npg

ORIGINAL ARTICLE

Life in the dark: metagenomic evidence that a microbial slime community is driven by inorganic nitrogen metabolism

Sasha G Tetu[1], Katy Breakwell[1], Liam DH Elbourne[1], Andrew J Holmes[2], Michael R Gillings[3] and Ian T Paulsen[1]

[1]Department of Chemistry and Biomolecular Sciences, Macquarie University, Sydney, New South Wales, Australia; [2]School of Molecular Bioscience, Sydney University, Sydney, New South Wales, Australia and [3]Department of Biological Sciences, Macquarie University, Sydney, New South Wales, Australia

Beneath Australia's large, dry Nullarbor Plain lies an extensive underwater cave system, where dense microbial communities known as 'slime curtains' are found. These communities exist in isolation from photosynthetically derived carbon and are presumed to be chemoautotrophic. Earlier work found high levels of nitrite and nitrate in the cave waters and a high relative abundance of Nitrospirae in bacterial 16S rRNA clone libraries. This suggested that these communities may be supported by nitrite oxidation, however, details of the inorganic nitrogen cycling in these communities remained unclear. Here we report analysis of 16S rRNA amplicon and metagenomic sequence data from the Weebubbie cave slime curtain community. The microbial community is comprised of a diverse assortment of bacterial and archaeal genera, including an abundant population of Thaumarchaeota. Sufficient thaumarchaeotal sequence was recovered to enable a partial genome sequence to be assembled, which showed considerable synteny with the corresponding regions in the genome of the autotrophic ammonia oxiser *Nitrosopumilus maritimus* SCM1. This partial genome sequence, contained regions with high sequence identity to the ammonia mono-oxygenase operon and carbon fixing 3-hydroxypropionate/4-hydroxybutyrate cycle genes of *N. maritimus* SCM1. Additionally, the community, as a whole, included genes encoding key enzymes for inorganic nitrogen transformations, including nitrification and denitrification. We propose that the Weebubbie slime curtain community represents a distinctive microbial ecosystem, in which primary productivity is due to the combined activity of archaeal ammonia-oxidisers and bacterial nitrite oxidisers.

Introduction

Life in the dark recesses of the planet comes in many strange forms. Exploration of these environments provides key insights into the metabolic flexibility of micro-organisms that do not rely on sunlight for energy. Along the southwestern coast of Australia lies the Nullarbor Plain, the world's largest contiguous karst system, covering more than $200\,000\,km^2$, which arose from the sea in the Middle Miocene (~ 14 MYA) (Webb and James, 2006). Although the surface environment is semiarid, a number of aquatic, subterranean caves extend to considerable distances below the Nullarbor karst, intersecting the

water table at a depth of around 100 m below the surface. Exploration of these underwater passages by cave divers has revealed the presence of unusual microbial biofilms, dense 'mantles' or 'curtains' of biological material, known as Nullarbor cave slimes (Holmes *et al.,* 2001). Cave slime communities consist of long, tendril-like structures or amorphous biofilms attached to cave surfaces. These microbial formations are observed in areas of complete darkness, deep within submerged regions of the cave systems. They comprise a high standing biomass and can be up to 1 m in length (Holmes *et al.,* 2001). No aquatic macrofauna has been recorded in these regions of the caves (Richards, 1971). The cave waters are unlikely to receive significant organic input from the surface, as rainfall in these regions is typically very low, ranging from 150–200 mm annually (Gillieson and Spate, 1992). Indeed, filtered water samples collected from the vicinity of the microbial material were previously shown to contain

no detectable organic carbon (James and Rogers, 1994). Chemical analysis of the water samples did show, however, that there are relatively high levels of nitrite as well as significant sulphate and nitrate in many of the caves (Holmes et al., 2001). The composition and metabolism of these communities is therefore of particular interest, with the microorganisms comprising these structures presumably capable of some form of chemolithotrophy.

Initial work by Holmes et al. (2001) provided the first assessment of the microbial communities associated with the Nullarbor cave slimes. In this study, the microbial community from one Nullarbor cave, Weebubbie, was sampled and bacterial 16S rRNA gene clone libraries were generated and analysed via restriction fragment length polymorphism mapping and sequencing of selected restriction fragment length polymorphism pattern representatives. The majority of phylotypes were found to belong to Proteobacteria, with *Pseudomonas* and *Pseudoalteromonas* related sequences being particularly abundant. A relatively high proportion ($\sim 12\%$) of clones were classified to the phylum *Nitrospira*, of which all characterised members are able to oxidise nitrite to nitrate (Holmes et al., 2001). This suggested that nitrite oxidation could have a significant role in the trophic structure of the slime communities. There were also a large number of clones which represented novel phylotypes, highlighting the unusual nature of this microbial ecosystem. Examination of the Weebubbie slime community was also carried out using scanning electron microscopy, revealing a dense matrix of unbranched filaments interspersed with spherical-, rod- and spiral-shaped cells and calcite microcrystals. The observation of a microscopic structure most likely comprising an abundant filamentous organism, whose morphology did not match with the known cell shape of any abundant clone type, led to the hypothesis that this environment might also contain additional key organisms not effectively detected using a PCR-based approach.

