Environmental Microbiology (2009) 11(5), 1208–1223 doi:10.1111/j.1462-2920.2008.01817.x

Diversity of hopanoids and squalene-hopene cyclases across a tropical land-sea gradient

Ann Pearson,[1]* William D. Leavitt,[1,2] James P. Sáenz,[3,4] Roger E. Summons,[3] Mandy C.-M. Tam[1] and Hilary G. Close[1]
Departments of [1]Earth and Planetary Sciences, and [2]Organismic and Evolutionary Biology, Cambridge, MA 02138, USA.
[3]Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.
[4]Department of Marine Chemistry and Geochemistry, Woods Hole Oceanographic Institution, Woods Hole, MA 02543, USA.

Summary

Bacterial hopanoids are ubiquitous in Earth surface environments. They hold promise as environmental and ecological biomarkers, if the phylogeny and physiological drivers of hopanoid biosynthesis can be linked with the distribution of hopanoids observed across a breadth of samples. Here we survey the diversity of hopanoid cyclases from a land-sea gradient across the island of San Salvador, in the easternmost part of the Bahamas. The distribution of lipids was determined for the same sites, for the first time overlaying quantification of bacteriohopanepolyols with *sqhC* phylogeny. The results are similar to previous reports: environmental *sqhC*s average < 65% translated amino acid identity to their closest named relatives, and sequences from putative *Proteobacteria* dominate. Additionally, a new and apparently ubiquitous group of marine hopanoid producers is identified; it has no identifiable close relatives. The greatest diversity of hopanoid lipids occurs in soil, but hopanoids represent a minor fraction of total soil-derived lipids. Marine samples contain fewer identifiable hopanoids, but they are more abundant as a fraction of the total extractable lipids. In soil, the dominant compounds are 35-aminobacteriohopane-32,33,34-triol and adenosylhopane. In an upper estuarine sample, bacteriohopanetetrol and 32,35-anhydrobacteriohopanetetrol dominate; while in lower estuarine and open marine samples, the most abundant are bacteriohopanetetrol and bacteriohopaneribonolactone. Cyclitol ethers are trace components in the soil, absent in the estuary, and of moderate abundance in the open marine setting, suggesting a dominant marine source. Conversely, aminotriol and aminotetrol decrease in abundance or disappear completely from land to ocean, while 2-methyldiplopterol shows the opposite trend. Small quantities of 2-methylbacteriohopanepolyols are detectable in all samples. The overall hopanoid distributions may correlate to the major phylogenetic families of hopanoid producers or to the environments in which they are found.

Received 12 August, 2008; accepted 7 October, 2008. *For correspondence. E-mail pearson@eps.harvard.edu; Tel. (+1) 617 384 8392; Fax (+1) 617 496 4387.

Introduction

A primary goal in geobiology is to reconstruct the co-evolution of life and environment. Ancient records of biological activity are obtained from lipid biomarkers in sedimentary rocks (Brocks *et al.*, 1999; Summons *et al.*, 1999; Dutkiewicz *et al.*, 2006). The most valuable of these lipids are both taxonomically and environmentally diagnostic, and microbial lipids called hopanoids are abundant among these molecules (Ourisson and Albrecht, 1992). Hopanoids are isoprenoid lipids which derive primarily from bacteria and have a fossil record that extends back billions of years. Their presence may reflect local biogeochemical conditions and the active microbial populations present at the time of deposition. Bacteriohopanepolyols (BHPs) have been proposed to be functional analogues to eukaryotic sterols (Rohmer *et al.*, 1979; Ourisson *et al.*, 1987), based primarily on their polycyclic structures, amphipathic properties and molecular size. In addition, these molecules may confer resistance to oxidative stress, dessication and/or heat (Berry *et al.*, 1993; Poralla *et al.*, 2000; Bosak *et al.*, 2008). In recent years, much effort has been devoted to understanding the distribution of unique polar head groups of BHPs, in order to identify new structures and to place environmental and/or phylogenetic constraints on the sources of hopanoids in sediments (Talbot *et al.*, 2003a,b; Talbot *et al.*, 2007a,b; 2008; Bednarczyk *et al.*, 2005; Sugden *et al.*, 2005).

Technical advancements in the analysis of complex BHP have been paralleled by culture-independent techniques to identify new microbial phylotypes. These have

© 2008 The Authors
Journal compilation © 2008 Society for Applied Microbiology and Blackwell Publishing Ltd

exposed a vast expanse of uncharacterized diversity (Hugenholtz and Pace, 1996) with unknown genetic potential for lipid biosynthesis and profound implications for our understanding of lipid biomarker proxies. This is because uncultured phylotypes may harbour an under-appreciated diversity of microbial lipids, with no easy means by which to characterize their organismic origins. Hopanoids are synthesized from the acyclic isoprenoid precursor, squalene (Woodward and Bloch, 1953) with the initial step catalysed by the enzyme squalene-hopene cyclase (SHC; Wendt *et al.*, 1997). SHCs can be identified readily based on their numerous conserved amino acids (Hoshino and Sato, 2002; Summons *et al.*, 2006; Fischer and Pearson, 2007) and specific functional motifs (Feil *et al.*, 1996; Wendt *et al.*, 2000). SHCs show a primarily vertical pattern of inheritance with limited lateral gene transfer (LGT). Trees based on sequences of 16S rRNA genes are similar to trees based on the cyclases (Pearson *et al.*, 2007; Frickey and Kannenberg, 2008). Some cases of gene duplication and/or instances of LGT exist, but they primarily are confined to the δ-*Proteobacteria*, which inevitably contain two copies of the gene squalene-hopene cyclase (*sqhC*) per genome, each apparently of different origin. Most other taxa maintain one copy of *sqhC*. We have recently estimated that fewer than one in 10 bacterial cells in soils and fewer than one in 20 bacterial cells in the ocean contains an *sqhC* (Pearson *et al.*, 2007; Pearson and Rusch, 2008). Among cultured bacteria that have had their genomes fully sequenced, approximately one in 10 species has a copy of *sqhC*.

The ability to synthesize hopanoids appears to be an uncommon physiological trait among bacteria, occurring less frequently than earlier studies suggested (Rohmer *et al.*, 1984). The presence of the gene in an individual species always correlates with the eventual detection of hopanoids in that species (Rohmer *et al.*, 1984; Perzl *et al.*, 1998; Hartner *et al.*, 2005) although in at least one case it is expressed only during a specific life stage of the organism (Poralla *et al.*, 2000), and in another case the gene was shown to encode a cyclase for regular polyprenes, not squalene (Bosak *et al.*, 2008; Kontnik *et al.*, 2008). It therefore remains of considerable interest to explore the taxonomy of bacterial hopanoid production and to link this to patterns of BHP distribution and abundance in natural environments. A better understanding of which organisms produce BHPs in contemporary marine communities, the specifics of their structures, and the associated environmental conditions will help to constrain geobiological interpretations. For example, ancient carbonate environments harbour large quantities of geohopanes (Knoll *et al.*, 2007; Eigenbrode *et al.*, 2008), including hopanes with a 2-methyl substituent. Their precursors, 2-methylbacteriohopanepolyols (2-MeBHPs) are common in *Cyanobacteria* (Rohmer *et al.*, 1984; Simonin *et al.*, 1996; Summons *et al.*, 1999; Talbot *et al.*, 2008). Due to the high 2-methylhopane concentrations found in association with carbonates and evaporites, these environments are proposed to record high densities of *Cyanobacteria* (Summons *et al.*, 1999). The recent discovery of 2-MeBHP in *Rhodopseudomonas palustris*, when grown in culture as an anoxygenic phototroph (Rashby *et al.*, 2007) complicates this interpretation. Thus, modern authigenic carbonate platforms are important sites for investigating the diversity of *sqhC* genes and associated BHPs.

Here we report the diversity of *sqhC* sequences in environmental samples from a tropical, carbonate-rich environment. Samples were taken along a land-sea gradient across the Bahamian island of San Salvador. Environments surveyed were an upland soil, waters from upstream and downstream in a tidal creek and the shallow open ocean (Fig. 1). We utilized our third-codon degenerate primer sets (Pearson *et al.*, 2007) to target conserved motifs of the translated amino acid sequence of SHC. These primers were designed to target almost the entire bacterial Domain, with the exception of the highly divergent *Planctomycetes*. Simultaneously, we profiled the distribution of free hopanols (e.g. diploptene, diplopterol and the related tetrahymanol) and functionalized BHPs in sediments accumulating in the same locations. The results increase the known diversity of hopanoid producers in natural communities and illuminate challenges for deciphering the global distribution of microbially derived lipids from complex communities.

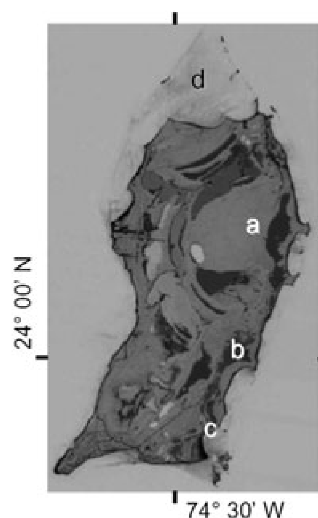

Fig. 1. Satellite image of San Salvador, modified from http://www.geraceresearchcentre.com/sansal.html and from Adams (1980). Sampling sites are labelled on and around San Salvador, Bahamas, showing (a) Soil, (b) Pigeon Creek Upper, (c) Pigeon Creek Lower and (d) Graham's Harbour.

© 2008 The Authors

Table 1. Bahamas *sqhC* clone libraries.

Sample	cyR	prA	prB	Total	Unique[a]
Soil	24u/28t	3u/19t	4u/33t	80	31
Pigeon Creek Upper	2u/27t	4u/11t	2u/16t	54	8
Pigeon Creek Lower	11u/32t	–	4u/20t	52	15
Graham's Harbour	3u/26t	–	1u/1t	27	4

a. Defined as < 98% identical at the amino acid level and counted as unique within a sample, yielding 58; across all samples, 3 of these are the same, resulting in 56 unique overall.

Results and discussion

Squalene-hopene cyclases

Primer sets cyR, prA and prB generated 58 unique (u) sequences out of 213 total (t) clones from the four geochemically distinct environments (Table 1). Primer set pr_R generated no sequences, but previous work suggests this primer targets primarily aerobic methanotrophs (Pearson *et al.*, 2007), which are not expected to be abundant in these samples. Classification of the amplicons by tBLASTn suggested that most sequences amplified from the Bahamas samples were apparent relatives of α-*Proteobacteria*, δ-*Proteobacteria* and *Acidobacteria*. Among these, sequences nominally related to α-*Proteobacteria* increased in relative abundance from land to ocean (Fig. 2). Conversely, δ-*Proteobacteria*-like sequences progressively decreased in relative abundance from land to ocean. Minor numbers of putative *Actinomycetes* and *Cyanobacteria* also were detected, as well as a sequence potentially similar to the anammox *Planctomycetes*, *Kuenenia stuttgartiensis*, whose SHC is different from all other known *Planctomycetes*. These were found primarily in the soil, which was also the sample that demonstrated the highest overall sequence diversity (Table 1). It is likely, however, that sequences of β- and λ-*Proteobacteria* were undersampled, due to potential biases in the amplification efficiency of the primers. *SqhC*s of β- and λ-*Proteobacteria* are found in metagenomic data obtained from similar environments (Pearson and Rusch, 2008) but are less common among the PCR amplicons. The distribution of other taxa identified in the PCR data is consistent with the relative distribution found in metagenomic data, including the marine trend toward α-*Proteobacteria* (Pearson and Rusch, 2008).

