environmental microbiology

Environmental Microbiology (2013) **15**(5), 1464–1475 doi:10.1111/j.1462-2920.2012.02809.x

Diversity of cyanobacterial biomarker genes from the stromatolites of Shark Bay, Western Australia

Tamsyn J. Garby,[1] Malcolm R. Walter,[1] Anthony W.D. Larkum[2] and Brett A. Neilan[1]*

[1]*Australian Centre for Astrobiology and School of Biotechnology and Biomolecular Sciences, The University of New South Wales, Sydney, NSW 2052, Australia.*

[2]*Plant Functional Biology and Climate Change Cluster, University of Technology Sydney, NSW 2007, Australia.*

Summary

Families of closely related chemical compounds, which are relatively resistant to degradation, are often used as biomarkers to help trace the evolutionary history of early groups of organisms and the environments in which they lived. Biomarkers derived from hopanoid variations are particularly useful in determining bacterial community compositions. 2-Methylhopananoids have been thought to be diagnostic for cyanobacteria, and 2-methylhopanes in the geological record are taken as evidence for the presence of cyanobacteria-containing communities at the time of sediment deposition. Recently, however, doubt has been cast on the validity of 2-methylhopanes as cyanobacterial biomarkers, since non-cyanobacterial species have been shown to produce significant amounts of 2-methylhopanoids. This study examines the diversity of *hpnP*, the hopanoid biosynthesis gene coding for the enzyme that methylates hopanoids at the C2 position. Genomic DNA isolated from stromatolite-associated pustular and smooth microbial mat samples from Shark Bay, Western Australia, was analysed for bacterial diversity, and used to construct an *hpnP* clone library. A total of 117 partial *hpnP* clones were sequenced, representing 12 operational taxonomic units (OTUs). Phylogenetic analysis showed that 11 of these OTUs, representing 115 sequences, cluster within the cyanobacterial clade. We conclude that the dominant types of microorganisms with the detected capability of producing 2-methylhopanoids within pustular and smooth microbial mats in Shark Bay are cyanobacteria.

Received 26 January, 2012; revised 21 May, 2012; accepted 22 May, 2012. *For correspondence. E-mail b.neilan@unsw.edu.au; Tel. (+61) 2 93853235; Fax (+61) 2 93851483.

Introduction

The evolutionary history of early life on Earth is difficult to establish due both to the scarcity of fossilized remains from microbial organisms and the non-specific morphologies of most of those that are found. Sedimentary structures known as stromatolites offer some additional information (Flannery and Walter, 2011) but this, too, is limited. As an alternative, geobiologists use biomarkers to help trace the history of different taxonomic groups and the environments in which they lived. Biomarkers are the remnants of molecules with a backbone structure that is highly resistant to degradation, and can persist in the rock record for billions of years. Ideally, biomarkers can be used to infer the taxonomic composition of the community that was present at the time of sediment deposition, by linking specific compounds to the taxonomic groups that are known to produce them. However, as it is difficult to gain a complete picture of the biosynthetic pathways of organisms that existed billions of years ago, we must base our interpretations on what is known of the biosynthesis of similar compounds in modern day organisms. Unfortunately, the current understanding of all biosynthetic pathways in modern day organisms is also far from complete, which has led to difficulties in interpreting the biomarker record.

Hopanoids are proving a particularly useful group of lipid biomarkers for tracing the evolutionary history of bacteria (Ourisson and Albrecht, 1992; Ourisson and Rohmer, 1992). Hopanes, the degraded products of hopanoids, are probably the most abundant polycyclic hydrocarbons found in crude oil of all ages and have a chemical backbone that is highly resistant to degradation and is readily incorporated into sediments (Summons *et al.*, 1999). The distribution of hopanoids throughout bacteria and the high number of variations that can be produced by different groups means that bacteriohopanepolyols have great potential as molecular markers for bacterial communities, and may be used to characterize both extinct and extant microbial populations. Despite their abundance, very little is known about the current distribution or biosynthesis of hopanoids. Their presence in bacterial cells was not even examined until after it was discovered that bacteriohopanepolyols are found extensively in sediments (Ourisson and Albrecht, 1992) and their biosynthesis is only just beginning to be elucidated (Welander *et al.*, 2012).

The role of hopanoids in bacterial physiology is still uncertain; however, there is a growing body of evidence suggesting that hopanoids play a role in membrane integrity, similar to that of sterols in eukaryotes (Ourisson *et al*., 1987). It is likely that hopanoids have a condensing effect on phospholipid bilayers, increasing structural rigidity and reducing membrane permeability (Kannenberg and Poralla, 1999). This role is supported by evidence showing the localization of hopanoids to membranes in a variety of bacteria (Jurgens *et al*., 1992; Simonin *et al*., 1996; Doughty *et al*., 2009; 2011). The presence and abundance of hopanoids is also likely to be affected by environmental conditions. It is possible that hopanoids can help protect cells from stress due to temperature (Poralla *et al*., 1984; Schmidt *et al*., 1986; Joyeux *et al*., 2004), pH (Poralla *et al*., 1984; Welander *et al*., 2009) or desiccation (Poralla *et al*., 2000). They have also been shown to be involved in the creation of an oxygen-excluding membrane surrounding nitrogen-fixing cells (Berry *et al*., 1993) and in protection against a reactive oxygen species (Bosak *et al*., 2008). Continued work with regards to the function of hopanoids in bacteria will allow us to better interpret the presence of hopanes in the geologic record in terms of the physiological processes and environmental conditions they might represent.

The 2-methylhopanoids in particular have attracted much attention for the potential usefulness of their degraded products as molecular markers. These hopanoids were thought to be produced mainly by cyanobacteria, and so the presence of 2-methylhopanes in the geological record has been used as evidence for the presence of cyanobacteria-containing communities at the time of sediment deposition. This is significant, as the emergence of oxygenic photosynthesis, which radically altered the composition of Earth's atmosphere, is thought to have occurred within cyanobacteria or their ancestors. 2-Methylhopanes are found extensively in both ancient and modern sediments. Summons and colleagues (1999) showed that 2-methylhopanoids are produced abundantly by cyanobacteria in culture and cyanobacteria-dominated microbial mats. Consequently, the presence of 2-methylhopanes in ancient sediments has been used to infer the presence of cyanobacteria, and hence oxygenic photosynthesis, as far back as 2.7 billion years ago, well before the Great Oxidation Event (Brocks *et al*., 1999; 2003a,b; Summons *et al*., 1999; Eigenbrode *et al*., 2008; Waldbauer *et al*., 2009). This interpretation is consistent with (scant) evidence from microfossils (Schopf and Walter, 1983), and from stromatolites (Flannery and Walter, 2011).