In this study, we have significantly expanded our understanding of this community by utilising pyrosequencing to generate metagenomic and 16S rRNA amplicon data sets for the Weebubbie microbial cave slime. We have found evidence for an abundant archaeal population with sequence simi-larity to *Nitrosopumilus maritimus* SCM1, an ammonia-oxidising archaeon, thus considerably extending our knowledge of both the phylogenetic composition and metabolic capabilities of this unusual community.

Materials and methods

Site description and sample preparation

Weebubbie cave is located in the Nullarbor Plain region of South Australia in the middle of the Great Australian Bight (Latitude -31.654180526, Longitude 128.774993896). The entrance of the cave is via a 70 m descent from the surface through a collapsed doline. At the base a ~ 200 m talus filled passage leads to two large lakes, which when traversed, lead into a number of submerged passages where the microbial slime communities are found (Figure 1).

DNA samples used in this work were aliquots of genetic material extracted by Holmes et al. (2001) that had been stored at $-80\,^{\circ}\text{C}$. Descriptions of sample collection, DNA extraction methodology, microscopy and water chemistry analyses are provided in Holmes et al. (2001). Physicochemical characteristics of Weebubbie cave water were moderate salinity falling in the brackish range (16.2%), relatively high levels of sulphate ($1200\,\text{mg}\,\text{l}^{-1}$), nitrate ($98\,\text{mg}\,\text{l}^{-1}$) and nitrite ($5–10\,\text{mg}\,\text{l}^{-1}$).

As a minimum of 1 µg of DNA was needed for metagenomic pyrosequencing, 33 ng of the remaining DNA from Weebubbie samples was amplified using the multiple strand displacement Phi29 DNA polymerase, following the manufacturer's instructions (GenomiPhi V2 DNA Amplification Kit, GE Healthcare, Little Chalfont, UK). Duplicate whole-genome amplifications were performed and pooled to reduce potential amplification bias. Each amplification used 1-µl template and was subsequently cleaned using ethanol precipitation and resuspended in a final volume of 75 µl Tris EDTA buffer. DNA concentrations for each amplification reaction were measured with the Nanodrop 2000 (Thermo Scientific, Bremen, Germany) and ~ 600 ng of DNA from each reaction was pooled, then sent to the Ramaciotti Centre for Gene Function Analysis for sequencing using the GS-FLX pyrosequencing platform (Roche, Mannheim, Germany).

Figure 1 Photographs of Weebubbie cave environment and microbial slime communities. (**a**) Aerial view of Weebubbie cave doline. (**b**) Submerged chamber of Weebubbie cave. (**c**) Image of slime 'tentacles' attached to cave walls and roof.

For 16S rRNA amplicon sequencing, original cave DNA (rather than Phi29-amplified DNA) was used as a template in PCR with the primer set F515 and R806 of Bates *et al.* (2011), which have been shown to successfully amplify an ~250 bp region of the 16S rRNA gene from both bacteria and archaea (Bates *et al.*, 2011). Reactions were performed in triplicate 25 µl reactions using the FastStart High Fidelity PCR System (Roche) according to the manufacturer's instructions with 1 µl template DNA. Reactions were carried out using the following cycling conditions: 35 cycles (95 °C, 30 s; 50 °C, 45 s; 72 °C, 60 s) after an initial denaturation 95 °C, 2 min. Pooled triplicate reactions were run out on a 2% agarose gel and ~250 bp bands were excised and purified using the Qiaquick gel cleanup kit (Qiagen, Valencia, CA, USA) before sequencing at the Ramaciotti Centre for Gene Function Analysis using the GS-FLX pyrosequencing platform (Roche).

Data analysis

Metagenome reads were assembled into contigs at the Ramaciotti Centre for Gene Function Analysis using Newbler assembly software (Margulies *et al.*, 2005). The unassembled sequence file was processed to remove potentially low quality sequences using the mothur software suite (Schloss *et al.*, 2009). Sequences with an average quality score less than 20, below 100 bp in length or containing ambiguous nucleotides were removed.

Exploring taxonomic and metabolic composition of Weebubbie community

Processed unassembled sequences were uploaded to the Metagenome Rapid Annotation using Subsystem Technology (MG-RAST) server, v3.1.2, for annotation (Meyer *et al.*, 2008). A taxonomic profile was generated using the phylogenetic identity of sequence matches to the M5 non-redundant protein database with a maximum e-value cutoff of $1e^{-15}$. A functional comparison was performed using the Weebubbie metagenome, four metagenomes that were publicly available on MG-RAST, and one metagenome that was obtained and uploaded on to MG-RAST as a private metagenome (Supplementary Table 1). Metagenomes were searched against the Subsystem database with a maximum e-value of $1e^{-5}$, using a minimum identity cutoff of 75%. Metabolic profiles were determined from normalised abundances of SEED subsystems, and a heatmap was produced using complete clustering with the Bray-Curtis metric.