Although the majority of the total sequences group with α- and δ-*Proteobacteria* and *Acidobacteria* by tBLASTn, most likely are from unknown genera, families or even uncultivated major lineages of *Bacteria*. The sequences affiliated putatively with α-*Proteobacteria* average 64% identity to their closest relatives, while those affiliated with δ-*Proteobacteria* average 58% and those with

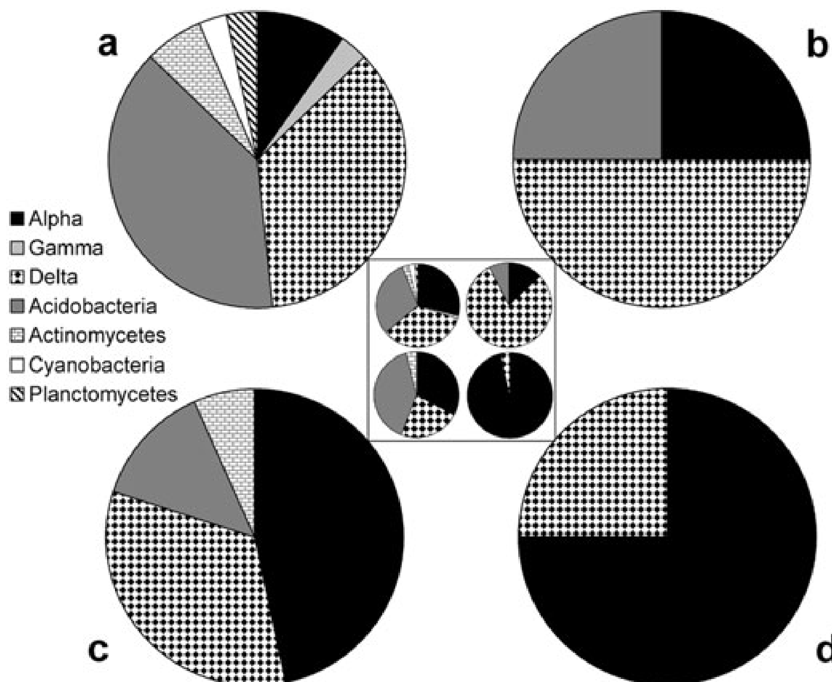

Fig. 2. Distribution of clonal phylogeny as classified by tBLASTn. Pie charts show clone distributions from (a) Soil, (b) upper Pigeon Creek, (c) lower Pigeon Creek and (d) Graham's Harbour. Small inset pies show the total number of clones obtained, while large pies represent the distribution of unique clones. See Table S1 for further sequence statistics.

© 2008 The Authors
Journal compilation © 2008 Society for Applied Microbiology and Blackwell Publishing Ltd,

Acidobacteria only 55% (Appendix S1, Table S1). Identities > 60% are typical for SHCs from organisms within the same major phylogenetic group (Fischer and Pearson, 2007). This suggests that the sequences associated here with α-*Proteobacteria* could represent uncultivated families of α-*Proteobacteria*. In contrast, the low identities for the other taxonomic groups (< 60%) suggest the phylogenetic affiliations assigned by tBLASTn are more uncertain. This highlights the fact that tBLASTn cannot account for bacterial divisions that have not yet been grown in culture or are not represented in genome sequencing projects (e.g. SAR-86; Fuhrman *et al*., 1993). All of the data have been forced to bin with the 'nearest' genomic data, regardless of the magnitude of that distance.

Further exploration by Bayesian inference shows a more detailed phylogeny (Fig. 3). Also included in Fig. 3

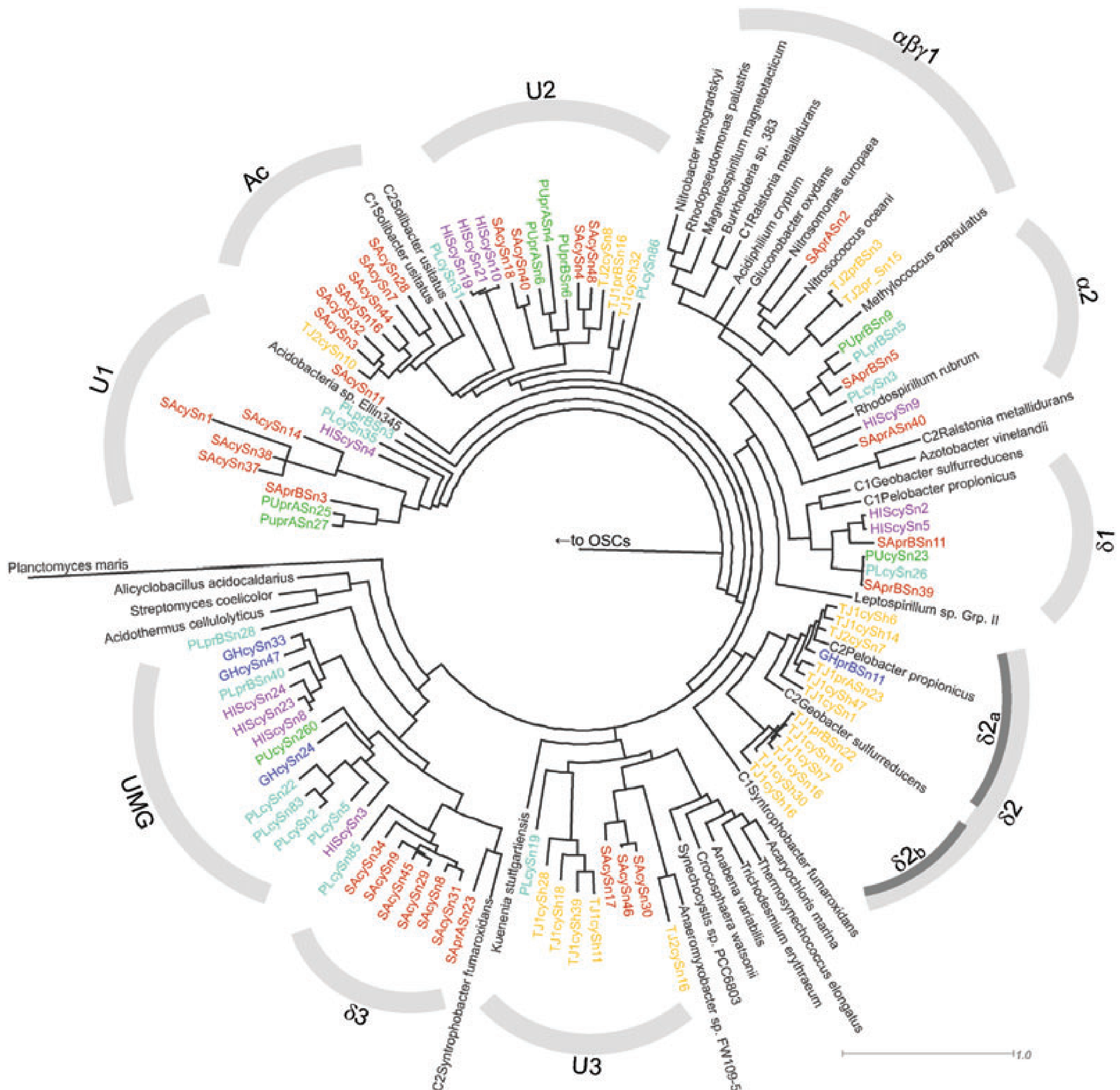

Fig. 3. Bayesian phylogenetic tree of all SHC sequences from environmental *sqhC* PCR amplicons. The tree includes all sequences from this work and also incorporates all sequences from Pearson and colleagues (2007). Sample origins are colour-coded: SA (upland soil, brown), PCU (upper Pigeon Creek, green), PCL (lower Pigeon Creek, aqua), GH (Graham's Harbour, blue), HIS (Hawaii surface, purple; Pearson *et al*., 2007), TJ (Lake Mishwam, orange; Pearson *et al*., 2007). After calculating for 2 million generations, final posterior probabilities for the overall tree were < 0.05 (similar to bootstrap > 95%) and at all individual nodes were < 0.50 (similar to bootstrap values > 50%).

© 2008 The Authors

are all sequences previously obtained from a fresh water lake and from surface waters of the Pacific Ocean, near Hawaii (Pearson *et al*., 2007). Genomic sequences for comparison include one species of each genus currently known to contain a squalene cyclase gene. The Bayesian analysis suggests that in some cases, the phylogeny indicated by the original tBLASTn classification may be incorrect, while in other cases, groups of environmental amplicons cluster with known taxa and have sequence identity that indicates they are affiliated with these taxa.

Acidobacteria species *Solibacter usitatus* and *Acidobacteria* sp. Ellin345 are soil bacteria. Several environmental SHCs cluster with the former and have average sequence identity of 66% with *S. usitatus* (Group Ac, Fig. 3). These amplicons primarily are from the Bahamas soil but also include one relative from the fresh water lake. A second group, while it includes the single-copy SHC from *Acidobacteria* sp. Ellin345, is highly divergent and is not structured as a coherent clade (Group U1). Sequences in this cluster have only 48% average identity to their closest relative. In some cases tBLASTn classifies this 'closest relative' not as *Acidobacteria*, but as δ-*Proteobacteria*. The low sequence identities, however, suggest that the reason for this discrepancy is that these SHCs derive from neither *Acidobacteria* nor δ-*Proteobacteria*, but instead represent lineage(s) not represented in GenBank to date. Sequences in this group come from the Bahamas soil, both Pigeon Creek samples, and also include one instance from Hawaii surface waters.

A second unknown cluster (U2), has no reference sequence from any genome. This prevents speculation as to the potential phylogenetic affiliation of these amplicons. They have 52% average identity to their nearest relatives, and tBLASTn alternately chooses among four options with no apparent pattern: *Acidobacteria*, δ-*Proteobacteria*, anammox *Planctomycetes* and *Actinomycetes* (Table S1). All locations other than Graham's Harbour (GH) also are represented within this cluster, indicating that these SHCs of unknown affiliation are geographically and environmentally widespread.

Among α-, β- and γ-*Proteobacteria* there is a cluster of SHCs affiliated with α-*Proteobacteria* that is phylogenetically incoherent (α2), grouping separately from other α-, β- and γ-*Proteobacteria* (αβγ1). This is consistently observed among SHC trees constructed both by maximum likelihood methods and Bayesian analysis, and regardless of the inclusion or exclusion of environmental sequences (Pearson *et al*., 2007). A possible cause is an ancient LGT that created group α2, which then persisted through time. Environmental amplicons in group αβγ1 have 61% identity to their closest relatives and specifically include aerobic methanotrophs from the fresh water meromictic lake (Pearson *et al*., 2007) and an apparent nitrosifyer from the Bahamas soil. Sequences in group α2 are widely found (Bahamas soil, both Pigeon Creek samples, Hawaii surface waters) and have 62% identity to *Rhodospirillum rubrum*, the genomic representative for this cluster. This cluster also is highly represented in the Global Ocean Sampling expedition metagenome (Rusch *et al*., 2007; Pearson and Rusch, 2008).