Recently, however, questions have been raised as to the validity of using 2-methylhopanes as cyanobacterial biomarkers. 2-Methylhopanoids have been shown to be produced by non-cyanobacterial species, such as *Methylobacterium* (Bisseret *et al*., 1985; Renoux and Rohmer, 1985; Talbot *et al*., 2007), *Bradyrhizobium* (Bravo *et al*., 2001), and *Beijerinckia* (Sahm *et al*., 1993), yet these instances were considered rare and the production of 2-methylhopanoids to be far more abundant in cyanobacteria. However, Rashby and colleagues (2007) found that an anoxygenic phototroph, *Rhodopseudomonas palustris* TIE-1, produces significant amounts of 2-methylhopanoids under anaerobic conditions, challenging the link between 2-methylhopanes, cyanobacteria and oxygenic photosynthesis. It also appears that the ability to synthesize hopanoids is far from ubiquitous among cyanobacteria. Environmental surveys of squalene-hopane cyclase (SHC), a gene that is required for the biosynthesis of all hopanoids, by Pearson and colleagues (2007; 2009) found no cyanobacterial sequences, while an investigation of SHC sequences from the Global Ocean Sampling data found only a single cyanobacterial SHC candidate from a hypersaline lagoon (Pearson and Rusch, 2009).

Thorough determination of the current taxonomic distribution, biosynthesis, function and potential origins of bacterial hopanoids is necessary in order to reliably assign hopane biomarkers from ancient sediments to taxonomic groups. Despite the abundance of hopanoids in sediments of all ages, little progress has been made in this area. This is partly because in the past geologists were mainly concerned with these molecules, while microbiologists took little interest until recently. Welander and colleagues (2010) identified, in the anoxic phototroph *Rhodopseudomonas palustris* TIE-1, a radical SAM methylase gene, termed *hpnP*, responsible for methylation at the C2 position in hopanoids. The identification of this gene provided a means to investigate the distribution and diversity of 2-methylhopanoid producing bacteria in modern environments. A search of public databases revealed only 30 organisms containing homologues for this gene. Ten of these 30 species have previously tested positive for 2-methylhopanoids (Vilcheze *et al*., 1994; Bravo *et al*., 2001; Rashby *et al*., 2007; Talbot *et al*., 2008). The majority of *hpnP* homologues are from members of the alphaproteobacteria, with one putative acidobacterial sequence, and only six identified from cyanobacterial genomes. This limited subset does not provide adequate information regarding the distribution of *hpnP* in bacteria, and it is not known how widespread the capability to produce 2-methylhopanoids is, or from which group this trait originated.

Abundant 2-methylhopanes have been found in ancient sediments which are thought to have originated in shallow-water, likely supporting abundant bacterial growth in the form of stromatolites. One of the best modern analogues for such ancient environments is in Shark Bay, Western Australia. Smooth and pustular

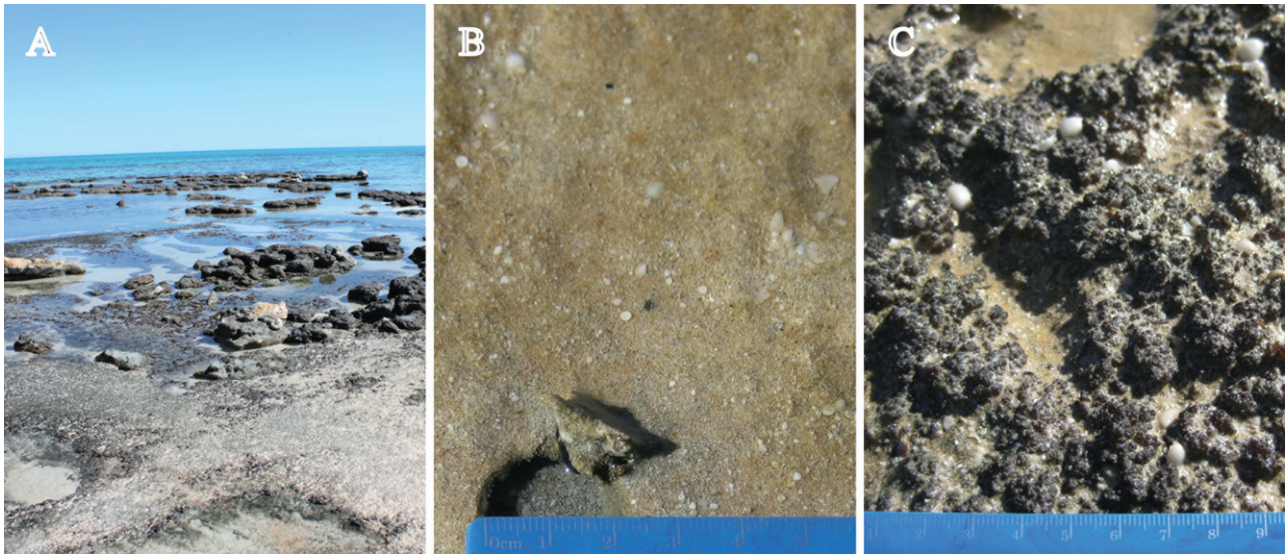

Fig. 1. Photographs of the field site at Carbla Point, Hamelin Pool, Shark Bay.
A. Low tide in the intertidal region at Carbla Point, showing exposed microbial mats.
B. Smooth mat showing area from which a sample core was taken, metric ruler for scale.
C. Pustular mat, metric ruler for scale.

microbial mats as well as columnar stromatolites from Hamelin Pool, Shark Bay support abundant bacterial growth, and have also previously been shown to produce 2-methylhopanoids (Allen *et al*., 2009; 2010). The combination makes this a particularly relevant environment in which to investigate the extant distribution of 2-methylhopanoid producing microorganisms.

Results and discussion

16S rRNA gene composition of microbial mat samples

To gain an approximation of the bacterial community composition of pustular and smooth mats from Hamelin Pool, Shark Bay (Fig. 1), 16S ribosomal RNA genes were sequenced from each sample. Overall, 5706 rRNA sequences from the smooth mat and 8349 sequences from the pustular mat samples were aligned and classified to the level of phylum. Rarefaction curves were generated for each of these two sample data sets to determine the theoretical coverage of 16S rRNA sequences. These indicated that sampling coverage was incomplete for both samples, as neither had reached saturation when operational taxonomic units (OTUs) were grouped by 96% nucleotide sequence identity (data not shown). The bacterial diversity observed within these two mat types is higher than that previously reported, most likely due in part to the better sampling coverage obtained via the high-throughput, tagged pyrosequencing method employed (Sogin *et al*., 2006). Previous work by Allen and colleagues (2009) sampled bacterial diversity of pustular and smooth mats through culturing and 16S rRNA gene clone libraries. The Shannon diversity indices, calculated with OTUs grouped at 96% sequence identity, for the pustular and smooth mats sampled in this study were 5.93 and 5.44 respectively, compared with previously calculated indices at the same level of 4.66 and 4.62 respectively (Allen *et al*., 2009). The composition of phyla observed was very similar between this study and previous work (Allen *et al*., 2009), even though the relative abundance of sequences within each phylum differed, probably due in part to the much larger sample size able to be obtained in the current study. In both samples, members of the alphaproteobacteria formed the largest group: 49% of total bacterial rRNA sequences from the pustular mat and 28% from the smooth mat sample (Fig. 2). In the pustular mat, cyanobacterial sequences made up only 7% of the total bacterial sequences, while in the smooth mat sample they accounted for 21% of the total bacterial sequences (Fig. 2). Alphaproteobacteria were much less abundant in the pustular mat analysed by (Allen *et al*., 2009), only 16% of total sequences. In the smooth mat, cyanobacteria and spirochaetes were found to be much more abundant in the present study; 21% and 10% of total sequences respectively, instead of 6% and < 1% reported by (Allen *et al*., 2009). On the other hand, planctomycetes previously accounted for 14% of sequences isolated from the smooth mat, whereas in the present study it accounted for < 1% of taxa.