For 16S rRNA, amplicon sequencing data reads were processed with both MG-RAST and QIIME software (Caporaso *et al.*, 2010), which binned reads into phylotypes (97% similarity) and assigned an identity based on comparison to sequences in the Ribosomal Database Project (Cole *et al.*, 2005). Taxonomic analysis of the metagenomic data set for comparison with the 16S rRNA data employed PhymmBL (Brady and Salzberg, 2009) and MG-RAST to identify the most probable taxonomic origin of each sequence read. The metagenomic sequence reads and contigs, as well as the 16S rRNA amplicon sequences are publically accessible through the MG-RAST server with the accession numbers 4448052.3, 4447951.3, and 4467995.3 respectively.

Examination of Weebubbie Thaumarchaeota population

The programme PhymmBL (Brady and Salzberg, 2009) was used to identify the most probable taxonomic origin of each contig sequence. Contigs binning to the phylum Thaumarchaeota were aligned against the *Ca.* Cenarchaeum symbiosum (Hallam *et al.*, 2006) genome draft and *N. maritimus* SCM1 (Walker *et al.*, 2010) genome using ProgressiveMauve (Darling *et al.*, 2010). Contigs binning to the phylum Thaumarchaeota were also searched against the Ribosomal Database Project database and two putative thaumarchaeotal 16S rRNA fragments were identified, one of which was of sufficient length for use in phylogenetic analysis (the second, containing 301 bp of 16S rRNA gene sequence at the other end of the gene, spanned insufficient variable regions to allow meaningful phylogenetic analysis). The first contig of 941 bp was aligned with 26 known archaea within the phyla Crenarchaeota, Euryarchaeota and Thaumarchaeota, as well as five uncultured marine Thaumarchaeota. A bacterial representative, *Thermotoga maritima* MSB-8, was included in the alignment as an out-group (accession numbers of all included gene sequences are provided in Supplementary Table 2). Sequences were aligned using ClustalX2 and trimmed to a total of 695 nucleotide positions. Phylogenetic analyses were performed using PhyML 3.0 (Guindon *et al.*, 2010), with the GTR model of nucleotide substitution, utilising a BioNJ tree as a starting tree, and the SPR heuristic implemented for tree topology searches. Confidence values at each node were calculated using the non-parametric bootstrap analysis over 1000 replicates. jModelTest (Darriba *et al.*, 2012) was utilised to assess the effect of variation in the substitution model and other parameters, the topology model-averaged phylogeny resulting from the 280 models in the 100% confidence range was identical to the tree shown in Figure 4.

Exploration of nitrogen cycling pathways in the Weebubbie community

Gene fragments of key enzymes involved in nitrogen metabolism were identified and combined with taxonomic assignations of the sequence reads to produce a metabolic pathway showing the relative abundances of bacteria and archaea contributing to each reaction. PhymmBL was used to identify the most probable taxonomic origin of each read sequence. A local database of representative protein

sequences was generated for key enzymes and processed metagenomic reads were searched against this database using BlastX with an e-value cutoff of $1e^{-15}$. The BlastX results and PhymmBL taxonomic assignments for each enzyme were used to determine the relative taxonomic abundances calculated for each enzyme.

Results

Taxonomic composition of Weebubbie cave microbial slime community

A total of 548 372 raw reads with an average length of 581 bp were generated from metagenomic sequencing of Weebubbie microbial slime community DNA. Following removal of low quality reads and those of less than 100 bp, a total of 475 608 high quality processed read sequences remained in the data set and were included in analyses. To extend the taxonomic analysis of the metagenomic data we additionally generated a 16S rRNA amplicon library; a total of 48 851 fragments spanning the V4 variable region were analysed. In both the metagenome and

the 16S amplicon data sets the Weebubbie slime communities were principally comprised of a mix of bacterial and archaeal members (Figure 2; Supplementary Figure 1). Both analyses showed similar sets of dominant taxa, although there were minor differences in the rank abundance order. The main difference between the 16S rRNA data and the binned metagenomic data was the relative abundances of Archaea compared with Bacteria, with the 16S rRNA data set showing a higher proportion of Archaea. Amongst the Bacteria, the phylum *Proteobacteria* had the highest relative abundance. Within the Proteobacterial assigned sequences, more than half belonged to gammaproteobacteria with almost all the remainder fairly evenly distributed between alpha, beta and delta subdivisions. Looking further into the gammaproteobacteria at a genus level, *Pseudoalteromonas* was the most commonly detected.