Most δ-*Proteobacteria* have two copies of SHC-encoding genes. One copy places coherently with other *Proteobacteria*, basal to α-, β- and γ-*Proteobacteria* (δ1). Environmental amplicons clustering with this group are found in Bahamas soil, both Pigeon Creek samples, and Hawaii surface waters and have 63% identity to *Geobacteraceae*. A different set of environmental sequences is found to cluster with the Copy 2 SHC of *Geobacter* and *Pelobacter* spp., representing the fresh water lake and GH samples. The average identity within this cluster (δ2) is 71%; although it contains two sub-clusters, one related to *Pelobacter* spp. (average 81% identity; $\delta 2_a$) and one more distant cluster (average 60% identity; $\delta 2_b$). A third cluster of putative SHCs of δ-*Proteobacteria* (δ3) is related to the Copy 2 gene of *Syntrophobacter* spp. at 62% identity. Phylogenetically, however, this group is distant from other *Proteobacteria*, aligning as a major division of the branch that also contains *Actinomycetes* (high GC content Gram-positive *Bacteria*). Thus group δ3 potentially records another case of ancestral LGT (Frickey and Kannenberg, 2008).

The PCR protocol used here should amplify *sqhC*s of environmental *Cyanobacteria*, as verified by our recent detection of cyanobacterial *sqhC* in a cryptogamic (desert) soil (Y. Pi, C. Hayes and A. Pearson, unpublished). However, here no amplicons are affiliated with *Cyanobacteria*, and similarly, zero cyanobacterial *sqhC*s are detected in all but one of the samples from the Global Ocean Sampling expedition metagenome (Rusch *et al*., 2007; Pearson and Rusch, 2008). Together, these results suggest that *sqhC* genes affiliated with *Cyanobacteria* are fractionally a small component of the total hopanoid cyclase gene content in contemporary marine environments.

A major cluster of sequences radiates basal to *Cyanobacteria* and includes one genomic affiliate, Anaeromyxobacter sp. FW109-5 (δ-*Proteobacterium* of the *Myxococcales*). This again indicates phylogenetic incoherence resulting from LGT among the multiple *sqhC* copies of δ-*Proteobacteria*. Within the environmental amplicons that group with this sequence (U3), tBLASTn classifies their 'closest relatives' alternately as δ-*Proteobacteria*, *Cyanobacteria* or *Actinomycetes*. Their average identity to any of these taxa is only 49%, however, suggesting this again represents an uncharacterized cluster of environmental *Bacteria* or additional cases of *sqhC* acquired by LGT. These sequences derive

© 2008 The Authors

primarily from the fresh water lake and the Bahamas soil, with one sequence from Pigeon Creek, suggesting their source primarily is terrigenous.

Finally, an unusual set of sequences groups with the branch that includes *Actinomycetes* (UMG, Unknown Marine Group). All of the marine aquatic sites are represented (GH, both Pigeon Creek samples and Hawaii surface waters), while none of the fresh water or soil samples appears. By tBLASTn, each amplicon invariably classifies with α-*Proteobacteria* (average 64% identity). Yet phylogenetically, they form a well-supported cluster distant from all other α-*Proteobacteria*. The reason for this discrepancy remains unknown. More information is critical about the origin of UMG sequences and the hopanoids produced by their hosts, as the data suggest this group is a marine-specific source of hopanoids. The UMG group also is responsible for much of the apparent increase in 'α-*Proteobacteria*' (Fig. 2) across the land-sea gradient.

Throughout Fig. 3, the majority of SHCs from named species cluster with other named species. In contrast, most environmental clones count other clones as their closest relatives, even when representing different geographies (e.g. Bahamian and Hawaii sequences appearing as closest relatives). This suggests that in addition to phylogenetic distance from known species, the total diversity of *sqhC*s is far greater than currently known and remains poorly constrained by culture collections. Many of the groups are represented by multiple clones from diverse locations (e.g. U2, UMG) yet nothing is known about their taxonomic identity within the *Bacteria*, more than 99.9% of which remain uncultured (Hugenholtz and Pace, 1996; Rappe and Giovannoni, 2003).

Bacteriohopanepolyols

The soil sample was a dry peat, composed primarily of organic detritus and having very low mineral content. Twelve different BHP structures could be identified in the soil TLE by APCI-HPLC-MS^n (Fig. 4; Fig. A1). The three BHPs bacteriohopanetetrol (BHT; **If**), 35-aminobacteriohopane-32,33,34-triol (**Ig**), and adenosylhopane (**Ip**), comprised 21%, 30% and 20% of total BHPs respectively (Fig. 5A, Table 2). The primary BHT was accompanied by a small amount of a second isomer (**If′**). Although we do not have an authentic standard, the less common epimer (32*R*, 33*R*, 34*R*)-bacteriohopanetetrol has been reported in *Acetobacter* and *Nostoc* spp. (Peiseler and Rohmer, 1992; Zhao *et al.*, 1996) along with the usual (32*R*, 33*R*, 34*S*)-bacteriohopanetetrol, and could account for the presence of **If′** in addition to **If**.

At least eight other minor BHPs could be detected in the soil TLE. These include two isomers of

Table 2. Abundance of BHPs and free hopanoids (μg compound/mg TLE).

Free hopanoids	**III**	**Ia**	**IV**	**IIb**	**Ib**	[a]**Ie′**	**Id**	**IIe**	**Ie**	Total free
Soil	0.00	0.84	*i*	*i*	*i*	*i*	*i*	*i*	*i*	n.a.
Pigeon Creek Upper	2.13	0.00	1.71	0.06	0.25	0.23	0.47	0.12	7.05	12.0
Pigeon Creek Lower	0.10	0.25	0.21	0.44	0.47	0.07	0.03	*i*	1.84	3.4
Graham's Harbour	0.04	0.25	0.25	1.80	0.95	0.01	0.03	*i*	1.28	4.6

BHPs	**If**	**If′**	**IIf**	**IIf′**	**Ig**	**Ih**	**Ii**	**Ik**	**Ij**	**Ij′**	**Im**	**Iq**	**Io**	**Io′**	**IIo**	**Ip**	Total BHPs
Soil	0.85	0.09	0.09	0.20	1.21	0.10	0.21	0.00	0.05	0.05	0.00	0.00	0.00	0.32	0.02	0.78	4.0
Pigeon Creek Upper	2.20	0.00	0.31	0.37	0.41	0.00	0.00	0.00	0.00	0.00	0.00	2.04	0.34	0.83	0.00	0.00	6.5
Pigeon Creek Lower	25.60	0.34	1.37	2.57	4.41	2.60	0.00	3.20	0.00	0.00	2.99	0.00	1.23	23.89	0.47	0.00	68.7
Graham's Harbour	33.70	0.00	3.05	2.62	1.80	6.39	0.00	2.85	0.00	0.00	3.19	3.51	1.10	26.08	0.58	0.00	84.9

Ratios	Me-hopanoids Free	Me-hopanoids BHPs	Amino-hopanoids BHPs	Lactone + anhydro BHPs
Soil	n.a.	0.08	0.36	0.08
Pigeon Creek Upper	0.02	0.11	0.06	0.49
Pigeon Creek Lower	0.15	0.06	0.06	0.37
Graham's Harbour	0.44	0.07	0.02	0.36

© 2008 The Authors

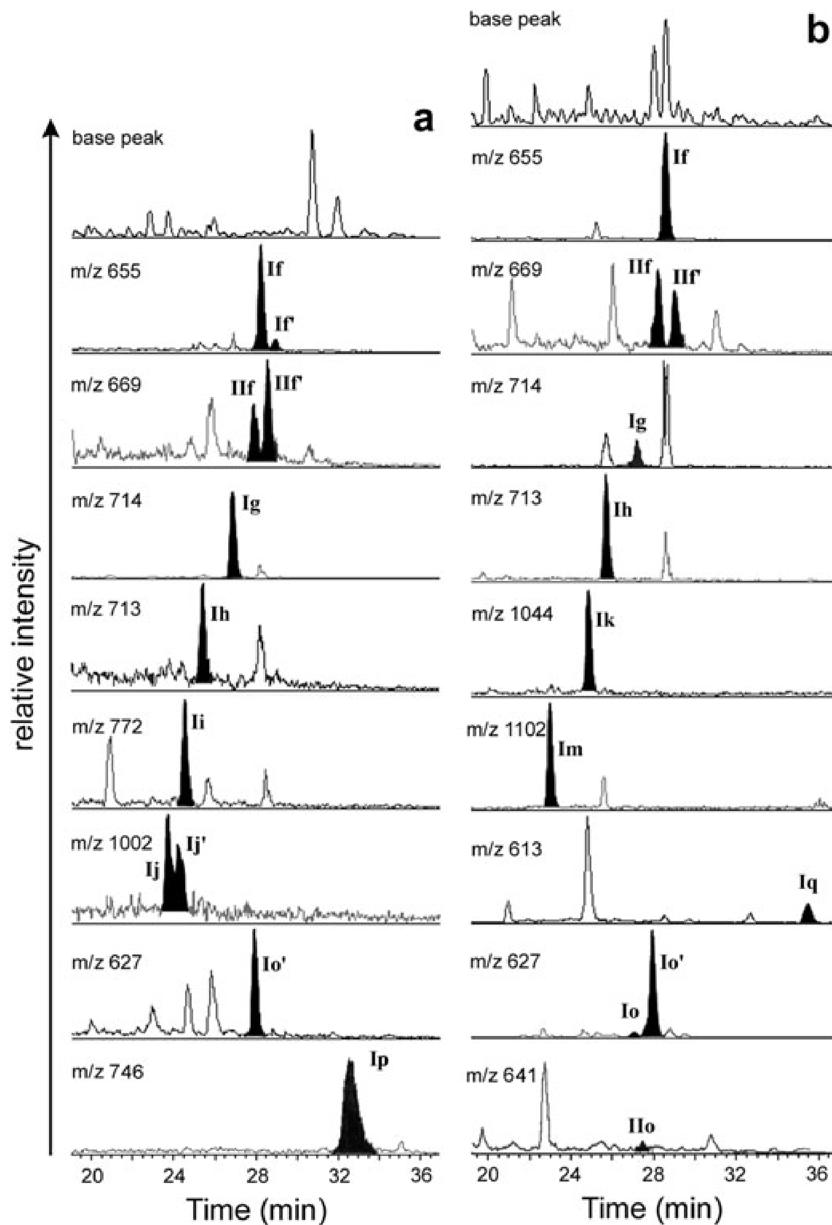

Fig. 4. Chromatograms of bacteriohopanepolyols obtained by APCI-HPLC-MS for (a) Soil and (b) Graham's Harbour. Compounds are identified by roman numerals corresponding to structures in the Fig. A1.

