planet Sc Lett* 273:323–331.
19. Waldbauer JR, Sherman LS, Sumner DY, Summons RE (2009) Late Archean molecular fossils from the Transvaal Supergroup record the antiquity of microbial diversity and aerobicity. *Precambrian Res* 169:28–47.
20. Rashby SE, Sessions AL, Summons RE, Newman DK (2007) Biosynthesis of 2-methylbacteriohopanepolyols by an anoxygenic phototroph. *Proc Natl Acad Sci USA* 104:15099–15104.
21. Neunlist S, Bissleret P, Rohmer M (1988) The hopanoids of the purple non-sulfur bacteria *Rhodospseudomonas palustris* and *Rhodospseudomonas acidophila* and the absolute configuration of bacteriohopanetetrol. *Eur J Biochem* 171:245–252.
22. Pace NR (1997) A molecular view of microbial diversity and the biosphere. *Science* 276:734–740.
23. Doughty DM, Hunter RC, Summons RE, Newman DK (2009) 2-Methylhopanoids are maximally produced in akinetes of *Nostoc punctiforme*: geobiological implications. *Geobiology* 7:524–532.
24. Perzl M, et al. (1998) Cloning of conserved genes from *Zymomonas mobilis* and *Bradyrhizobium japonicum* that function in the biosynthesis of hopanoid lipids. *Biochim Biophys Acta* 1393:108–118.
25. Zundel M, Rohmer M (1985) Prokaryotic triterpenoids. 3. The biosynthesis of 2  $\beta$ -methylhopanoids and 3  $\beta$ -methylhopanoids of *Methylobacterium organophilum* and *Acetobacter pasteurianus* ssp. *pasteurianus*. *Eur J Biochem* 150:35–39.
26. Bravo JM, Perzl M, Hartner T, Kannenberg EL, Rohmer M (2001) Novel methylated triterpenoids of the gammacerane series from the nitrogen-fixing bacterium *Bradyrhizobium japonicum* USDA 110. *Eur J Biochem* 268:1323–1331.
27. Vilcheze C, Llopiz P, Neunlist S, Poralla K, Rohmer M (1994) Prokaryotic triterpenoids: new hopanoids from the nitrogen-fixing bacteria *Azotobacter vinelandii*, *Beijerinckia indica* and *Beijerinckia mobilis*. *Microbiology* 140:2749–2753.
28. Knani M, Corpe WA, Rohmer M (1994) Bacterial hopanoids from pink pigmented facultative methylotrophs (PPFMs) and from green plant surfaces. *Microbiology* 140:2755–2759.
29. Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52:696–704.
30. Abascal F, Zardoya R, Posada D (2005) ProtTest: Selection of best-fit models of protein evolution. *Bioinformatics* 21:2104–2105.
31. Morrison DA, Ellis JT (1997) Effects of nucleotide sequence alignment on phylogeny estimation: a case study of 18S rDNAs of apicomplexa. *Mol Biol Evol* 14:428–441.
32. Wong KM, Suchard MA, Huelsenbeck JP (2008) Alignment uncertainty and genomic analysis. *Science* 319:473–476.
33. Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797.
34. Notredame C, Higgins DG, Heringa J (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* 302:205–217.
35. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673–4680.
36. Lee MSY (2001) Unalignable sequences and molecular evolution. *Trends Ecol Evol* 16:681–685.
37. Talavera G, Castresana J (2007) Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol* 56:564–577.
38. Smith AB (1994) Rooting molecular trees—problems and strategies. *Biol J Linn Soc* 51:279–292.
39. Shimodaira H (2002) An approximately unbiased test of phylogenetic tree selection. *Syst Biol* 51:492–508.
40. Mazon G, et al. (2004) Reconstruction of the evolutionary history of the LexA-binding sequence. *Microbiology* 150:3783–3795.
41. Beiko RG, Harlow TJ, Ragan MA (2005) Highways of gene sharing in prokaryotes. *Proc Natl Acad Sci USA* 102:14332–14337.
42. Tomitani A, Knoll AH, Cavanaugh CM, Ohno T (2006) The evolutionary diversification of cyanobacteria: molecular-phylogenetic and paleontological perspectives. *Proc Natl Acad Sci USA* 103:5442–5447.
43. Williams KP, Sobral BW, Dickerman AW (2007) A robust species tree for the alphaproteobacteria. *J Bacteriol* 189:4578–4586.
44. Pearson A, et al. (2009) Diversity of hopanoids and squalene-hopene cyclases across a tropical land-sea gradient. *Environ Microbiol* 11:1208–1223.
45. Talbot HM, Farrimond P (2007) Bacterial populations recorded in diverse sedimentary biohopanoid distributions. *Org Geochem* 38:1212–1225.
46. Green PN (2006) *The Prokaryotes*, eds M Dworkin, S Falkow, E Rosenberg, KH Schleifer, and E Stackebrandt (Springer, New York), pp 257–265.
47. Imhoff JF (2006) *The Prokaryotes*, eds M Dworkin, S Falkow, E Rosenberg, KH Schleifer, and E Stackebrandt (Springer, New York), pp 41–64.
48. Ward NL, et al. (2009) Three genomes from the phylum *Acidobacteria* provide insight into the lifestyles of these microorganisms in soils. *Appl Environ Microbiol* 75:2046–2056.
49. Ettema TJ, Andersson SG (2009) The alpha-proteobacteria: The Darwin finches of the bacterial world. *Biol Letters* 5:429–432.
50. Abeliovich A (2006) *The Prokaryotes*, eds M Dworkin, S Falkow, E Rosenberg, KH Schleifer, and E Stackebrandt (Springer, New York), pp 861–872.
51. Oren A, Padan E (1978) Induction of anaerobic, photoautotrophic growth in the cyanobacterium *Oscillatoria limnetica*. *J Bacteriol* 133:558–563.
52. Jiao Y, Newman DK (2007) The *pio* operon is essential for phototrophic Fe(II) oxidation in *Rhodospseudomonas palustris* TIE-1. *J Bacteriol* 189:1765–1773.
53. Parales RE, Harwood CS (1993) Regulation of the *pcalJ* genes for aromatic acid degradation in *Pseudomonas putida*. *J Bacteriol* 175:5829–5838.
54. Summons RE, Jahnke LL (1992) *Biomarkers in Sediments and Petroleum*, eds M JM, A P, and PR P (Prentice Hall, Englewood Cliffs, NJ), pp 189–200.
55. Ten Haven HL, Rohmer M, Rullkötter J, Bissleret P (1989) Tetrahymanol, the most likely precursor of gammacerane, occurs ubiquitously in marine sediments. *Geochim Cosmochim Acta* 53:3073–3079.
56. Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* 17:540–552.
57. Shimodaira H, Hasegawa M (2001) CONSEL: For assessing the confidence of phylogenetic tree selection. *Bioinformatics* 17:1246–1247.