Interestingly, in both the metagenomic data and the 16S rRNA data, there is evidence of an abundant population of Thaumarchaeota related to the ammonia oxidising *N. maritimus* SCM1 and *Candidatus* Cenarchaeum symbiosum.
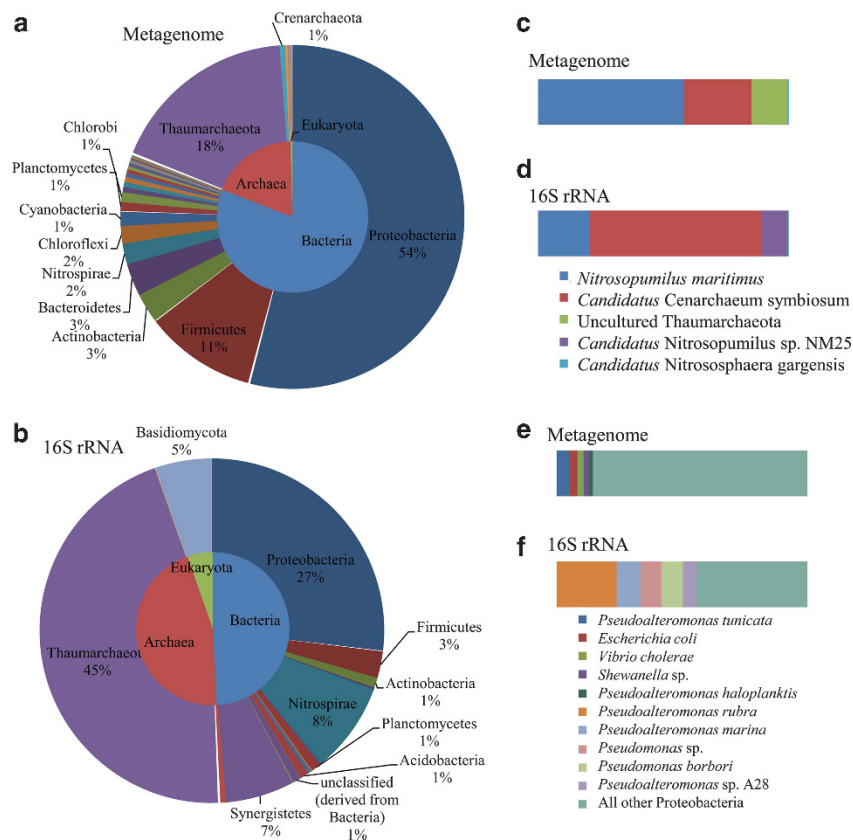

Figure 2 The inferred taxonomic composition of the Weebubbie cave community. Taxonomic designations were assigned at phylum level based on the (**a**) complete metagenomic data set, and (**b**) 16S rRNA amplicon data set; species level for the phylum Thaumarchaeota based on the (**c**) complete metagenomic data set, and (**d**) 16S rRNA amplicon data set; and species level for the phylum Proteobacteria based on (**e**) the complete metagenomic data set, and (**f**) the 16S rRNA amplicon data set. The taxonomic analyses based on the metagenome data used MG-RAST, and sequences were compared with the M5 non-redundant protein database with a maximum e-value cutoff of $1e^{-15}$. Unclassified sequences accounted for 7.31% of metagenome reads, and were not included in this figure. The taxonomic analyses based on the 16S rRNA data used MG-RAST, and sequences were compared with the M5 rRNA non-redundant rRNA database with a maximum e-value cutoff of $1e^{-15}$. Unclassified sequences accounted for 26.35% of 16S rRNA amplicon reads. Full colour breakdown of the organisms represented in the pie charts in (**a**) and (**b**) can be found in Supplementary Figure 1.

Comparison of the Weebubbie thaumarchaeotal population with other studied Thaumarchaeota

As a high proportion of reads were taxonomically assigned to the phylum Thaumarchaeota, and the sequenced genome of *N. maritimus* SCM1 and draft genome of *Ca.* C. symbiosum both recruited a high number of reads, alignments were produced to compare the thaumarchaeotal component of the metagenome with these genomes. From assembly of metagenomic data, a set of 20 233 contigs with a minimum of 4 × coverage, ranging in size from 100 bp to 67 Kbp (1697 contigs > 1 Kbp, 84 contigs > 5 Kbp), was generated. The subset of contigs that binned to Thaumarchaeota was used to generate alignments (which included almost all of the largest contigs, 82/84 of those greater than > 5 Kbp). The global alignment showing the greatest degree of synteny with the Weebubbie thaumarchaeotal contigs was observed with *N. maritimus* SCM1 (Figure 3). Alignment against *Ca.* C. symbiosum genome showed substantially less synteny (Supplementary Figure 2).