2-methylbacteriohopanetetrol (**IIf** and **IIf′**); bacteriohopanepentol (**Ih**); 35-aminobacteriohopane-31,32,33,34-tetrol (**Ii**); two isomers of a cyclitol ether heptaacetate (base peak *m/z* 1002 $[M+H]^+$ with weak *m/z* 330 $[T]^+$ terminal fragments, i.e. neither appears to be a glycoside; **Ij** and **Ij′**); and hopane ribonolactone (**Io′**) (Seemann *et al.*, 1999), plus an apparent 2-methyl homologue of this compound (**IIo**). Novel compound **IIo** has a mass spectrum consistent with an A-ring methylated version of **Io**, with base peak *m/z* 641 $[M+H]^+$ and major fragments at *m/z* 581 $[M+H-CH_3COOH]^+$, and *m/z* 435 and 205 (cleavage through the C-ring; Fig. S1). We presume that it is the 2-methyl (rather than 3-methyl) version, based on the relative retention times of 2-methyl analogues of other BHPs in the sample.

The absolute abundance of BHPs in Bahamas soil reflects the small influence of mineral components and the dominantly organic content of the sample. The three most abundant hopanoids are present at 25–40 μg g^{-1} total sample (Fig. 5B), the highest absolute content of hopanoids per sample mass among any of the four locations. Conversely, the total extractable lipid content of the soil sample was very high, resulting in a low abundance of hopanoids per mg of TLE (< 2 μg mg^{-1} TLE; Fig. 5C). Thus these hopanoid quantifications represent extreme outliers, and this sample is not comparable to the

© 2008 The Authors
Journal compilation © 2008 Society for Applied Microbiology and Blackwell Publishing Ltd,

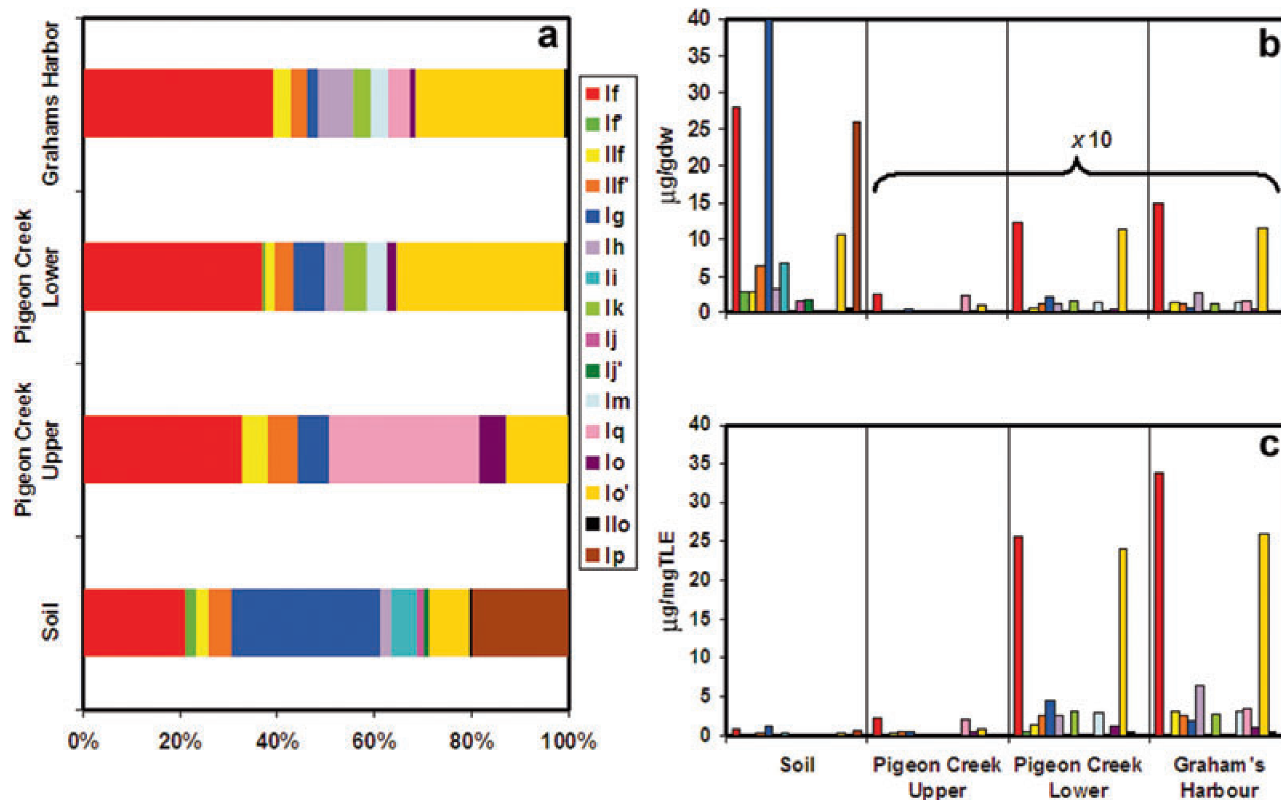

Fig. 5. Abundances of bacteriohopanepolyols (BHPs) shown (a) as the percentage of total BHPs (b) as µg hopanoids per gram dry weight of sample, with values multiplied ×10 for all samples except the soil sample, and (c) as µg of hopanoids per mg of TLE. Compounds are identified by roman numerals corresponding to structures in the Fig. A1.

results obtained for the aquatic samples, all of which were rich in $CaCO_3$ minerals.

There was a smaller diversity of BHPs in the Pigeon Creek Upper (PCU) sample (Fig. 5A). BHT (**If**) was the most abundant component at 33%. Nearly as abundant was 32,35-anhydrobacteriohopanetetrol (**Iq**; 31%). This compound has been proposed to be a reductive degradation product of **Ip**, adenosylhopane (Talbot *et al.*, 2005). Although **Ip** was not detected in this sample, significant proportions of the related hopane ribonolactone (a proposed oxidative degradation product or biosynthetic intermediate; (Seemann *et al.*, 1999) again were detected (two isomers, **Io** and **Io′**). Together with BHT, these compounds account for 83% of the total hopanoids identified in the Pigeon Creek sample. The remaining minor hopanoids in Pigeon Creek were the two isomers of 2-methylbacteriohopanetetrol (**IIf** and **IIf′**) and a small amount of aminotriol (**Ig**).

For BHP analysis of PCU, the amount of TLE injected on-column was comparable to the other samples analysed. Thus, while the limited diversity of BHPs detected in this sample possibly could have been affected by instrumental detection limits, we think this explanation is unlikely. Two complete replicates (extraction, separation and analysis) of the Pigeon Creek sample were performed with similar results. Pigeon Creek Lower (PCL) has the lowest absolute concentration of hopanoids per gram of sample (< 0.3 µg g^{-1}; Fig. 5B, shown ×10 in the figure for visibility), but does not have the lowest concentration of hopanoids relative to TLE (soil is lower; Fig. 5C).

The BHP samples from PCL and GH are virtually identical, suggesting their similar depositional environment is a greater determinant than their geography (Fig. 1). Graham's Harbour is a shallow, open marine bay. The PCL sample was taken at the open mouth of the estuary, where seawater fully intrudes. Both samples are dominated by BHT (**If**) and hopane ribonolactone (**Io′**), with 37% **If** and 35% **Io′** in PCL and 40% **If** and 31% **Io′** in GH. They also contain minor amounts of a second isomer of the lactone (**Io**) and of the putative 2-methyl lactone (**IIo**). Both samples contain 2-methylbacteriohopanetetrol (**IIf**). There are minor amounts of 35-aminotriol (**Ig**), as well as some bacteriohopanepentol (**Ih**), but apparently no 35-aminotetrol (**Ii**). Both samples contain very similar relative amounts of peracetylated BHT-cyclitol ether (**Ik**) and its pentafunctionalized counterpart [**Im**; *m/z* 713 diagnostic fragment from cleavage of the glycosidic moiety (Talbot *et al.*, 2003b)]. Graham's Harbour contains measurable

© 2008 The Authors

32,35-anhydrobacteriohopanetetrol (**Iq**), but PCL does not. Both TLE samples were analysed at approximately the same concentrations and were rich in total BHPs, so these differences (e.g. lack of **Iq**) apparently are real.

Finally, both marine samples show absolute abundances of BHPs up to 2 $\mu g\ g^{-1}$ total sample, which is nearly an order of magnitude more concentrated than in PCU (Fig. 5B). Concentrations of the major BHPs relative to total extractable lipids are similar in these samples, and much richer than soil or PCU, with values approaching 35 $\mu g\ mg^{-1}$ TLE (Fig. 5C). The minor hopanoids also are both more abundant as well as more diverse in PCL and GH relative to PCU.

Free hopanols and diploptene

Free hopenes, hopanols and tetrahymanol also were detected in each sample (Table 2). Biogenic hopenes included diploptene (**Ia**) and 22,29,30-trisnorhop-17(21)-ene (**III**), both of which appeared more abundant in the terrigenously influenced samples. Diploptene, in particular, has been invoked as a tracer of land-derived organic matter (Prahl *et al.*, 1992). Free hopanols included diplopterol (**Ib**), 2-methyldiplopterol (**IIb**) and tetrahymanol (**IV**), all of which nearly co-eluted but could be distinguished by comparison of their mass spectra to literature data and to the spectra and retention times of triterpenoids present in previously characterized bacteria.

All samples also contained the common C_{32} hopanol, bishomohopanol (**Ie**) and its C_{31} analogue, homohopanol (**Id**), which are degradation products of tetrafunctionalized and pentafunctionalized BHPs respectively. Similarly, minor amounts of putative 2-methylbishomohopanol (**IIe**) could be detected in some samples, but the analogous 2-methylhomohopanol (**IId**) was not detected in any sample. This is consistent with the apparent absence of any pentafunctionalized 2-MeBHP compound that could act as a precursor, i.e. we did not detect structures **IIh**, **IIi** or **IIm**. An additional C_{32} hopanol (**Ie′**) was detected in most samples, but its mass spectrum was inconsistent with methylation on the A-ring (no *m/z* 205). Instead, its mass spectrum was similar to bishomohopanol, and it could be the uncommon biogenic form 17α(H), 21β(H)-bishomohopanol (Rosa-Putra *et al.*, 2001). This compound has been reported in *Frankia* spp., a soil bacterium, and here is relatively more abundant in samples proximal to terrigenous sources (Table 2).