**Fig. S4.** Unrooted tree showing the major clusters of HpnP from different bacterial phyla. The alignment was generated by Muscle and maximum likelihood tree inferred using PhyML as described in *Methods*.

**Table S1. Presence of 2-methylhopanoids in bacteria with an HpnP homolog**

Bacterial genome	2-Me C30*	2-Me C35 <sup>†</sup>	Ref.	Isolation habitat
<i>α</i> -Proteobacteria				
<i>R. palustris</i> TIE-1	+	+	(1)	Iron mat
<i>R. palustris</i> CGA009	+	+	(1)	Freshwater sediment
<i>R. palustris</i> BisB18	+	+	This study	Freshwater sediment
<i>R. palustris</i> BisA53	+	+	This study	Freshwater sediment
<i>R. palustris</i> BisB5	+	+	This study	Freshwater sediment
<i>R. palustris</i> HaA2	+	+	This study	Freshwater sediment
<i>Bradyrhizobium</i> sp. ORS278	nt <sup>‡</sup>	nt		Stem nodules
<i>Bradyrhizobium</i> sp. BTAi1	+	–	(2)	Stem nodules
<i>Bradyrhizobium japonicum</i> USDA110	+	–	(2)	Root nodules
<i>Nitrobacter hamburgensis</i> X14	nt	nt		Soil
<i>Nitrobacter winogradskyi</i> Nb-255	nt	nt		Soil
<i>Nitrobacter</i> sp. Nb-311A	nt	nt		Coastal marine waters
<i>O. carboxidovorans</i> OM5	nt	nt		Wastewater
<i>Beijerinckia indica</i> ATCC 9039	+	–	(3)	Acidic soil
<i>M. silvestris</i> BL2	nt	nt		Acidic forest soil
<i>Methylobacterium extorquens</i> AM1 <sup>§</sup>	nt	nt		Airborne contaminant
<i>M. extorquens</i> PA1 <sup>§</sup>	nt	nt		
<i>M. extorquens</i> DM4 <sup>§</sup>	nt	nt		Industrial wastewater
<i>Methylobacterium chloromethanicum</i> CM4 <sup>§</sup>	nt	nt		Contaminated soil
<i>M. nodulans</i> ORS 2060 <sup>§</sup>	nt	nt		Root nodules
<i>Methylobacterium populi</i> BJ001 <sup>§</sup>	nt	nt		Poplar tree
<i>Methylobacterium</i> sp. 4–46 <sup>§</sup>	nt	nt		Root nodules
<i>Methylobacterium radiotolerans</i> JCM 2831 <sup>§</sup>	nt	nt		
Cyanobacteria				
<i>G. violaceus</i> PCC 7421	+	+	(4)	Calcareous rocks
<i>T. elongatus</i> BP-1	nt	nt		Hot spring
<i>Cyanothece</i> sp. 7424 <sup>§</sup>	nt	nt		Rice field
<i>Cyanothece</i> sp. 7425 <sup>§</sup>	nt	nt		Rice field
<i>Cyanothece</i> sp. PCC7822 <sup>§</sup>	nt	nt		Rice field
<i>Nostoc punctiforme</i> PCC73102	+	+	(4)	Gymnosperm symbiont
Acidobacteria				
<i>Candidatus Koribacter versatilis</i> Ellin345	nt	nt		Soil