The cyanobacterial sequences of each mat sample were further classified to the genus level and analysed in terms of abundance. In the pustular mat sample, *Euhalothece* sequences formed the most abundant group,

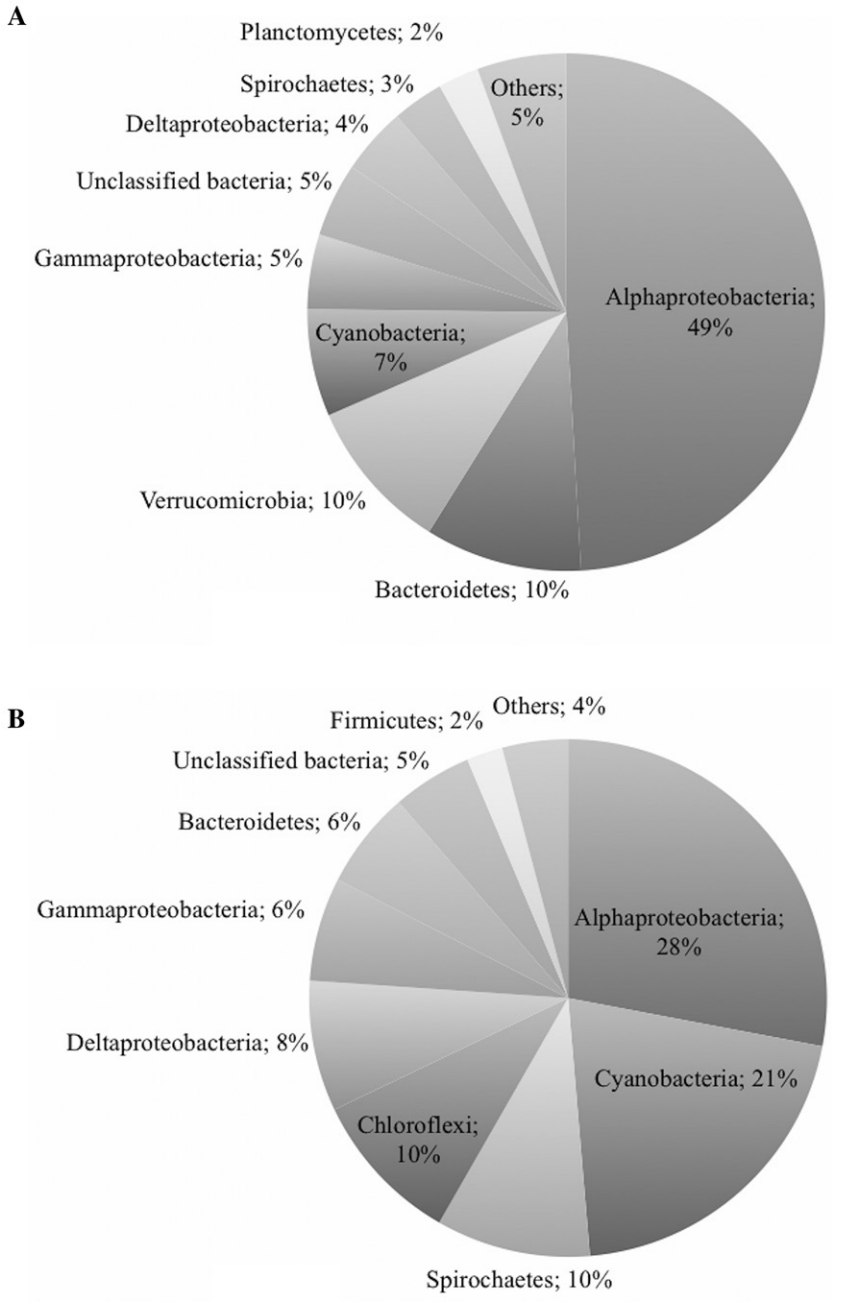

Fig. 2. Bacterial community composition of pustular (A) and smooth (B) microbial mat samples based on 16S ribosomal RNA gene sequence analysis. A total of 8349 sequences were analysed from the pustular mat, and 5706 from the smooth mat.

equal to 50% of total cyanobacterial sequences (Table 1). In the smooth mat sample, *Coleofasciculus* (= *Microcoleus chthonoplastes*) sequences predominated, constituting 69% of all cyanobacterial sequences (Table 1). Novel sequences also contributed substantially to the cyanobacterial sample; 13% of total cyanobacterial rRNA sequences from the smooth mat could not be classified. This suggested the existence of novel stromatolite or hypersaline-associated cyanobacteria, consistent with studies from the Highborne Cay stromatolites (Baumgartner *et al*., 2009) and Guerrero Negro hypersaline microbial mats (Ley *et al*., 2006).

The abundance and community composition of cyanobacterial 16S rRNA gene sequences found is not consistent with that reported in previous work by Allen and colleagues (2009), although in both cases sampling coverage remains incomplete. The proportions of *Euhalothece* and *Spirulina* detected within the pustular mat sample were much higher than previously reported; 50% and 23% respectively in this study versus 4% reported previously for both (Allen *et al*., 2009). In the case of the smooth mat, the proportion of *Coleofasciculus* detected was much higher than previously reported; 69% in this study versus 24% reported previously (Allen *et al*.,

Table 1. Classification by genus of cyanobacterial 16S rRNA gene sequences isolated from pustular and smooth mat samples from Shark Bay.