This example is one of several trends in relative abundance that were detected among the C_{30} hopanoids and their extended counterparts, the BHPs. At least three classes of compounds decrease in relative abundance from land to ocean. The primary example is adenosylhopane, which could only be detected in the soil sample. In addition, amino-functionalized BHPs also decline from

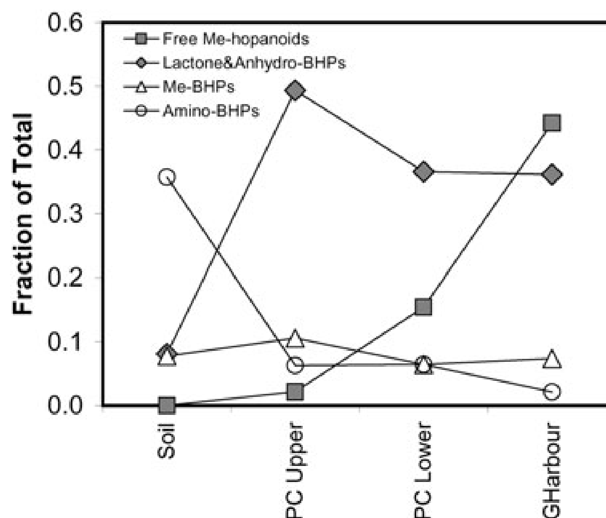

Fig. 6. Land-sea distribution metrics developed for free hopanoids, 2-MeBHPs, amino-BHPs and lactone and anhydro-BHPs. Locations are shown from left to right in order of increasing marine character of the sample site.

the soil, through the estuary, and into GH (Fig. 6), and there is some indication that intact, A-ring methylated BHPs also decrease between the soil and the ocean. In contrast, the lactone and the 32,35-anhydrohopanoid increase as samples become progressively more marine, as does the abundance of 2-methyldiplopterol. The extent to which these trends are biosynthetic or diagenetic is difficult to determine. However, as total intact BHPs are more abundant and total hopanols (plus diploptene) are less abundant as samples become marine in character, this suggests that the increase in 2-methyldiplopterol might indeed be biosynthetic, because it goes against this trend.

Relating phylogeny and physiology to BHP distributions

Presently there is no clear trend that can relate the presence or absence of BHPs in a phylogenetic group to its general physiology. Little is known about the role(s) of BHPs in the cell, although evidence is now emerging that polycylic terpenoids in bacteria may serve a variety of functions, and these functions could vary by taxonomy or environment or physiology. Notably, hopanoids are produced in *Streptomyces coelicolor* during the formation of aerial hyphae and sporulation (Poralla *et al.*, 2000); whether the role is structural, prevents dessication or reduces oxidative stress is unknown. In related cases, the diazotrophic soil bacteria *Frankia* spp. produce heterocyst-like vesicles coated with BHPs, apparently to protect nitrogenase from exposure to oxygen (Berry *et al.*, 1993), and *Bacillus* spp. produce regular polyprenoids known as sporulenes (Kontnik *et al.*, 2008) to protect their

© 2008 The Authors
Journal compilation © 2008 Society for Applied Microbiology and Blackwell Publishing Ltd,

spores against oxidative stress (Bosak *et al*., 2008). Reports that *Zymomonas mobilis* increases its hopanoid content as a function of exposure to ethanol have been questioned, but there is evidence for a temperature association (Bringer *et al*., 1985; Schmidt *et al*., 1986; Hermans *et al*., 1991; Joyeux *et al*., 2004). This lack of understanding about hopanoid function means that linking environmental phylogenetic distributions of *sqhC* genes to physiological roles for BHPs – either specific BHPs or all BHPs in aggregate – remains an enigma.

The potential to use BHPs as bacterial biomarkers for phylogeny has been studied in a diverse range of cultured strains (e.g. Neunlist and Rohmer, 1985a; Neunlist *et al*., 1988; Llopiz *et al*., 1992; Rohmer and Bisseret, 1994; Cvejic *et al*., 2000; Talbot *et al*., 2008). Notably, we did not identify any novel BHPs in high abundance, although this was partly a constraint imposed by the MS^n with Dynamic Exclusion methodology we used. All of the BHPs found in the Bahamas samples are known products of previously characterized species, with the one exception of the A-methylbacteriohopaneribonolactone. Although we show for the first time profiles of BHPs with simultaneous characterization of the diversity of squalene cyclases, given the novelty of most of the environmental *sqhC* sequences, it remains a significant challenge to make the association between *sqhC* sequence → species → BHP profile. One clone from the soil environment groups closely with *Nitrosomonas europaea*, which is known to produce the ribonolactone and adenosyl-BHPs observed in this sample (Seemann *et al*., 1999). Previous work in a methane-rich lake detected sequences closely affiliated with aerobic methanotrophs (Pearson *et al*., 2007). BHP abundances recently were reported from this lake, and aminotriol was detected in the metalimnion and surface sediments (Ertefai *et al*., 2008), consistent with expectations that amino-BHPs would be found (Neunlist and Rohmer, 1985c,b; Cvejic *et al*., 2000; Coolen *et al*., 2008). These are among the few predictions that could be made from the observed *sqhC* and BHP data.

The abundance of the BHP lactone is particularly striking, both in its absolute concentration, which is far in excess of previous reports of this compound in bacterial cultures (Seeman *et al*., 1999), and in its relative proportion in samples of marine character (Figs 5 and 6). To our knowledge, this is the first report of this compound in an environmental sample. The reduced analogue, 32,35-anhydro-BHP, has been reported more often and in somewhat higher abundance (Bednarczyk *et al*., 2005; van Dongen *et al*., 2006). In our samples it is less abundant than the lactone and appears to be replaced by the lactone as samples become increasingly marine (Fig. 5). If the source is biosynthetic rather than diagenetic, among the environmental sequences in Fig. 3, only group UMG appears to be a uniquely marine ecotype. This group is a promising target for study of BHPs that could be used as specific markers for marine production. However, the specific hopanoids made by UMG cannot yet be determined.

Such uncertainty poses a concern for previous correlations between specific BHP structures and their phylogenetic groups of origin. For example, until recently it was thought that *Cyanobacteria* were the most widespread taxonomic group to synthesize BHP methylated at the C-2 position. For that reason we had expected to identify cyanobacterial *sqhC* genes in most of our study sites based on the presence of 2-MeBHPs; the subsequent failure to detect such genes is striking. The Global Ocean Sampling expedition (Venter *et al*., 2004; Rusch *et al*., 2007) also does not contain any cyanobacterial *sqhC* sequences in similar types of samples (Pearson and Rusch, 2008). This raises the possibility that other sources of methylated BHPs may be important in tropical carbonate platforms. The recent discovery of 2-methylhopanoids in *R. palustris* (Rashby *et al*., 2007) and *B. japonicum* (Talbot *et al*., 2007a) reveals potential alternative sources. Alternatively, episodic blooms may be an important factor in the ecology of hopanoid-producing marine *Cyanobacteria*, and the 2-methylhopanoids we observe here in sediments may indeed be cyanobacterial. The water column at the time of sampling could have had low numbers of *Cyanobacteria*. This is feasible given that all of the presently known marine species of hopanoid-producing *Cyanobacteria* are diazotrophic (Talbot *et al*., 2008) and these tend to proliferate in short blooms. Conversely, no marine *Cyanobacteria* have yet been found to produce 2-methyl-BHPs. Although we expected diazotrophs to be prevalent in the summer months during our field expedition (Capone *et al*., 1997; Zehr *et al*., 2001; 2007), bloom events are difficult to capture in water-column sampling. On the other hand, sediments, which integrate extended periods of water-column production and export, provide a valuable record of temporal variability in the proliferation of BHP producers.

What is the potential degree of species-specificity of the BHP compounds observed here? Given that the minimum number of hopanoid-producing 'species' in these samples is 56 (Table 1) and the total number of different compounds observed is 21 (counting hopenes, diplopterol, tetrahymanol and BHPs, but not diagenetic products such as C_{31} and C_{32} hopanols; Table 2), it seems reasonable to ask if any taxonomically unique BHPs exist. Given that cultivated hopanoid producers that have been screened in prior studies generally make between 2 and 10 different BHP structures, the structural diversity observed in Bahamas samples could be accounted for by 1–5 hopanoid producing species per environment. An alternative phrasing of this question is to ask whether the diversity of *sqhC* sequences at a given site predicts the diversity of co-occurring BHP compounds, under the

© 2008 The Authors

assumption that the more *sqhC*-containing organisms there are, the more potential capacity there could be to create diverse BHP head groups. The number of *sqhC* sequences detected in each sample is certainly less than the total *sqhC* diversity. Estimates of rarefaction curves from previous work (Pearson *et al*., 2007), as well as the data here, suggest a minimum of several tens of *sqhC*s in most environments. The data also likely underestimate β and γ-*Proteobacteria*, and only capture a brief temporal snapshot of genetic data (whereas lipids are integrative). These factors would potentially drive estimates of genetic versus lipid diversity even higher.

Based on current data, the soil contains the largest number of unique *sqhC*s (31), and it contains 12 BHPs. The sample from PCU is less diverse for *sqhC* (8) and contains only six BHPs. These two samples alone would suggest (i) that *sqhC* diversity predicts BHP diversity, (ii) that soils are more diverse and produce more hopanoids than aquatic settings, and (iii) that the prospect for making taxonomic connections between *sqhC*s and BHPs would be optimistic. However, the samples from PCL and GH contain 15 and 4 identifiable *sqhC*s, respectively, but both contain 11 BHPs. Given the prospect for offshore transport of organic matter, there could be a disconnect between the presence of *sqhC*-containing cells in a local environment and the accumulation of exogenous BHPs, and this could overwhelm the *in situ* BHP signal in GH. However, evidence against this comes from the steady decrease in amino-BHPs (Fig. 6), and especially from the disappearance of the purely terrigenous adenosyl-BHP, unless diagenesis converts the latter to either BHP ribonolactone or to 32,35-anhydro-BHP. Graham's Harbour also is not proximal to Pigeon Creek (Fig. 1). The reason for the small number of identifiable *sqhC*s in GH remains unknown.

BHPs and environments

The similarity between GH and PCL could suggest that environmental conditions are a significant factor affecting BHP diversity. Determining the BHP distributions in different settings might be more important than characterizing *sqhC*s. It is possible that specific environmental conditions trigger the production of different structures by a given species or an entire community, and that individuals vary their BHPs under different conditions (e.g. Poralla *et al*., 2000; Rashby *et al*., 2007). Presently, there are few BHPs that are diagnostic for specific kinds of environments. Adenosyl-BHP has only been observed in soils and rarely in lacustrine sediments (Talbot and Farrimond, 2007; Cooke *et al*., 2008a,b). Compared with some environments (e.g. Lake Druzby; Talbot and Farrimond, 2007), the diversity of BHPs here is relatively low; and, with the exception of the adenosyl-BHP and traces of pentafunctionalized BHPs, there are no compounds unique to a particular sample. Could this be a function of the relative uniformity of the Bahamian environment (oxic, warm and sunlit)?

Laboratory studies suggest that at least in some cases there are associations between particular BHPs and organisms expressing specific metabolisms. For example, amino-BHPs are prevalent in methanotrophs. Aminotetrol-BHP has only been found in methanotrophs and *Desulfovibrio* spp. (Blumenberg *et al*., 2006) and is most frequently observed in settings where methanogenesis is likely to occur. We observe the highest fraction of amino-BHPs in the soil environment (Fig. 6), and the only occurrence of aminotetrol-BHP, suggesting that methanotrophs are more prevalent in the soil. This is reasonable, given the oxygen and sulfate-rich marine waters of the Bahamas, whereas the organic-rich and sulfate-poor soils would provide an advantage for methanogens. Conversely, the 2-methylhopanoids are thought to be primarily associated with cyanobacterial oxygenic photosynthesis (Summons *et al*., 1999); however, we do not observe a clear trend towards higher abundance of 2-MeBHPs in marine environments.