\*2-MeC30 indicates 2-methylhopenes.

<sup>†</sup>2-MeC35 indicates 2-methylbacteriohopanetetrol.

<sup>‡</sup>Indicates not tested.

<sup>§</sup>2-methylhopanoid production has been shown in other *Methylobacterium* and *Cyanothece* strains (4, 5).

- Rashby SE, et al. (2007) Biosynthesis of 2-methylbacteriohopanepolyols by an anoxygenic phototroph. *Proc Natl Acad Sci USA* 104:15099–15104.
- Bravo JM, et al. (2001) Novel methylated triterpenoids of the gammacerane series from the nitrogen-fixing bacterium *Bradyrhizobium japonicum*. *Eur J Biochem* 268:1323–1331.
- Vilcheze C, et al. (1994) Prokaryotic triterpenoids: new hopanoids from the nitrogen-fixing bacteria *Azotobacter vinelandii*, *Beijerinckia indica* and *Beijerinckia mobilis*. *Microbiology* 140:2749–2753.
- Talbot HM, et al. (2008) Cyanobacterial bacteriohopanepolyol signatures from cultures and natural environmental settings. *Org Geochem* 39:232–263.
- Knani M, Corpe WA, Rohmer M (1994) Bacterial hopanoids from pink pigmented facultative methylotrophs (PPFMs) and from green plant surfaces. *Microbiology* 140:2755–2759.



**Table S3. Sensitivity of rooted HpnP topology to alignment parameters**

Trial	Total taxa incl. outgroup(s)	Alignment program	Filtering by Gblocks	No. of sites	-lnL	Best topology*
1	83	ClustalW	None	622	35,976.33	C
2	83	ClustalW	Stringent	399	21,615.26	C
3	83	ClustalW	Relaxed	469	27,187.91	A
4	118	ClustalW	None	678	57,342.19	A
5	118	ClustalW	Stringent	280	22,884.90	A
6	118	ClustalW	Relaxed	422	38,833.93	A
7	83	Muscle	None	626	36,221.90	A
8	83	Muscle	Stringent	392	20,958.59	A
9	83	Muscle	Relaxed	468	26,874.87	B
10	118	Muscle	None	704	58,004.70	B
11	118	Muscle	Stringent	306	25,363.81	A
12	118	Muscle	Relaxed	411	36,969.22	A
13	83	T-Coffee	None	689	35,924.79	A
14	83	T-Coffee	Stringent	382	20,832.01	A
15	83	T-Coffee	Relaxed	452	25,723.67	A
16	118	T-Coffee	None	812	57,199.16	B
17	118	T-Coffee	Stringent	312	25,942.82	A
18	118	T-Coffee	Relaxed	394	34,708.95	A

\*A indicates a topology similar to Fig. 4A, with early branching cyanobacteria. B, topology similar to Fig. 4B, with monophyletic cyanobacteria sister to the  $\alpha$ -proteobacteria. C, topology similar to Fig. 4C, with cyanobacteria branching within the  $\alpha$ -proteobacteria. We did not consider the position of the lone acidobacterium sequence as it was not well-supported in most trials.