It is evident that there are numerous syntenous regions in the Weebubbie Thaumarchaeota contigs compared with *N. maritimus* SCM1 genome, however the similarity profiles (indicated by vertical line height along alignment) indicate nucleic acid sequence similarities were relatively low in many regions (Figure 3a). There are also a number of instances where genes with sequence similarity to those of *N. maritimus* SCM1 were present in the Weebubbie metagenome but showed a different global arrangement (illustrated in Figure 3a by lines connecting regions of each alignment which have shifted in their relative location). Also evident in this alignment are a large number of gaps, representing both sequence present in the *N. maritimus* SCM1 genome and absent in the Weebubbie metagenome data and regions of Weebubbie thaumarchaeotal contigs, which contain genes not found in the *N. maritimus* SCM1 genome (Figure 3b).

Perhaps one of the most interesting regions of the *N. maritimus* SCM1 genome is the ammonia
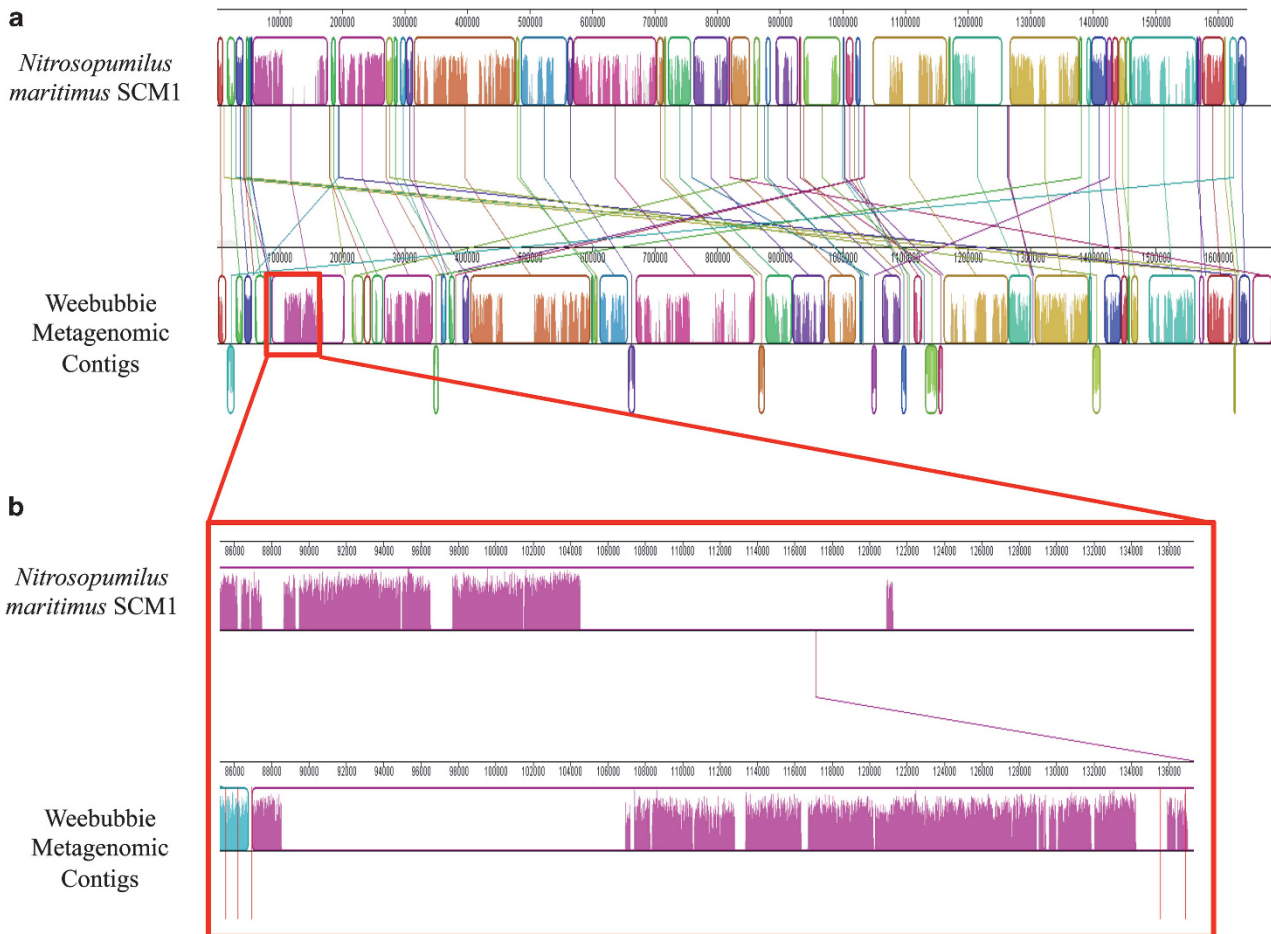