To interpret these environmental trends in future studies, it will be essential to combine quantitative molecular analyses of *sqhC* genes with continuing BHP measurements, while simultaneously measuring the key environmental parameters (e.g. temperature, salinity, pH, light intensity, redox potential, nutrients, and C_1 and C_2 compounds such as methane and acetate). A major outstanding question is the balance between species diversity and environmental control on the ultimate composition and diversity of BHP assemblages. Simultaneous assessment of multiple variables may result in a predictive capacity for the role of hopanoids and better constraints on paleoenvironmental conditions. This will provide necessary context for studies on BHP function within specific organisms. Finally, complementary analyses of stable isotope ratios and/or other major classes of biomarkers will continue to play an essential role for interpreting the sources of hopanoids in the geologic record (e.g. Jahnke *et al*., 1999; Pancost *et al*., 2000; Thiel *et al*., 2001; Hinrichs *et al*., 2003; Boot *et al*., 2006; Wakeham *et al*., 2007; Coolen *et al*., 2008).

Experimental procedures

Location and sampling

Samples were taken from San Salvador, The Bahamas, in June 2007 (Table 3, Fig. 1). Water samples were collected in HCl-cleaned, fluorinated 20 l carboys and were vacuum filtered onto 153 mm Supor (0.2 μm) filters either for freezing or for immediate lysis. Sediment samples from the same locations were collected using an Ekman grab. Soil was

© 2008 The Authors

Table 3. Bahamas station data.

	Location	Salinity	pH	Temperature (°C)	NH_4^+ (μM)	NO_3^- (μM)	PO_4^{3-} (μM)
Upland Soil	24°2.293′N, 74°27.068′W	n.a.	n.a.	27[a]	n.a.	n.a.	n.a.
Upper Pigeon Creek	24°0.741′N, 74°28.098′W	34.7	8.06	26.8	0.006 ± 0.001	2.3 ± 0.6	0.03 ± 0.05
Lower Pigeon Creek	23°57.834′N, 74°29.219′W	32.1	8.09	25.9	0.003 ± 0.001	2.7 ± 0.1	0.05 ± 0.00
Graham's Harbour	24°07.568′N, 74°28.024′W	35.2	8.14	27.9	–0.002 ± 0.003	1.4 ± 1.1	0.00 ± 0.00

a. Mean monthly temperature, June.

recovered using a sterile, stainless steel scoop. All samples were stored in sterile Whirlpak bags, pre-combusted glass jars or directly in CTAB DNA lysis buffer. Samples in CTAB were stored and shipped at 4°C, while all others were stored and shipped at –70°C. Waters for nutrient analysis were collected into Vacutainers, poisoned with benzalkonium chloride and stored at 4°C in the field. Nutrient concentrations were measured on site within 8 h using the MCM LTER Limnological Methods Manual (http://www.homepage.montana.edu/~lkbonney/). All other samples were transported to the USA for further analysis.

DNA isolation and PCR amplification

Isolation and purification of environmental DNA (eDNA) was performed on approximately 10 g of soil from the upland soil (SA) and 20 l of water from the upstream (PCU) and downstream (PCL) tidal creek, and the shallow open ocean (GH). Details of DNA extraction and PCR amplification, cloning and sequencing appear in Appendix S1. Briefly, DNA was extracted and partial *sqhC* genes were amplified by a two-step nested PCR protocol using conditions and primers previously described (Pearson *et al*., 2007). PCR products were cloned using the Invitrogen TOPO TA Cloning Kit for Sequencing with cloning vector pCR2.1 with TOP10 competent cells, and individual clones were screened for correct inserts using vector primers M13F and M13R. DNA was sequenced at the Dana-Farber/Harvard Cancer Center DNA Resource Core (http://dnaseq.med.harvard.edu/).

Reciprocal BLAST

SHC protein sequences were translated using the ExPASy Translate Tool (http://www.expasy.ch/tools/dna.html). The closest relative to each sequence was then identified among all genomic data in the public domain using the nr/nt nucleotide database and the NCBI (http://www.ncbi.nlm.nih.gov/blast/Blast.cgi) implementation of tBLASTn (Altschul *et al*., 1997). Typical expectation values for 200-mer fragments compared with sequences of known SHCs are $< 10^{-50}$ and typical sequence identities are > 40% (Table S1).

Alignments and tree construction

Complete sequences of known SHCs were obtained from the Integrated Microbial Genomes database of the Joint Genomes Institute (http://img.jgi.doe.gov/cgi-bin/m/main.cgi) and from the National Center for Biotechnology (http://www.ncbi.nlm.nih.gov/). Parameters for tBLASTn were the reference SHC sequence for *Alicyclobacillus acidocaldarius* (GI: 2851526); standard genetic code; expectation value cut-off 10^{-5}. All genomic and environmental partial SHCs were aligned using CLUSTALW through the portal http://clustalw.genome.ad.jp/. Multiple alignment parameters were: gap open penalty, 13.0; gap extension penalty 0.05; BLOSUM weight matrix for proteins; indels treated as single substitutions for use in a masked alignment or as individual substitutions in an unmasked alignment.

Two trees, reflecting the two approaches, subsequently were compared. The alignment spanned residues 309–535 of *A. acidocaldarius*, corresponding to the amplicon region. Included in this alignment were one species of every genus of cultured organisms possessing an SHC-encoding gene, all formerly obtained PCR amplicons (Pearson *et al*., 2007), and all data from this work. Bayesian trees were calculated using Mr Bayes (Huelsenbeck and Ronquist, 2001) using a mixed amino acid substitution model, *lset* equal to gamma and using four chains. Tree iterations were ceased when the posterior probabilities reached < 0.05 at approximately 2.2 million generations.

Lipid extraction and separation

Total lipid extracts (TLEs) were obtained from 16 g of SoilA, 86 g of PCU (dry weights), and ~240 g each wet weight of PCL and GH surface sediments. Sediments, rather than suspended particulate matter from the overlying water column, were used to maximize yields of TLE. Frozen sediment and mat samples were extracted using a modified Bligh and Dyer method. Freeze-dried material was re-hydrated in a monophasic mixture containing 2:1:0.8 $CH_3OH/CH_2Cl_2/H_2O$ with 1% trichloroacetic acid (Bligh and Dyer, 1959; Sturt *et al*., 2004), except for samples GH and PCL, which were extracted prior to drying. All samples were sonicated for 1 h and then shaken at room temperature for 12 h. After centrifuging to remove particulates, the supernatant was separated into aqueous and organic phases using 1:1 CH_2Cl_2/H_2O. We performed an additional separation on the remaining aqueous phase by first blowing it down to near dryness under N_2 in order to remove CH_3OH, then adding 10 ml water and extracting again with 10 ml CH_2Cl_2, in order to minimize loss of poorly extractable hopanoids (Herrmann *et al*., 1996). Total TLEs were passed over Na_2SO_4, taken to dryness under N2 and weighed.

Aliquots for high-performance liquid chromatography (HPLC) were acetylated in 10:10:1 pyridine/acetic anhydride/CH_2Cl_2, heated for 1 h at 70°C, and left at room temperature overnight. Following derivatization, TLEs were brought up in 60:40 methanol/isopropanol and filtered through a 0.2 μm syringe filter. Underivatized aliquots of TLE for gas chromatography-mass spectrometry (GC/MS) were separated by gravity SiO_2-gel (100–200 mesh) columns into

© 2008 The Authors

apolar (10% ethyl acetate in hexane), free alcohol (25% ethyl acetate in hexane) and polar (100% methanol) fractions. All solvents were Burdick and Jackson GC^2 grade.

GC/MS and HPLC-APCI-MS

Details of GC and HPLC separations and the accompanying MS methods appear in Appendix S1. Briefly, volatile lipids were analysed on an Agilent Technologies 6890 gas chromatograph coupled to a 5973 mass selective detector, equipped with a J and W Scientific DB-5MS column. Intact BHPs were analysed by high performance liquid chromatography atmospheric pressure chemical ionization mass spectrometry (HPLC-APCI/MS^n) on a Finnigan LTQ ion trap MS at the Woods Hole Oceanographic Institution. LC-MS conditions were adapted from published methods (Talbot *et al.*, 2007a). For both GC/MS and HPLC-APCI-MS, extracts of pure cultures of *R. palustris* CGA009, *Gemmata obscuriglobus* UQM2246 (A. Pearson, L. Wolchok and R. Summons, unpublished), *Crocosphaera watsonii* WH8501 and *Trichodesmium erythraeum* were run for verifications of specific molecular structures, and whenever possible, mass spectra and retention times also were checked for correspondence previous identifications (Talbot *et al.*, 2001; Talbot *et al.*, 2003a,b; 2005; 2007a,b; 2008). All data were analysed with AMDIS version 2.64 software. Our data are based on the total ion current of the full scan data compared with that of an acetylated glucosyl sitosterol external standard, assuming uniform response factors. Especially in cases of N-containing head groups, the absolute abundances are only semi-quantitative.

Acknowledgements

We thank the Government of The Commonwealth of The Bahamas and The Gerace Research Center for their hospitality, and especially Tom Rothfus for his assistance during our visit. We thank Felisa Wolfe-Simon for instruction in Mr Bayes. Laboratory assistance was provided by Susan Carter. Helen Talbot, Kai-Uwe Hinrichs and an anonymous reviewer provided numerous helpful comments on the manuscript. W.D.L and J.P.S. acknowledge support from NSF graduate research fellowships. J.P.S and R.E.S. were supported by a grant from the American Chemical Society Petroleum Research Fund. R.E.S. also thanks the Hanse-Wissenshaftskolleg and the Alexander von Humboldt Foundation for support. This work was funded by NSF EAR-0641899, the Merck Foundation and The David and Lucille Packard Foundation (to A.P.). We acknowledge M. Soule, E. Kujawinski and the Funding Sources of the WHOI FT-MS Users' Facility (NSF OCE-0619608 and the Gordon and Betty Moore Foundation).

References

Adams, R.W. (1980) General guide to the geological features of San Salvador. In *Field Guide to the Geology of San Salvador*. Gerace, D.T. (ed.). San Salvador, Bahamas: Bahamian Field Station, pp. 1–66.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J.H., Zhang, Z., Miller, W., and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25:** 3389–3402.

Bednarczyk, A., Hernandez, T.C., Schaeffer, P., Adam, P., Talbot, H.M., Farrimond, P., *et al.* (2005) 32,35-Anhydrobacteriohopanetetrol: an unusual bacteriohopanepolyol widespread in recent and past environments. *Org Geochem* **36:** 673–677.

Berry, A.M., Harriott, O.T., Moreau, R.A., Osman, S.F., Benson, D.R., and Jones, A.D. (1993) Hopanoid lipids compose the *Frankia* vesicle envelope, presumptive barrier of oxygen diffusion to nitrogenase. *Proc Natl Acad Sci USA* **90:** 6091–6094.

Bligh, E.G., and Dyer, W.J. (1959) A rapid method of total lipid extraction and purification. *Can J Biochem Physiol* **37:** 911–917.

Blumenberg, M., Kruger, M., Nauhaus, K., Talbot, H.M., Oppermann, B.I., Seifert, R., *et al.* (2006) Biosynthesis of hopanoids by sulfate-reducing bacteria (genus *Desulfovibrio*). *Environ Microbiol* **8:** 1220–1227.