**Table S4. Comparing alternative tree topologies using Shimodaira's Approximately Unbiased test**

Alignment	No. of sites	-lnL for topology*			p-value for topology†		
		A	B	C	A	B	C
ClustalWno filtering	622	36,001.91	35,997.57	35,976.33	0.20	0.26	0.81
ClustalW relaxed	469	27,187.91	27,199.24	27,199.77	0.69	0.32	0.38
ClustalW stringent	399	21,621.43	21,634.54	21,615.26	0.46	0.18	0.68
Muscleno filtering	626	36,221.90	36,244.39	36,259.22	0.84	0.24	0.10
Musclerelaxed	468	26,884.17	26,874.87	26,899.22	0.41	0.73	0.16
Musclestringent	392	20,958.59	20,982.64	20,964.40	0.72	0.14	0.42

\*For each alignment, the best tree was found by PhyML and its likelihood is shown in bold. The two alternative topologies were fixed for each alignment to topology A, B, or C from Fig. 4 and then branch lengths and substitution parameters were optimized on that topology, giving the likelihood values shown. These three trees were then compared using the AU test.

†p-values <0.05 allow us to reject that tree as the most likely. If no p-value is <0.05, then all three trees belong to the confidence set and the true tree could be any one of them.

**Table S5. Strains and plasmids used in this study**

Strains or plasmids	Genotype, markers, or characteristics
<b>Strains</b>	
<i>E. coli</i> S17-1	<i>thi pro hdsR hdsM<sup>+</sup> recA</i> ; chromosomal insertion of RP4-2 (Tc::Mu Km::Tn7)
<i>R. palustris</i> TIE-1	Isolated from Woods Hole, MA
<i>R. palustris</i> $\Delta$ hpnP	<i>R. palustris</i> TIE-1 $\Delta$ hpnP
<i>R. palustris</i> $\Delta$ hpnP + pPVW24	<i>R. palustris</i> TIE-1 $\Delta$ hpnP transformed to Gm <sup>r</sup> with pPVW24
<b>Plasmids</b>	
pJQ200SK	Mobilizable suicide vector; <i>sacB</i> Gm <sup>r</sup>
pBBR1MCS-5	5.1-kb broad host range plasmid; Gm <sup>r</sup> <i>lacZ</i>
pPVW9	1-kb <i>hpnP</i> upstream region PCR amplified with primers PW39 and PW40 and cloned into pJQ200SK at <i>NotI</i> and <i>XbaI</i>
pPVW10	1-kb <i>hpnP</i> downstream region PCR amplified with primers PW41 and PW42 and cloned into pJQ200SK at <i>XbaI</i> and <i>SpeI</i>
pPVW17	<i>XbaI</i> and <i>SpeI</i> cut <i>hpnP</i> downstream fragment from pPVW10 subcloned into pPVW9
pPVW24	3.5-kb <i>hpnP</i> complementation fragment PCR amplified with primers PW85 and PW86 and cloned into pBBR1MCS-5 at <i>HindIII</i> and <i>SpeI</i>
Gm, gentamicin	

**Table S6. Data for hopanoid analysis of *R. palustris* strains by high-temperature GC-MS**

Analyte	Rt*	M <sup>+</sup> (RI)	Base peak	Characteristic fragments (RI)
2-methylhopene	26.45	424(18)	189	205(73), 355(23), 381(25)
hopene	26.59	410(18)	189	191(93), 341(27), 367(23), 410(19)
2-methyldiplopterol acetate	27.68	484	149	189(92), 205(80), 381(5), 424(4)
diplopterol acetate	27.82	470	191	149 (93), 189(95), 367(5), 410(4)
2-methyltetrahymanol acetate	28.49	484	205	189(32), 263(5), 355(4), 424(4)
tetrahymanol acetate	28.64	470(6)	191	189(46), 249(6), 395(5), 410(4)
2-methylBHtetrol acetate	42.84	728(1)	205/95	253(10), 383(19), 433(5), 493(38), 713(1)
BHtetrol acetate	43.81	714(.5)	191	253(9), 369(18), 433(4), 493(34), 699(1)

\*Retention time (min) on 30 m DB-XLB column and using an Agilent 6890GC/5975MSD.