Figure 3 MAUVE alignment of the Weebubbie contigs and *N. maritimus* SCM1 genome. Coloured, boxed sections show regions of synteny, with the connecting lines between the two alignments linking matching regions in the *N. maritimus* SCM1 genome and Weebubbie contigs. The height of the coloured lines within these regions indicates how similar the gene sequences of the matching genes are, with identical gene sequences having a bar that extends to the top border of the region. Gaps indicate regions of sequence present in one genome, but absent in the other. Panel (**a**) shows the complete alignment over the entire *N. maritimus* SCM1 genome while (**b**) shows a subsection, where a region of aligned Weebubbie contigs contains sequence not observed in *N. maritimus* (in this panel extended red vertical lines indicate contig boundaries, with one large contig spanning most of the length of this region).

mono-oxygenase operon, which indicates the potential for these organisms to have a role in nitrogen cycling. In the Weebubbie metagenome contigs were recovered which span large regions of this operon, including a single large contig which covers both the *amoB* and *amoC* genes, and a second contig which contains part of the *amoA* gene. Alignments of each of these translated genes from *N. maritimus* SCM1 and *Ca.* C. symbiosum genome A with the relevant region of Weebubbie contigs showed protein sequence identities of 85% for *amoB* and 95% for *amoC*, while the fragment of *amoA* showed protein identities of 92% and 93% (over 131 of 216 amino acids) with *N. maritimus* SCM1 and *Ca.* C. symbiosum, respectively.

Two contigs containing putative thaumarchaeotal 16S rRNA gene fragments were identified in the Weebubbie metagenome data set, one of which contained a sufficiently long region of 16S rRNA gene identity for phylogenetic analyses. An alignment of this sequence with archaeal representatives from all three lineages Crenarcheaota, Euryarchaeota and Thaumarchaeota was used to generate a neighbour-joining phylogenetic tree (Figure 4). On the basis of this analysis there is good support for placing the Weebubbie 16S rRNA sequence within

the Thaumarchaeota. This sequence groups closely with *Ca.* C. symbiosum and an uncultured archaeal representative recovered from a metagenomic fosmid library of hydrothermal vent water (0.2 μm filtrate, 1000 m depth, Adriatic sea, accession number EU686633.2 (Martin-Cuadrado *et al.,* 2008)) (Figure 4). Also within this group are other well-studied group 1.1a Thaumarchaota, including *N. maritimus* SCM1 and *Ca.* Nitrosoarchaeum limnia SFB1, recently identified in a low-salinity environment in San Francisco Bay (Blainey *et al.,* 2011), while representatives of group 1.1b, *Ca.* Nitrososphaera spp. EN76 and *Ca.* Nitrososphaera gargensis, formed a separate group within the thaumarchaotal branch.

Comparison of metabolic profiles from other habitats
The Weebubbie cave environment has salinity levels in the brackish range, high levels of sulphate, nitrate and nitrite, no detectable organic carbon and a complete absence of sunlight (Holmes *et al.,* 2001). The metabolic profile of this community was compared with that of other aquatic metagenomes, including representatives from a freshwater aquifer located in South Australia, deep-sea and other
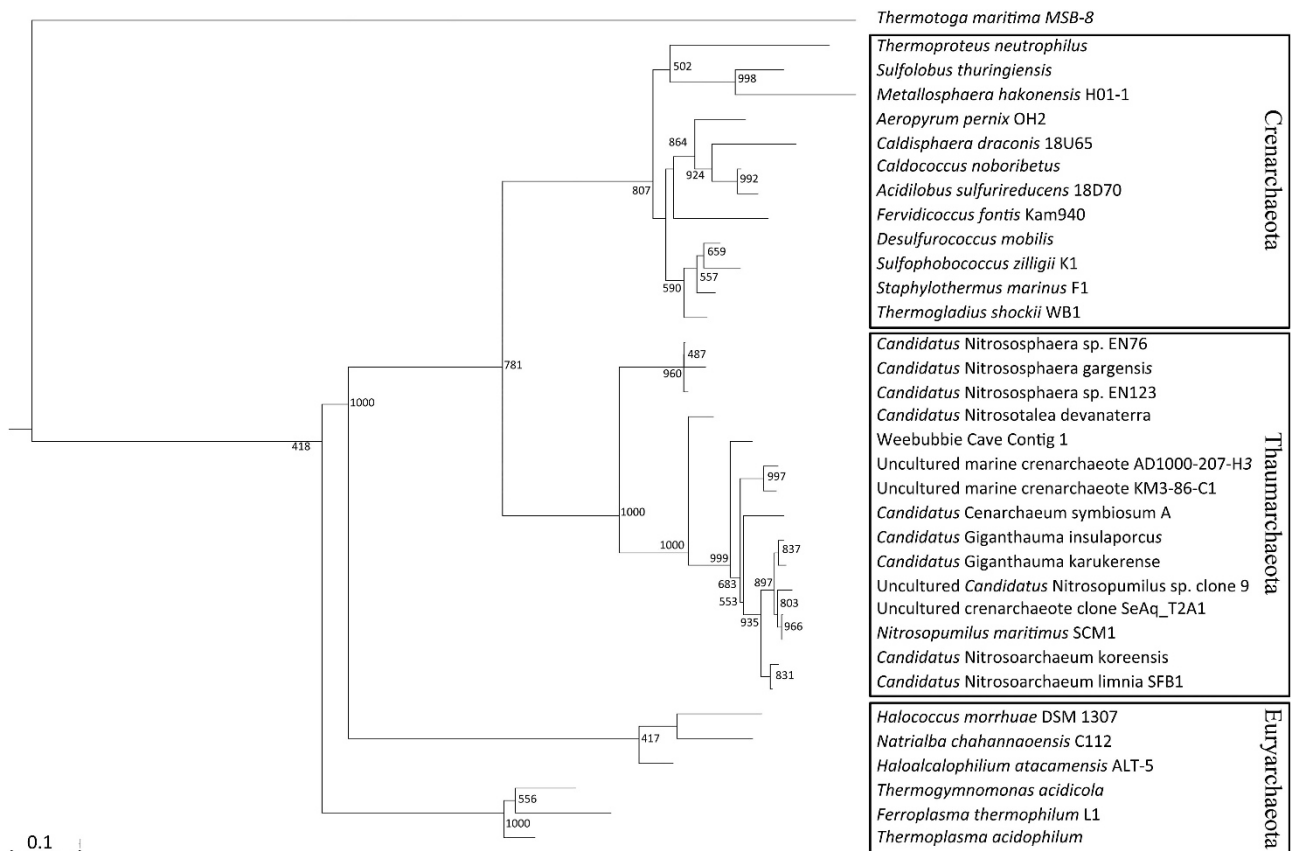