Pustular mat		Smooth mat	
Genus	Number of sequences	Genus	Number of sequences
Euhalothece	297	*Coleofasciculus*	842
Spirulina	139	Unclassified cyanobacteria	161
Coleofasciculus	50	*Euhalothece*	81
Other *Halothece* cluster	17	*Halothece*	63
Phormidium	9	*Xenococcus*	23
Pleurocapsa	9	*Brasilonema*	13
Merismopedia	7	*Crinalium*	8
Unclassified cyanobacteria	7	*Jaaginema*	7
Brasilonema	7	*Stanieria*	6
Other Oscillatoriales	7	*Aphanothece*	4
Leptolyngbya	6	*Cyanothece*	3
Cyanothece	5	*Cylindrospermum*	3
Stanieria	5	*Halospirulina*	3
Aphanothece	5	Other *Halothece* cluster	2
Geitlerinema sp. BBD	3	*Rivularia*	2
Limnothrix	3	*Lyngbya*	2
Plectonema	3	*Merismopedia*	1
Tolypothrix	2	*Leptolyngbya*	1
Cyanospira	2	*Plectonema*	1
Cylindrospermum	2	*Spirulina*	1
Other Nostocales	2		
Jaaginema	2		
Lyngbya	2		
Coleodesmium	1		
Myxosarcina	1		
Xenococcus	1		
Total	594	Total	1227

2009). A notable absence from this study was *Chroococcidiopsis*; this genus was previously found to account for 25% of all pustular mat sequences and 5% of all smooth mat sequences reported by Allen and colleagues (2009), but was absent from both mat types in the present study. Several other genera detected by Allen and colleagues (2009) were not detected in one or both mat types in the present study. These include *Gleocapsa* and *Halothece*, which each accounted for 13% of all pustular mat sequences isolated by Allen and colleagues (2009), but were not detected at all in the pustular mat in this study (Table 1). The samples used by Allen and colleagues (2009) were collected in November and December 2002, whereas the samples used in the present study were collected in May 2010. The differences in community composition observed could therefore be due to seasonal differences, or changes in environmental conditions that have occurred between 2002 and 2010, as well being a direct consequence of the different sampling, extraction and analytical techniques used, and the substantial difference in number of sequences analysed. In addition, the degree of community variability over small spatial scales between subsamples of the same morphological mat type is unknown, and may be responsible for some of the differences observed between samples.

Primer design and PCR amplification of hpnP

To ensure maximum detection of *hpnP* homologues, multiple degenerate primer sets were designed to amplify both cyanobacterial and other bacterial sequences. The primer sets were demonstrated to amplify a cyanobacterial *hpnP* from a control culture of *Nostoc punctiforme* PCC 73102, and alphaproteobacterial *hpnP* genes from an external environment control, a soil rhizosphere genomic DNA extract. Degenerate primer sets hpnpF1 plus hpnpCR5, as well as hpnpF1 plus hpnpR2B both amplified the *N. punctiforme hpnP* sequence. From the soil rhizosphere extract, primer set hpnpF1 plus hpnpCR5 amplified a sequence that, when analysed phylogenetically, clustered with known *Bradyrhizobium* sequences, while hpnpF1 plus hpnpR2B amplified a sequence that clustered with known *Nitrobacter* sequences. Even with these efforts to ensure complete *hpnP* homologue coverage, the majority of sequences amplified from the Shark Bay mat extracts grouped with the previously described cyanobacteria *hpnP* sequences.

hpnP *clone library analysis*

Amplification of partial *hpnP* genes and clone library construction was carried out using total community genomic

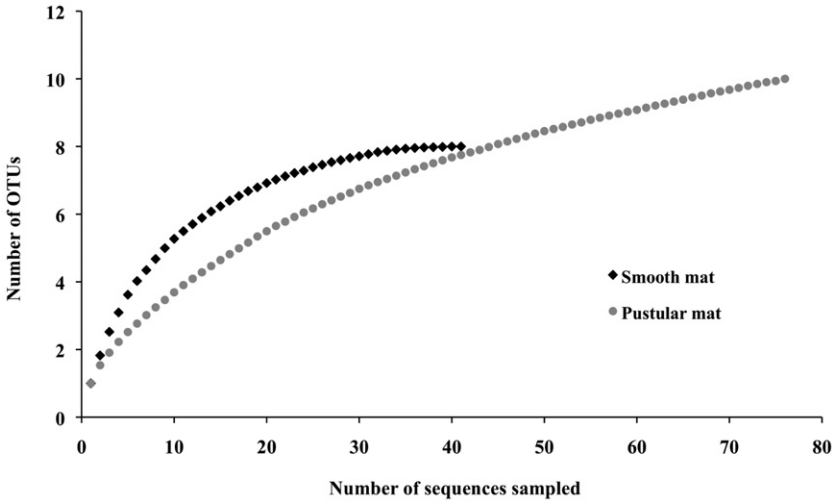

Fig. 3. Rarefaction curves for the clone library *hpnP* sequence data sets isolated from the pustular and smooth mats, generated with OTUs grouped at 95% sequence identity.

DNA from both pustular and smooth mat samples from Hamelin Pool. PCR amplification of *hpnP* segments was also attempted from total genomic DNA extracts of tufted mat and columnar stromatolite samples, but was unsuccessful. A total of 117 partial *hpnP* clones were sequenced from the smooth and pustular mat DNA extracts. Forty-one partial *hpnP* clones were sequenced from the smooth mat library, representing eight OTUs at a 95% level of nucleotide sequence identity. Seventy-six partial *hpnP* clones were sequenced from the pustular mat library, representing 10 OTUs at a 95% nucleotide sequence identity. A total of 12 distinct OTUs were detected from both samples combined, as some OTUs were found in both the pustular and smooth mat samples. Six OTUs were detected in both samples, two were unique to the smooth mat sample and four unique to the pustular mat sample.

Rarefaction curves generated for both *hpnP* libraries indicated that for the population of sequences the *hpnP* primers were able to amplify, sampling coverage was complete for the smooth mat library, but not for the pustular mat library (Fig. 3). This indicated that the pustular mat sample has greater *hpnP* gene diversity than the smooth mat sample. This is consistent with the higher cyanobacterial genus richness observed within the pustular mat sample, even though the smooth mat sample had a greater overall proportion of cyanobacterial 16S rRNA gene sequences (Table 1). The greater diversity of *hpnP* genes in pustular mats may also be the result of environmental factors, as these mats are known to undergo prolonged periods of desiccation, and 2-methylhopanoids have a proposed role in stress tolerance (Rashby *et al.*, 2007; Welander *et al.*, 2009).

Phylogenetic analysis of HpnP sequences

Representative sequences from each of the *hpnP* sequence OTUs isolated from the pustular and smooth mat clone libraries were translated and placed into an amino acid alignment with known HpnP sequences (Fig. 4). The alignment shows that the partial sequence of HpnP is highly conserved, indicating that 2-methylhopanoids likely play an active role in the physiology of the bacteria in which *hpnP* genes are found. The most divergent amino acid sequences of HpnP between cyanobacteria and alphaproteobacteria share 57% sequence identity over this alignment, and the most similar share 67% sequence identity. The most divergent cyanobacterial sequences (including all of those characterized in this study) share 72% sequence identity.

An amino acid alignment of representative translated partial *hpnP* sequences isolated in this study along with previously identified HpnP sequences was used to generate a maximum likelihood tree (Fig. 5). Eleven of the 12 *hpnP* OTUs from pustular and smooth mat samples grouped within the cyanobacteria clade. According to the 16S rRNA gene data analysis, the alphaproteobacteria represented the largest group of bacteria in both mat samples (Fig. 2). Therefore, it was expected that if any *hpnP* genes belonging to alphaproteobacteria were present, they would be amplified using the highly degenerate PCR primers. However, nearly all isolated sequences grouped with the cyanobacteria, the exception being those from smooth mat OTU 1 (two sequences), which clustered with neither cyanobacteria nor alphaproteobacteria. Additionally, none of the alphaproteobacteria genera in which *hpnP* sequences have been identified to date were detected in either the pustular or smooth mat samples (Table S1). These results indicate that the alphaproteobacteria present likely do not contribute significantly to the 2-methylhopanoid production occurring in Shark Bay microbial mats.

