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

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.
Identification of a methylase required for 2-methylhopanoid production and implications for the interpretation of sedimentary hopanes

Paula V. Welander¹,¹, Maureen L. Coleman¹,¹, Alex L. Sessions⁵, Roger E. Summons¹, and Dianne K. Newman¹,⁴,²

¹Department of Biology, Massachusetts Institute of Technology, 77 Massachusetts Avenue, 68-380, Cambridge, MA 02139; ²Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, MC100-23, 1200 East California Boulevard, Pasadena, CA 91125; ³Department of Earth, Atmospheric and Planetary Science, Massachusetts Institute of Technology, 77 Massachusetts Avenue, E25-633, Cambridge, MA 02139; and ⁴Howard 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-methylbacteriohopanepolys (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 α-proteobacteria and acido-bacteria. 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 α-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. Sedimentary hopanes are biomarkers whose distribution has been studied extensively. These hydrocarbon molecules are recognized as the diagenetic products of bacteriohopanepolys, 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, Rhodopseudomonas 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-methylhopanoids 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


The authors declare no conflict of interest. This article is a PNAS Direct Submission.

¹P.V.W. and M.L.C. contributed equally to this work. E-mail: dkn@mit.edu.

This article contains supporting information online at 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. L4). 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. L4). 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* methylene 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 α-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 α-proteobacteria, two γ-proteobacteria, two δ-proteobacteria, two β-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-methylhopanones 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-methylhopanones (*hpnP*-22(29)-ene, *hpnP*-17(21)-ene, and *hpnP*-21-ene), 2-methylbacteriohopanato-l, 2-methyltetrahymanol, and 2-methylbacteriohopanetetrol.
strategies (relaxed, stringent, and no filtering) to remove uncer-
tain alignment positions (Fig. S2). Although filtering may in-
crease 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 *Methyllobacterium* species, the *Bradyrhizo-
bium/Nitrobacter/Rhodopseudomonas* group, and the cyanobac-
teria (Fig. S4). The branching order of these groups, how-
ever, varied significantly between trials. Three different rooted
topolologies emerged (Fig. 4 and Table S3). Most commonly,
the α-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 an-
cestral HpnP resided in a cyanobacterium (Fig. 4A). In some
cases, the cyanobacterial and α-proteobacterial sequences
formed sister clades (Fig. 4B), or the cyanobacterial sequences
were nested within the α-proteobacterial sequences (Fig. 4C).
This ambiguity suggests that there is insufficient phylogenetic sig-
nal to confidently resolve the branching order of the major clades.
We tested whether the best topology recovered for a given align-
ment 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 phylogenetic distribu-
tion could arise two ways: either it was present in the ancestor of
cyanobacteria and α-proteobacteria and was repeatedly lost,
or it was horizontally transferred between cyanobacteria and
α-proteobacteria. Gene transfer between these phyla has been
observed (40), and cyanobacteria most often share genes with the
α-proteobacteria compared to other phyla (41). Within the cy-
nobacteria, HpnP is sporadically distributed and the HpnP phy-
logeny 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 cy-
nobacteria produced 2-methylhopanoids. The position of the
lone acido bacterium in the HpnP tree is also uncertain. Within
the α-proteobacteria, the HpnP and species phylogenies are
similar (43), implying vertical descent within this group. From
this phylogenetic analysis, it 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 indica-
tors 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 α-proteobacteria (specifically the rhizobiales)
and one acidobacterium.

The presence of this methylase in the rhizobiales and acidobac-
teria reinforces the need for caution in using 2-methylhopanoids
as indicators of cyanobacteria in modern environments. For ex-
ample, 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 pre-

cence 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 acidobac-
teria 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
methylation was most likely found in cyanobacteria, α-proteobac-
teria, or acidobacteria. However, the small number of HpnP

![Fig. 2. GC-MS total ion chromatograms of acetylated total lipid extracts of *R. palustris* cultures. The Δ*hpnP* mutant does not produce any C-2 methylated hopanoids. (A) *R. palustris* TIE-1. (B) *R. palustris* Δ*hpnP*. (C) *R. palustris* Δ*hpnP* complemented with the *hpnP* gene on a self-replicating plasmid. Numbered compounds: I, 2-methylhopenes; II, hopenes; III, 2-methyl/diplopterol; IV, di-
protopterol; V, 2-methyltetrahymanol; VI, tetrahymanol; VII, 2-methylbacteria-
hopenetetrol; 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
Δ*hpnP* strain did produce 2-methylbacteriohopanetetrol (compound VIII)
as verified by analysis of the 205 Da mass chromatogram and the mass spec-
trum; however, the peak is too broad to be seen at the resolution of this fig-
ure of the GC-MS total ion current.

methyllase in a yet untested marine cyanobacterium. However,
the α-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 ProtTest (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, Mus-
cle, and T-Coffee) (33–35). In addition, we used three filtering
strategies (relaxed, stringent, and no filtering) to remove uncer-
tain alignment positions (Fig. S2). Although filtering may in-
crease 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).
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 phyyla may have been responsible for the deposition of ancient sedimentary 2-methylhopanoids. 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 photoautotrophs such as cyanobacteria (10, 19). Although the majority of the α-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 α-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-methylhopanoids 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-methylhopanoids with confidence.

In conclusion, four lines of evidence suggest that 2-methylhopanoids 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-methylhopanoids 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. Rhodopseudomonas 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 μg/mL (E. coli) or 800 μg/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.
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-XLB column as previously described (15). Compounds were identified by comparison of retention times and mass spectra to authentic compounds (tetrahymanol 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 using ClustalW, Muscle, or T-Coffee (56). Alignments were trimmed by Muscle and 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 GBloks. The appropriateness of the two outgroup families was confirmed by a more exhaustive tree of related prokaryotic C-2 methyltransferases. The appropriateness of the two outgroup families was confirmed by a more exhaustive tree of related prokaryotic C-2 methyltransferases.

Evolutionary relationships among HpnP homologs were inferred from the aligned C-2 methyltransferases. The outgroup families were Brevundimonas sp.; diplopterol (56) 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. Maximum likelihood trees were constructed by PhyML (29) using the LG 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 GBloks 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 sho 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.W.V.), and an Agouron Institute Geobiology Postdoctoral Fellowship (M.L.C.). D.K.N. is an Investigator of The Howard Hughes Medical Institute.