Figure 4 Phylogenetic tree showing the relationship of recovered archaeal Weebubbie cave 16S rRNA gene to that of other archaeal representatives. The maximum likelihood tree was constructed from an alignment of 695 nucleotide positions in the 16S rRNA gene. The tree was inferred with the GTR genetic distance model, using *Thermotoga maritima* MSB-8 as an outgroup. Bootstraps were calculated based on a total of 1000 replications, with bootstrap values displayed at the nodes.
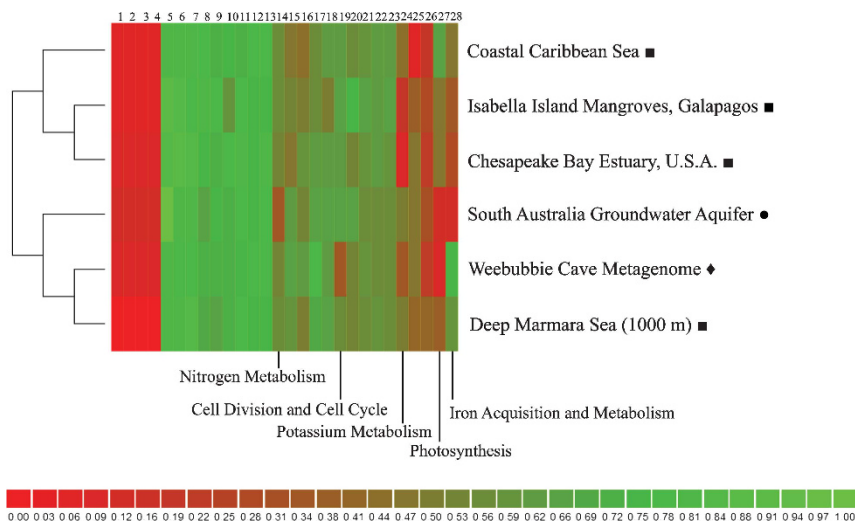
Figure 5 Functional comparison of the Weebubbie cave metagenome with other selected aquatic metagenomes. A complete cluster analysis with Bray-Curtis distances was performed using normalised abundances of SEED Subsystem categories, comparing the Weebubbie cave metagenome (♦) with freshwater (●) and marine (■) metagenomes (scale bar indicates log transformed, normalised abundance values). SEED categories from left to right: (1) Fatty Acids, Lipids, and Isoprenoids; (2) Phages, Prophages, Transposable Elements, Plasmids; (3) Cofactors, Vitamins, Prosthetic Groups, Pigments; (4) Virulence, Disease and Defence; (5) Carbohydrates; (6) Protein Metabolism; (7) Clustering-based Subsystems; (8) Respiration; (9) Stress Response; (10) DNA Metabolism; (11) Amino Acids and Derivatives; (12) Miscellaneous; (13) RNA Metabolism; (14) Nitrogen Metabolism; (15) Cell Wall and Capsule; (16) Metabolism of Aromatic Compounds; (17) Membrane Transport; (18) Nucleosides and Nucleotides; (19) Cell Division and Cell Cycle; (20) Sulphur Metabolism; (21) Motility and Chemotaxis; (22) Regulation and Cell Signalling; (23) Phosphorus Metabolism; (24) Potassium Metabolism; (25) Secondary Metabolism; (26) Dormancy and Sporulation; (27) Photosynthesis; (28) Iron Acquisition and Metabolism.

marine environments (Figure 5). The Weebubbie cave metagenome grouped most closely with the deep Marmara Sea metagenome and together with the South Australia Groundwater Aquifer metagenome form a separate branch from the other examined marine metagenomes, which represent photic zone habitats (Figure 5).