Although many of the genera of cyanobacteria present in the pustular and smooth mat samples were identified (Table 1), we are presently unable to reliably assign the

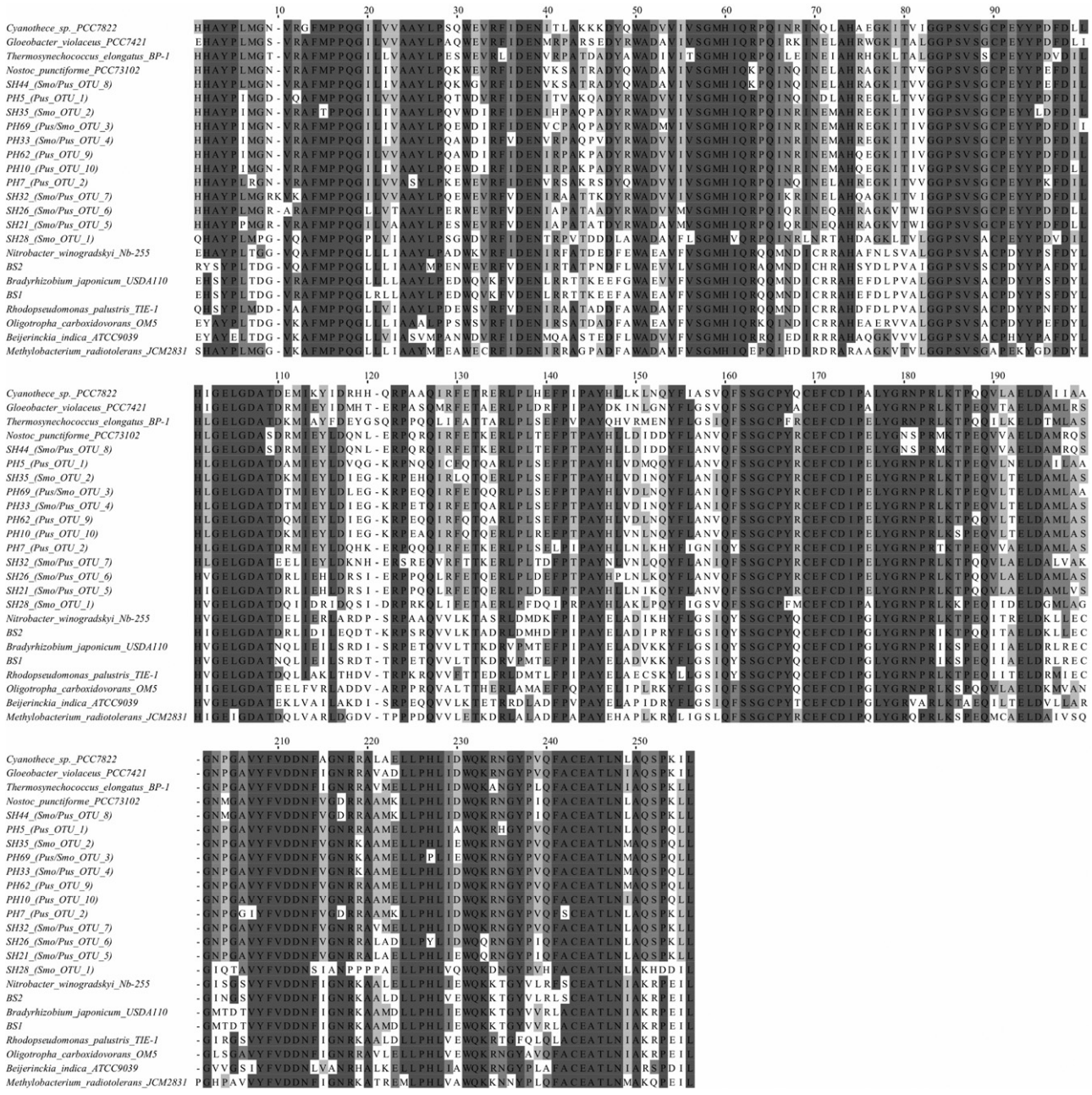

Fig. 4. Amino acid alignment of partial translated *hpnP* sequences from selected cyanobacterial and alphaproteobacteria species, and translated representative sequences from each *hpnP* OTU isolated from the pustular and smooth mat samples. Sequences from the mat samples were amplified using primers hpnpF1 and hpnpCR5.

HpnP sequences isolated from these samples to any phylogenetic group lower than cyanobacteria based on the phylogeny of HpnP sequences from known sources, as presented in Fig. 5. The genomes of the cyanobacteria present in these mat samples have not been sequenced; hence a larger database of known *hpnP* sequences, against which to compare novel sequences, is required in order to propose more useful taxonomic information regarding the distribution of 2-methylhopanoids in these environmental samples. Several species, representing different genera of cyanobacteria, have previously tested positively for 2-methylhopanoids (reviewed in Talbot *et al*., 2008). Species from *Cyanothece*, *Halothece* (*Synechococcus* ATCC 29534) and *Phormidium* have been shown to produce 2-methylhopanoids (Summons *et al*., 1999; Jahnke *et al*., 2004). Representatives from *Cyanothece* were found in both the smooth and pustular mats, representatives from *Halothece* were found in the smooth mat,