Boot, C.S., Ettwein, V.J., Maslin, M.A., Weyhenmeyer, C.E., and Pancost, R.D. (2006) A 35,000 year record of terrigenous and marine lipids in Amazon Fan sediments. *Org Geochem* **37:** 208–219.

Bosak, T., Losick, R.M., and Pearson, A. (2008) A polycyclic terpenoid that alleviates oxidative stress. *Proc Natl Acad Sci USA* **105:** 6725–6729.

Bringer, S., Hartner, T., Poralla, K., and Sahm, H. (1985) Influence of ethanol on the hopanoid content and the fatty-acid pattern in batch and continuous cultures of *Zymomonas mobilis*. *Arch Microbiology* **140:** 312–316.

Brocks, J.J., Logan, G.A., Buick, R., and Summons, R.E. (1999) Archean molecular fossils and the early rise of eukaryotes. *Science* **285:** 1033–1036.

Capone, D.G., Zehr, J.P., Paerl, H.W., Bergman, B., and Carpenter, E.J. (1997) *Trichodesmium*, a globally significant marine cyanobacterium. *Science* **276:** 1221–1229.

Cooke, M.P., Talbot, H.M., and Wagner, T. (2008a) Tracking soil organic carbon transport to continental margin sediments using soil-specific hopanoid biomarkers: a case study from the Congo fan (ODP site 1075). *Org Geochem* **39:** 956–971.

Cooke, M.P., Talbot, H.M., and Farrimond, P. (2008b) Bacterial populations recorded in bacteriohopanepolyol distributions in soils from Northern England. *Org Geochem* **39:** 1347–1358.

Coolen, M.J.L., Talbot, H.M., Abbas, B.A., Ward, C., Schouten, Volkman, J.K., and Sinninghe Damsté, J.S. (2008) Sources for sedimentary bacteriohopanepolyols as revealed by 16S rDNA stratigraphy. *Environ Microbiol* **10:** 1783–1803.

Cvejic, J.H., Bodrossy, L., Kovacs, K.L., and Rohmer, M. (2000) Bacterial triterpenoids of the hopane series from the methanotrophic bacteria *Methylocaldum* spp. phylogenetic implications and first evidence for an unsaturated aminobacteriopanepolyol. *FEMS Microbiology Lett* **182:** 361–365.

van Dongen, B.E., Talbot, H.M., Schouten, S., Pearson, P.N., and Pancost, R.D. (2006) Well preserved *Palaeogene* and *Cretaceous* biomarkers from the Kilwa area, Tanzania. *Org Geochem* **37:** 539–557.

© 2008 The Authors

Dutkiewicz, A., Volk, H., George, S.C., Ridley, J., and Buick, R. (2006) Biomarkers from Huronian oil-bearing fluid inclusions: An uncontaminated record of life before the Great Oxidation Event. *Geology* **34:** 437–440.

Eigenbrode, J., Summons, R.E., and Freeman, K.H. (2008) Methylhopanes in Archean sediments. *Earth Planetary Sci Lett* **273:** 323–331.

Ertefai, TF., Fisher MC., Fredricks, HF., Lipp, JS., Birgel D., Udert KM., *et al.* (2008) Vertical distribution of microbial lipids and functional genes in chemically distinct layers of a highly polluted meromictic lake. *Org Geochem* **39:** 1572–1588.

Feil, C., Sussmuth, R., Jung, G., and Poralla, K. (1996) Site-directed mutagenesis of putative active-site residues in squalene-hopene cyclase. *Eur J Biochem* **242:** 51–55.

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

Frickey, T., and Kannenberg, E. (2008) Phylogenetic analysis of the triterpene cyclase protein family in prokaryotes and eukaryotes suggests bidirectional lateral gene transfer. *Environ Microbiol* (in press).

Fuhrman, J.A., McCallum, K., and Davis, A.A. (1993) Phylogenetic diversity of subsurface marine microbial communities from the Atlantic and Pacific Oceans. *Appl Environ Microbiol* **59:** 1294–1302.

Hartner, T., Straub, K.L., and Kannenberg, E. (2005) Occurrence of hopanoid lipids in anaerobic Geobacter species. *FEMS Microbiol Lett* **243:** 59–64.

Hermans, M.A.F., Neuss, B., and Sahm, H. (1991) Content and composition of hopanoids in *Zymomonas mobilis* under various growth conditions. *J Bacteriol* **173:** 5592–5595.

Herrmann, D., Bisseret, P., Connan, J., and Rohmer, M. (1996) A non-extractable triterpenoid of the hopane series in *Acetobacter xylinum*. *FEMS Microbiol Lett* **135:** 323–326.

Hinrichs, K.U., Hmelo, L.R., and Sylva, S.P. (2003) Molecular fossil record of elevated methane levels in late pleistocene coastal waters. *Science* **299:** 1214–1217.

Hoshino, T., and Sato, T. (2002) Squalene-hopene cyclase: catalytic mechanism and substrate recognition. *Chem Commun* **4:** 291–301.

Huelsenbeck, J.P., and Ronquist, F. (2001) MRBAYES: bayesian inference of phylogenetic trees. *Bioinformatics* **17:** 754–755.

Hugenholtz, P., and Pace, N.R. (1996) Identifying microbial diversity in the natural environment: a molecular phylogenetic approach. *Trends Biotechnol* **14:** 190–197.

Jahnke, L.L., Summons, R.E., Hope, J.M., and des Marais, D.J. (1999) Carbon isotopic fractionation in lipids from methanotrophic bacteria II: the effects of physiology and environmental parameters on the biosynthesis and isotopic signatures of biomarkers. *Geochim Cosmochim Acta* **63:** 79–93.

Joyeux, C., Fouchard, S., Llopiz, P., and Neunlist, S. (2004) Influence of the temperature and the growth phase on the hopanoids and fatty acids content of *Frateuria aurantia* (DSMZ 6220). *FEMS Microbiol Ecol* **47:** 371–379.

Knoll, A.H., Summons, R.E., Waldbauer, J.R., and Zumberge, J. (2007) The geological succession of primary producers in the oceans. In Falkowski, P., and Knoll, A.H. (eds). *The Evolution of Primary Producers in the Sea*. Boston, MA, USA: Academic Press, pp. 133–163.

Kontnik, R., Bosak, T., Butcher, R.A., Brocks, J.J., Losick, R.M., Clardy, J., and Pearson, A. (2008) Sporulenes, heptaprenyl metabolites from *Bacillus subtilis* spores. *Org Lett* **10:** 3551–3554.

Llopiz, P., Neunlist, S., and Rohmer, M. (1992) Prokaryotic triterpenoids – *o*-α-D-glucuronopyranosyl bacteriohopanetetrol, a novel hopanoid from the bacterium *Rhodospirillum rubrum*. *Biochem J* **287:** 159–161.

Neunlist, S., Bisseret, P., and Rohmer, M. (1988) The hopanoids of the purple non-sulfur bacteria *Rhodopseudomonas palustris* and *Rhodopseudomonas acidophila* and the absolute configuration of bacteriohopanetetrol. *European J Biochem* **171:** 245–252.

Neunlist, S., and Rohmer, M. (1985a) A novel hopanoid, 30-(5′-adenosyl) hopane, from the purple non-sulfur bacterium *Rhodopseudomonas acidophila*, with possible DNA interactions. *Biochem J* **228:** 769–771.

Neunlist, S., and Rohmer, M. (1985b) The hopanoids of *Methylosinus trichosporium* – aminobacteriohopanetriol and aminobacteriohopanetetrol. *J Gen Microbiol* **131:** 1363–1367.

Neunlist, S., and Rohmer, M. (1985c) Novel hopanoids from the methylotrophic bacteria *Methylococcus capsulatus* and *Methylomonas methanica* – (22S)-35-aminobacteriohopane-30,31,32,33,34-pentol and (22S)-35-amino-3-β-methylbacteriohopane-30,31,32,33,34-pentol. *Biochem J* **231:** 635–639.

Ourisson, G., and Albrecht, P. (1992) Hopanoids. 1. Geohopanoids – the most abundant natural products on Earth. *Accounts Chem Res* **25:** 398–402.

Ourisson, G., Rohmer, M., and Poralla, K. (1987) Prokaryotic hopanoids and other polyterpenoid sterol surrogates. *Annu Rev Microbiol* **41:** 301–333.

Pancost, R.D., Damste, J.S.S., de Lint, S., van der Maarel, M., Gottschal, J.C., and Medinaut Shipboard Sci, P. (2000) Biomarker evidence for widespread anaerobic methane oxidation in Mediterranean sediments by a consortium of methanogenic archaea and bacteria. *Appl Environ Microbiol* **66:** 1126–1132.

Pearson, A., and Rusch, D.B. (2008) Distribution of microbial terpenoid lipid cyclases in the global ocean metagenome. *ISME J* doi:10.1038/ismej.2008.116.

Pearson, A., Page, S.R.F., Jorgenson, T.L., Fischer, W.W., and Higgins, M.B. (2007) Novel hopanoid cyclases from the environment. *Environ Microbiol* **9:** 2175–2188.

Peiseler, B., and Rohmer, M. (1992) Prokaryotic triterpenoids of the hopane series – bacteriohopanetetrols of new side-chain configuration from *Acetobacter* species. *J Chem Res* **9:** 298–299.

Perzl, M., Reipen, I.G., Schmitz, S., Poralla, K., Sahm, H., Sprenger, G.A., and Kannenberg, E.L. (1998) Cloning of conserved genes from *Zymomonas mobilis* and *Bradyrhizobium japonicum* that function in the biosynthesis of hopanoid lipids. *Biochim Biophys Acta Lipids Lipid Metab* **1393:** 108–118.

Poralla, K., Muth, G., and Hartner, T. (2000) Hopanoids are formed during transition from substrate to aerial hyphae in *Streptomyces coelicolor* A3(2). *FEMS Microbiol Lett* **189:** 93–95.

© 2008 The Authors

Prahl, F.G., Hayes, J.M., and Xie, T.M. (1992) Diploptene – an indicator of terrigenous organic carbon in Washington coastal sediments. *Limnol Oceanogr* **37:** 1290–1300.

Rappe, M.S., and Giovannoni, S.J. (2003) The uncultured microbial majority. *Annu Rev Microbiol* **57:** 369–394.

Rashby, S.E., Sessions, A.L., Summons, R.E., and Newman, D.K. (2007) Biosynthesis of 2-methylbacteriohopanepolyols by an anoxygenic phototroph. *Proc Natl Acad Sci USA* **104:** 15099–15104.

Rohmer, M., and Bisseret, P. (1994) Hopanoid and other polyterpenoid biosynthesis in Eubacteria – phylogenetic significance. In *Isopentenoids and Other Natural Products*. Nes, W.D. (ed.). Washington DC, USA: American Chemical Society, pp. 31–43.

Rohmer, M., Bouvier, P., and Ourisson, G. (1979) Molecular evolution of biomembranes – structural equivalents and phylogenetic precursors of sterols. *Proc Natl Acad Sci USA* **76:** 847–851.

Rohmer, M., Bouviernave, P., and Ourisson, G. (1984) Distribution of hopanoid triterpenes in prokaryotes. *J Gen Microbiol* **130:** 1137–1150.