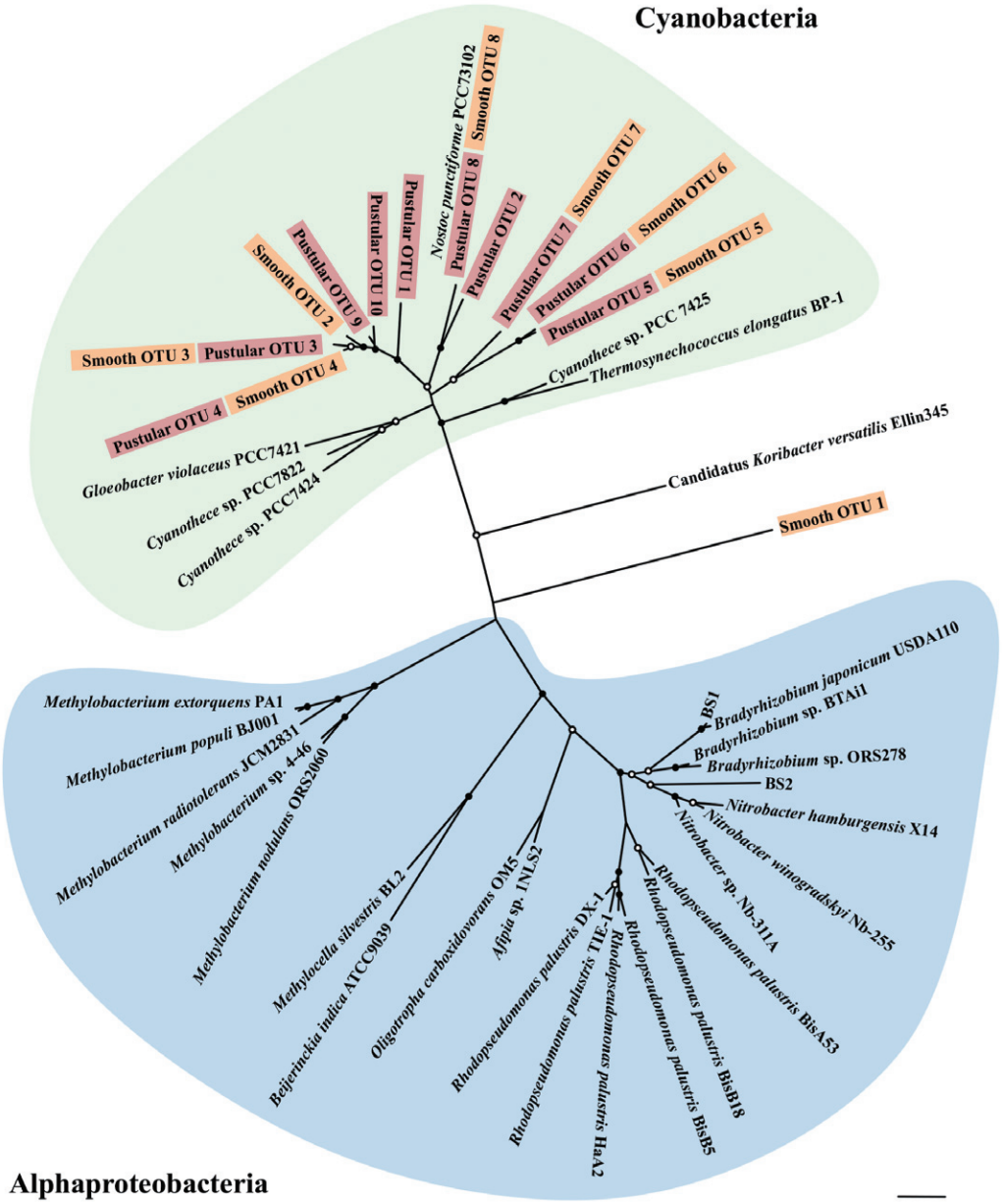

Fig. 5. Unrooted phylogenetic tree based on an amino acid alignment of translated *hpnP* sequences isolated in this study and previously identified *hpnP* sequences, using maximum likelihood methods. Representative sequences from each of the OTUs found within the pustular and smooth mat samples were used for alignment and tree construction. The two major groups, cyanobacteria and alphaproteobacteria, are indicated. Sequences isolated from the pustular mat sample are shown in pink, while those from the smooth mat sample are shown in orange. Scale bar represents 0.1 substitutions per nucleotide site. Filled circles at nodes represent aLRT support values of > 0.90 and empty circles at nodes represent aLRT support values of 0.60–0.90.

and representatives from *Phormidium* were found in the pustular mat (Table 1). Although members from these three genera did not represent a large portion of the bacterial community within the mats, species from *Cyanothece*, *Halothece* and *Phormidium* are nonetheless possible contributors to the 2-methylhopanoids previously detected in samples from these mat types (Allen *et al.*, 2010). However, it is likely that there are other cyanobacteria within the mats also producing 2-methylhopanoids, as to date relatively few species have actually been tested for the presence of 2-methylhopanoids or the genetic capability to produce them. Screening of cultured cyanobacterial species isolated from Shark Bay or from other hypersaline environments for 2-methylhopanoids would be a useful step in determining which species of cyanobacteria are producing these compounds in this environment.

Conclusions

The discovery that cyanobacteria are not the only group of bacteria capable of producing significant amounts of 2-methylhopanoids (Rashby *et al*., 2007) has challenged the interpretation of 2-methylhopanes as biomarkers for cyanobacteria in ancient sediments. More information about the present day function and distribution of 2-methylhopanoids is required in order to accurately determine from what types of organism–environment combinations ancient 2-methylhopanes are likely to have originated. However, based on this initial study, the majority of 2-methylhopanoid producing bacteria living within pustular and smooth microbial mats in the Hamelin Pool area of Shark Bay are cyanobacteria. Other 2-methylhopanoid-producing bacteria appear to constitute a very small portion of the total bacterial community of the mats. As Shark Bay is one of the best modern analogues for the type of microbial mat-rich environments that many ancient sediments are thought to have originated from, this provides support for a cyanobacterial origin of the 2-methylhopane biomarkers found within those ancient sediments.

Experimental procedures

Environmental sample collection

Samples of stromatolite-associated microbial mats were collected with permission from the Department of Environment and Conservation, from Hamelin Pool, Shark Bay, Western Australia. Collection was performed on 8 May 2010 at Carbla Point (26.26374°S, 114.2159°E), on the eastern shore of Hamelin Pool. Approximately 1.5 cm deep cores were taken from smooth, pustular and tufted mat types from the intertidal region, and cores of colloform mat from the tops of large domes were collected in the sub-tidal region. All samples were collected with sterile plastic or aluminum corers, then transferred immediately to sterile plastic jars and saturated in RNAlater (Ambion). Samples were left at ambient temperature overnight, then stored at –20°C until processing.

DNA extraction

Microbial mat samples were homogenized using a sterile mortar and pestle. Approximately 400 mg (wet weight) of material was used for DNA extraction. Extractions were carried out using a FastDNA® Spin Kit for Soil (MP Biomedicals) according to the manufacturer's instructions. DNA concentration and purity was measured with a NanoDrop 1000 spectrophotometer (Thermo Scientific). Concentrations of contaminants in these extracts were found to be too high for PCR amplification, so additional purification steps were performed. Briefly, Tris-EDTA buffer was added to the initial DNA extracts to bring the total volume up to 535 μl. To this, 5 μl of proteinase K (Amresco) (50 mg ml^{-1}) and 60 μl of SDS (20%) were added, and the solution left to incubate at 37°C for 1 h. After incubation, 150 μl of NaCl (5 M) and 0.1 volumes of 10% CTAB were added. The solution was incubated at 65°C for 10 min, and then extracted twice with chloroform: isoamyl alcohol (24:1). DNA was precipitated with one volume of isopropanol, the resultant pellet washed with 70% ethanol, and resuspended in Tris-EDTA buffer.