Rosa-Putra, S., Nalin, R., Domenach, A.M., and Rohmer, M. (2001) Novel hopanoids from *Frankia* spp. and related soil bacteria – squalene cyclization and significance of geological biomarkers revisited. *European J Biochem* **268:** 4300–4306.

Rusch, D.B., Halpern, A.L., Sutton, G., Heidelberg, K.B., Williamson, S., Yooseph, S., *et al.* (2007) The Sorcerer II Global Ocean Sampling expedition: Northwest Atlantic through Eastern Tropical Pacific. *Plos Biology* **5:** 398–431.

Schmidt, A., Bringermeyer, S., Poralla, K., and Sahm, H. (1986) Effect of alcohols and temperature on the hopanoid content of *Zymomonas mobilis*. *Appl Microbiology Biotechnology* **25:** 32–36.

Seemann, M., Bisseret, P., Tritz, J.P., Hooper, A.B., and Rohmer, M. (1999) Novel bacterial triterpenoids of the hopane series from *Nitrosomonas europaea* and their significance for the formation of the C-35 bacteriohopane skeleton. *Tetrahedron Lett* **40:** 1681–1684.

Simonin, P., Jurgens, U.J., and Rohmer, M. (1996) Bacterial triterpenoids of the hopane series from the Prochlorophyte *Prochlorothrix hollandica* and their intracellular localization. *European J Biochem* **241:** 865–871.

Sturt, H.F., Summons, R.E., Smith, K., Elvert, M., and Hinrichs, K.U. (2004) Intact polar membrane lipids in prokaryotes and sediments deciphered by high-performance liquid chromatography/electrospray ionization multistage mass spectrometry – new biomarkers for biogeochemistry and microbial ecology. *Rapid Commun Mass Spectrom* **18:** 617–628.

Sugden, M.A., Talbot, H.M., and Farrimond, P. (2005) Flash pyrolysis – a rapid method for screening bacterial species for the presence of bacteriohopanepolyols. *Org Geochem* **36:** 975–979.

Summons, R.E., Bradley, A.S., Jahnke, L.L., and Waldbauer, J.R. (2006) Steroids, triterpenoids and molecular oxygen. *Philosoph Transactions Royal Soc B-Biolog Sci* **361:** 951–968.

Summons, R.E., Jahnke, L.L., Hope, J.M., and Logan, G.A. (1999) 2-Methylhopanoids as biomarkers for cyanobacterial oxygenic photosynthesis. *Nature* **400:** 554–557.

Talbot, H.E., and Farrimond, P. (2007) Bacterial populations recorded in diverse sedimentary biohopanoid distributions. *Org Geochem* **38:** 1212–1225.

Talbot, H.M., Watson, D.F., Murrell, J.C., Carter, J.F., and Farrimond, P. (2001) Analysis of intact bacteriohopanepolyols from methanotrophic bacteria by reversed-phase high-performance liquid chromatography-atmospheric pressure chemical ionisation mass spectrometry. *J Chromatogr A* **921:** 175–185.

Talbot, H.M., Squier, A.H., Keely, B.J., and Farrimond, P. (2003a) Atmospheric pressure chemical ionisation reversed-phase liquid chromatography/ion trap mass spectrometry of intact bacteriohopanepolyols. *Rapid Commun Mass Spectrom* **17:** 728–737.

Talbot, H.M., Summons, R., Jahnke, L., and Farrimond, P. (2003b) Characteristic fragmentation of bacteriohopanepolyols during atmospheric pressure chemical ionisation liquid chromatography/ion trap mass spectrometry. *Rapid Commun Mass Spectrom* **17:** 2788–2796.

Talbot, H.M., Farrimond, P., Schaeffer, P., and Pancost, R.D. (2005) Bacteriohopanepolyols in hydrothermal vent biogenic silicates. *Org Geochem* **36:** 663–672.

Talbot, H.M., Rohmer, M., and Farrimond, P. (2007a) Rapid structural elucidation of composite bacterial hopanoids by atmospheric pressure chemical ionisation liquid chromatography/ion trap mass spectrometry. *Rapid Commun Mass Spectrom* **21:** 880–892.

Talbot, H.M., Rohmer, M., and Farrimond, P. (2007b) Structural characterisation of unsaturated bacterial hopanoids by atmospheric pressure chemical ionisation liquid chromatography/ion trap mass spectrometry. *Rapid Commun Mass Spectrom* **21:** 1613–1622.

Talbot, H.M., Summons, R.E., Jahnke, L.L., Cockell, C.S., Rohmer, M., and Farrimond, P. (2008) Cyanobacterial bacteriohopanepolyol signatures from cultures and natural environmental settings. *Org Geochem* **39:** 232–263.

Thiel, V., Peckmann, J., Richnow, H.H., Luth, U., Reitner, J., and Michaelis, W. (2001) Molecular signals for anaerobic methane oxidation in Black Sea seep carbonates and a microbial mat. *Mar Chem* **73:** 97–112.

Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A., *et al.* (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304:** 66–74.

Wakeham, S.G., Amann, R., Freeman, K.H., Hopmans, E.C., Jorgensen, B.B., Putnam, I.F., *et al.* (2007) Microbial ecology of the stratified water column of the Black Sea as revealed by a comprehensive biomarker study. *Org Geochem* **38:** 2070–2097.

Wendt, K.U., Feil, C., Lenhart, A., Poralla, K., and Schulz, G.E. (1997) Crystallization and preliminary x-ray crystallographic analysis of squalene-hopene cyclase from *Alicyclobacillus acidocaldarius*. *Protein Sci* **6:** 722–724.

Wendt, K.U., Schulz, G.E., Corey, E.J., and Liu, D.R. (2000) Enzyme mechanisms for polycyclic triterpene formation. *Angew Chem Int Ed* **39:** 2812–+.

Woodward, R.B., and Bloch, K. (1953) The cyclization of squalene in cholesterol synthesis. *J Am Chem Soc* **75:** 2023–2024.

Zehr, J.P., Waterbury, J.B., Turner, P.J., Montoya, J.P., Omoregie, E., Steward, G.F., *et al.* (2001) Unicellular

© 2008 The Authors
Journal compilation © 2008 Society for Applied Microbiology and Blackwell Publishing Ltd,

cyanobacteria fix N-2 in the subtropical North Pacific Ocean. *Nature* **412:** 635–638.

Zehr, J.P., Bench, S.R., Mondragon, E.A., McCarren, J., and DeLong, E.F. (2007) Low genomic diversity in tropical oceanic N-2-fixing cyanobacteria. *Proc Natl Acad Sci USA* **104:** 17807–17812.

Zhao, N., Berova, N., Nakanishi, K., Rohmer, M., Mougenot, P., and Jurgens, U.J. (1996) Structures of two bacteriohopanoids with acyclic pentol side-chains from the cyanobacterium *Nostoc* PCC 6720. *Tetrahedron* **52:** 2777–2788.

Appendix

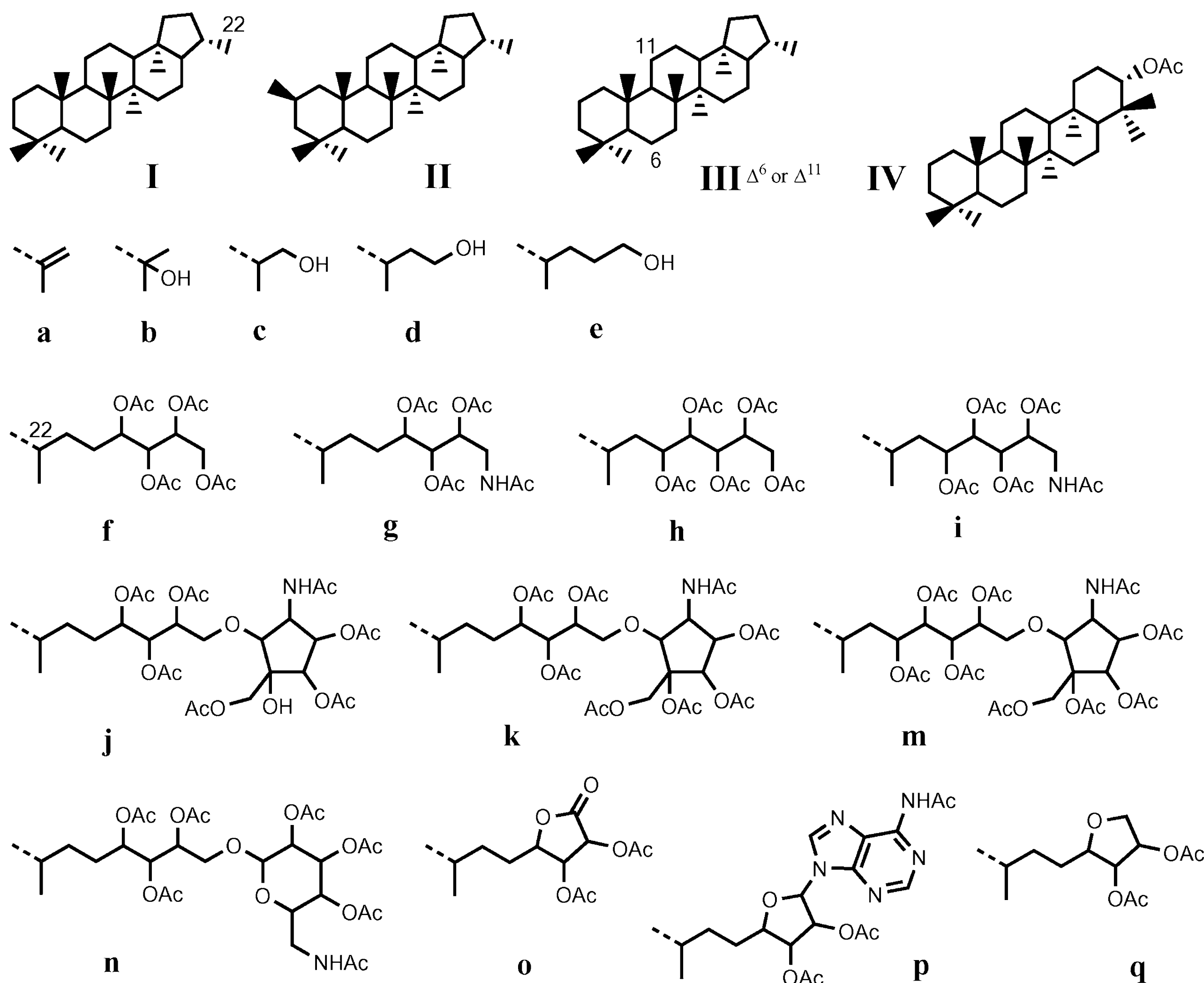

Fig. A1. Structures of hopanoids discussed in the text and shown in Figs 4 and 5 and Table 2. Analysis by LC-MS does not allow precise stereochemical assignment and therefore we assume all compounds identified in this study have the same configuration as previously characterized structures. The occurrence of additional/alternative isomers cannot be excluded.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Fig. S1. Mass spectrum of A-ring-methylbacteriohopaneribonolactone *m*/*z* 641 [M+H]+.

Table S1. Sequence data for Bahamas SHCs.

Table S2. Table of fragment ions for A-ring-methybacterohopaneribonolactone.

Appendix S1. Detailed methods.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

© 2008 The Authors