16S rRNA gene sequence data analysis

The bacterial community composition of each sample was determined by 454 ribosomal RNA gene tagged pyrosequencing at the Research and Testing Laboratory (Lubbock, TX). Bacteria-specific primers 27F (Weisburg *et al*., 1991) and 519R (Lane *et al*., 1985) were used to amplify a ~ 500 bp product spanning the V1–V3 variable regions of the bacterial 16S rRNA gene. Sequence analysis was performed using the mothur v.1.20.1 suite of programs (Schloss *et al*., 2009). Sequences were aligned to the SILVA 16S rRNA reference bacterial alignment, and trimmed so that all sequences covered the same region without gaps. Potential chimeric sequences were removed using UCHIME (Edgar *et al*., 2011). The final alignments contained 5706 sequences from the smooth mat and 8349 sequences from the pustular mat samples, of about 300 bp in length. These were classified using a Bayesian classifier method (Wang *et al*., 2007) implemented by mothur, against the SILVA 16S rRNA gene reference bacterial alignment. Sequences were classified to the level of phylum, with taxonomy for the sequences in the alignment assigned according to the RDP database, with confidence estimates of 60 or higher. Alphaproteobacterial sequences were further classified by this method to the level of genus, and cyanobacterial sequences were also classified to the level of genus, but using the taxonomy assigned by the NCBI database. Rarefaction curves and Shannon diversity indices were generated for each sample data set.

Primer design and PCR amplification of hpnP

An alignment of 30 known *hpnP* sequences obtained from public databases was used to design multiple forward and reverse degenerate primers (Table S2) that amplify segments of the gene approximately 530–1070 bp long, incorporating the region in which the active site cysteine residues are found. Known *hpnP* sequences, along with other types of radical SAM methylases, were aligned using ClustalW (Chenna *et al*., 2003) within Geneious v5.4.6 (Drummond *et al*., 2011). Highly conserved *hpnP*-specific segments of sequence were identified from this alignment, and the consensus sequence of these segments generated by Geneious was used by the authors as a guide to design degenerate *hpnP*-specific primers.

Partial *hpnP* genes were amplified from environmental DNA extracts; typical cycling conditions consisted of an initial denaturing step of 94°C for 3 min, followed by 35 cycles of denaturing for 20 s at 94°C, annealing for 30 s at 58°C, and extension at 72°C for 45 s, with a final extension step of 7 min, and then a hold at 20°C until the amplification products were analysed by agarose gel electrophoresis. Bands of the appropriate size were excised from the gel and purified using a Zymoclean™ gel DNA recovery kit, according to the manufacturer's instructions, and eluted in a final volume of 8 μl of distilled water. DNA products were then either sequenced

directly using the primers with which they were amplified, or cloned into vector as outlined below.

Clone library construction and sequencing

Clone libraries of the partial *hpnP* sequences amplified from the pustular and smooth mat DNA extractions were constructed. To facilitate TA cloning, adenine nucleotides were added to the 3′ ends of PCR products following gel purification. A 7 μl reaction containing 5.65 μl of purified PCR product, 0.7 μl of 10× buffer, 0.25 μl of 50 mM $MgCl_2$, 0.2 μl of 10 mM dNTPs, and 0.2 μl of 1 U μl^{-1} *Taq* polymerase was incubated at 72°C for 10 min. This reaction mixture was then used directly in a cloning reaction using a TOPO-TA cloning kit (Invitrogen) according to the manufacturer's instructions. The ligation mix was used to transform 5-alpha competent *Escherichia coli* cells (NEB) according to the manufacturer's instructions. Cells were spread onto Luria–Bertani (LB) agar plates containing ampicillin, Xgal and IPTG, and incubated overnight at 37°C. Individual transformed colonies were screened for the *hpnP* insert by PCR using M13 forward and reverse primers. Colonies that were positive for the insert were used to inoculate 3 ml of LB media, and incubated overnight at 37°C with shaking. Plasmids were extracted from overnight cultures using a PureLink™ quick plasmid miniprep kit (Invitrogen). Plasmid inserts were sequenced at the Ramaciotti Centre for Gene Function Analysis (University of New South Wales) using an Applied Biosystems 3730 DNA Analyzer.

Representative sequences from the *hpnP* clone libraries are available under GenBank Accession numbers JQ434076–JQ434087.

hpnP *bioinformatics and phylogenetic analyses*

Sequences isolated from the pustular and smooth mat *hpnP* clone libraries were grouped into OTUs based on a 95% level of nucleotide sequence identity. Rarefaction curves for each clone library data set were generated using mothur v.1.20.1 (Schloss *et al*., 2009) to determine if the sampling coverage of each was complete. A single representative sequence from each OTU found in each of the two samples was used for alignments and tree-building. Nucleotide sequences obtained using the primers hpnpF1 and hpnpCR5 were translated into amino acid sequences and aligned with previously described HpnP sequences using ClustalW, then checked manually using Bioedit (Hall, 1999). FindModel (Tao *et al*., 2005) was used to analyse the alignments and determine the most suitable phylogenetic model for the data set. Maximum likelihood trees were constructed from the amino acid alignments with PhyML (Guindon and Gascuel, 2003), using the LG model, with aLRT SH-like branch support.

Acknowledgements

The authors would like to acknowledge the assistance provided by the Western Australia Department of Environment and Conservation, which made possible the collection of samples from within the Shark Bay World Heritage area. This research was funded by Australian Research Council Grants DP1093106 and FF0883440.

References

Allen, M.A., Goh, F., Burns, B.P., and Neilan, B.A. (2009) Bacterial, archaeal and eukaryotic diversity of smooth and pustular microbial mat communities in the hypersaline lagoon of Shark Bay. *Geobiology* **7:** 82–96.

Allen, M.A., Neilan, B.A., Burns, B.P., Jahnke, L.L., and Summons, R.E. (2010) Lipid biomarkers in Hamelin Pool microbial mats and stromatolites. *Org Geochem* **41:** 1207–1218.

Baumgartner, L.K., Spear, J.R., Buckley, D.H., Pace, N.R., Reid, R.P., Dupraz, C., and Visscher, P.T. (2009) Microbial diversity in modern marine stromatolites, Highborne Cay, Bahamas. *Environ Microbiol* **11:** 2710–2719.

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.

Bisseret, P., Zundel, M., and Rohmer, M. (1985) Prokaryotic triterpenoids 2. 2β-methylhopanoids from *Methylobacterium organophilum* and *Nostoc muscosum*, a new series of prokaryotic triterpenoids. *Eur J Biochem* **150:** 29–34.

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

Bravo, J.M., Perzl, M., Hartner, T., Kannenberg, E.L., and Rohmer, M. (2001) Novel methylated triterpenoids of the gammacerane series from the nitrogen-fixing bacterium *Bradyrhizobium japonicum* USDA 110. *Eur J Biochem* **268:** 1323–1331.

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.

Brocks, J.J., Buick, R., Logan, G.A., and Summons, R.E. (2003a) Composition and syngeneity of molecular fossils from the 2.78 to 2.45 billion-year-old Mount Bruce Supergroup, Pilbara Craton, Western Australia. *Geochim Cosmochim Acta* **67:** 4289–4319.

Brocks, J.J., Buick, R., Summons, R.E., and Logan, G.A. (2003b) A reconstruction of Archean biological diversity based on molecular fossils from the 2.78 to 2.45 billion-year-old Mount Bruce Supergroup, Hamersley Basin, Western Australia. *Geochim Cosmochim Acta* **67:** 4321–4335.

Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T.J., Higgins, D.G., and Thompson, J.D. (2003) Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res* **31:** 3497–3500.

Doughty, D.M., Hunter, R.C., Summons, R.E., and Newman, D.K. (2009) 2-Methylhopanoids are maximally produced in akinetes of *Nostoc punctiforme*: geobiological implications. *Geobiology* **7:** 524–532.

Doughty, D.M., Coleman, M.L., Hunter, R.C., Sessions, A.L., Summons, R.E., and Newman, D.K. (2011) The RND-family transporter, HpnN, is required for hopanoid localization to the outer membrane of *Rhodopseudomonas palustris* TIE-1. *Proc Natl Acad Sci USA* **108:** E1045–E1051.

Drummond, A.J., Ashton, B., Buxton, S., Cheung, M., Cooper, A., Duran, C., *et al*. (2011) Geneious v5.4. [WWW document]. URL http://www.geneious.com/

Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C., and Knight, R. (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27:** 2194–2200.

Eigenbrode, J.L., Freeman, K.H., and Summons, R.E. (2008) Methylhopane biomarker hydrocarbons in Hamersley Province sediments provide evidence for Neoarchean aerobiosis. *Earth Planet Sci Lett* **273:** 323–331.

Flannery, D.T., and Walter, M.R. (2011) Archean tufted microbial mats and the Great Oxidation Event: new insights into an ancient problem. *Aust J Earth Sci* **59:** 1–11.

Guindon, S., and Gascuel, O. (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52:** 696–704.

Hall, T.A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* **41:** 95–98.

Jahnke, L.L., Embaye, T., Hope, J., Turk, K.A., Van Zuilen, M., Des Marais, D.J., *et al*. (2004) Lipid biomarker and carbon isotopic signatures for stromatolite-forming, microbial mat communities and *Phormidium* cultures from Yellowstone National Park. *Geobiology* **2:** 31–47.

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.

Jurgens, U.J., Simonin, P., and Rohmer, M. (1992) Localization and distribution of hopanoids in membrane systems of the cyanobacterium *Synechocystis* PCC-6714. *FEMS Microbiol Lett* **92:** 285–288.

Kannenberg, E.L., and Poralla, K. (1999) Hopanoid biosynthesis and function in bacteria. *Naturwissenschaften* **86:** 168–176.

Lane, D.J., Pace, B., Olsen, G.J., Stahl, D.A., Sogin, M.L., and Pace, N.R. (1985) Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proc Natl Acad Sci USA* **82:** 6955–6959.

Ley, R.E., Harris, J.K., Wilcox, J., Spear, J.R., Miller, S.R., Bebout, B.M., *et al*. (2006) Unexpected diversity and complexity of the Guerrero Negro hypersaline microbial mat. *Appl Environ Microbiol* **72:** 3685–3695.

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

Ourisson, G., and Rohmer, M. (1992) Hopanoids. 2. Biohopanoids – a novel class of bacterial lipids. *Acc Chem Res* **25:** 403–408.

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

Pearson, A., and Rusch, D.B. (2009) Distribution of microbial terpenoid lipid cyclases in the global ocean metagenome. *ISME J* **3:** 352–363.

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.

Pearson, A., Leavitt, W.D., Saenz, J.P., Summons, R.E., Tam, M.C.M., and Close, H.G. (2009) Diversity of hopanoids and squalene-hopene cyclases across a tropical land-sea gradient. *Environ Microbiol* **11:** 1208–1223.

Poralla, K., Hartner, T., and Kannenberg, E. (1984) Effect of temperature and pH on the hopanoid content of *Bacillus acidocaldarius*. *FEMS Microbiol Lett* **23:** 253–256.

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.

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.

Renoux, J.M., and Rohmer, M. (1985) Prokaryotic triterpenoids: new bacteriohopanetetrol cyclitol ethers from the methylotrophic bacterium *Methylobacterium organophilum*. *Eur J Biochem* **151:** 405–410.

Sahm, H., Rohmer, M., Bringermeyer, S., Sprenger, G.A., and Welle, R. (1993) Biochemistry and physiology of hopanoids in bacteria. *Adv Microb Physiol* **35:** 247–273.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., *et al*. (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75:** 7537–7541.

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

Schopf, J.W., and Walter, M.R. (1983) Archean microfossils: new evidence of ancient microbes. In *The Earth's Earliest Biosphere: Its Origin and Evolution*. Schopf, J.W. (ed.). Princeton, NJ, USA: Princeton University Press, pp. 214–239.

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

Sogin, M.L., Morrison, H.G., Huber, J.A., Mark Welch, D., Huse, S.M., Neal, P.R., *et al*. (2006) Microbial diversity in the deep sea and the underexplored 'rare biosphere'. *Proc Natl Acad Sci USA* **103:** 12115–12120.

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.M., Rohmer, M., and Farrimond, P. (2007) 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., 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.

Tao, N., Richardson, R., Bruno, W., and Kuiken, C. (2005) FindModel. [WWW document]. URL http://www.hiv.lanl.gov/content/sequence/findmodel/findmodel.html

Vilcheze, C., Llopiz, P., Neunlist, S., Poralla, K., and Rohmer, M. (1994) Prokaryotic triterpenoids – new hopanoids from the nitrogen-fixing bacteria *Azotobacter vinelandii, Beijerinckia indica* and *Beijerinckia mobilis*. *Microbiology* **140:** 2749–2753.

Waldbauer, J.R., Sherman, L.S., Sumner, D.Y., and Summons, R.E. (2009) Late Archean molecular fossils

from the Transvaal Supergroup record the antiquity of microbial diversity and aerobiosis. *Precambrian Res* **169:** 28–47.

Wang, Q., Garrity, G.M., Tiedje, J.M., and Cole, J.R. (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* **73:** 5261–5267.

Weisburg, W.G., Barns, S.M., Pelletier, D.A., and Lane, D.J. (1991) 16S ribosomal DNA amplification for phylogenetic study. *J Bacteriol* **173:** 697–703.

Welander, P.V., Hunter, R.C., Zhang, L.C., Sessions, A.L., Summons, R.E., and Newman, D.K. (2009) Hopanoids play a role in membrane integrity and pH homeostasis in *Rhodopseudomonas palustris* TIE-1. *J Bacteriol* **191:** 6145–6156.

Welander, P.V., Coleman, M.L., Sessions, A.L., Summons, R.E., and Newman, D.K. (2010) Identification of a methylase required for 2-methylhopanoid production and implications for the interpretation of sedimentary hopanes. *Proc Natl Acad Sci USA* **107:** 8537–8542.

Welander, P.V., Doughty, D.M., Wu, C.H., Mehay, S., Summons, R.E., and Newman, D.K. (2012) Identification and characterization of *Rhodopseudomonas palustris* TIE-1 hopanoid biosynthesis mutants. *Geobiology* **10:** 163–177.

Supporting information

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

Table S1. Classification by genus of alphaproteobacterial 16S rRNA gene sequences isolated from pustular and smooth mat samples from Shark Bay.

Table S2. Nucleotide sequences and corresponding amino acid motifs of oligonucleotide primers used in this study.