The grouping of the Weebubbie metagenome metabolic profile with that of the Marmara deep-sea environment reflects similar abundances for many SEED categories. The 'iron acquisition and metabolism' group in particular stands out as relatively abundant in these two environments compared with the others analysed (Figure 5). Indeed the relative abundance of iron acquisition gene fragments was even higher in Weebubbie than the deep Marmara Sea, suggesting an environment low in biologically available iron. Examination of the predicted gene products within this subsystem indicated that it includes numerous iron transporter components, receptors and siderophores.

Nitrogen and Carbon Cycling in the Weebubbie Cave Community

Previous examination of the Weebubbie cave environment and slime community indicated that they may represent chemolithotrophic communities reliant on nitrite oxidation (Holmes *et al.*, 2001). The apparent abundance of an archaeal population in our Weebubbie metagenome related to ammonia-oxidising *N. maritimus* SCM1 indicated that ammonia oxidation might also be an important source of

energy in this environment. To gain a more complete appreciation of the nitrogen cycling processes carried out by the Weebubbie slime community the metagenome reads were analysed to determine the abundance and likely taxonomic affiliation of key nitrogen metabolism enzymes (Figure 6). This analysis indicated that genes encoding enzymes for each step of the nitrogen cycle, except nitrogen fixation, were present in this community. Of the reads matching ammonia mono-oxygenase subunit-encoding genes, 82% were most likely of archaeal origin (best match to *N. maritimus* SCM1 *amo* genes), while the remainder were most likely encoded by Nitrosomonadales representatives. A small number of reads matching hydroxylamine oxidase, which carries out the next step in nitrification in bacteria were also observed in the metagenome, further evidence that bacteria were responsible for only a small proportion of ammonia oxidation in this environment. As archaeal ammonia-oxidisers (AOA) are thought to utilise a divergent pathway for this second step, for which there is little information at present (Walker *et al.*, 2010), no candidate gene(s) were available to search for evidence of this part of the process in AOA. There is also evidence for the second stage of nitrification, oxidation of nitrite to nitrate, in a number of sequence hits to nitrate oxidoreductase, with both bacterial and archaeal forms observed in the metagenome. A relatively large number of putative nitrate reductase encoding genes were also observed in the Weebubbie metagenome sequence, the majority of which were likely bacterial in origin.
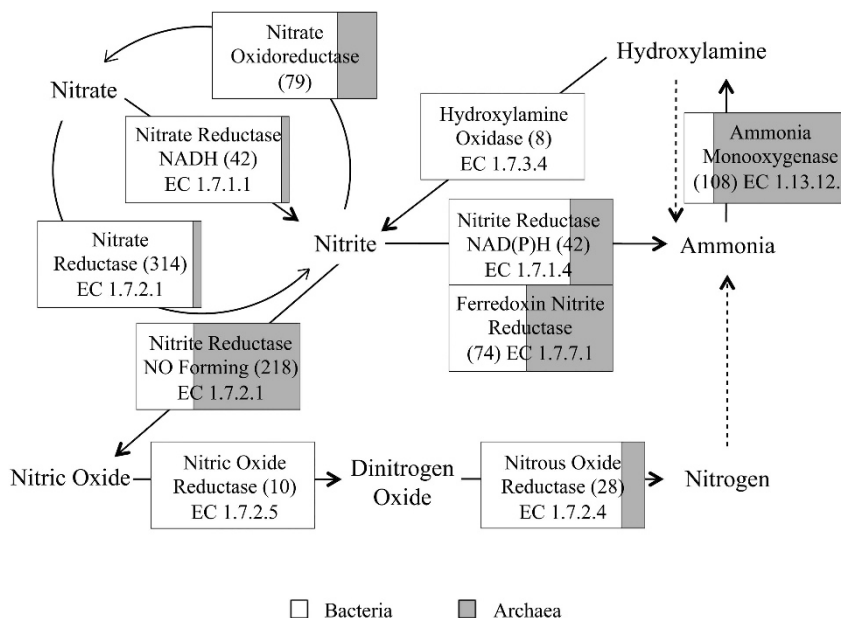
Figure 6 Bacterial and archaeal abundances for key nitrogen metabolism enzymes in the Weebubbie cave metagenome. The total abundance of gene fragments for each enzyme is shown in parentheses next to the enzyme name. The grey and white shading with the boxes indicates the relative proportions of bacterial and archaeal reads matching each enzyme. Solid arrows represent reactions performed by a known enzyme that was found in the Weebubbie cave metagenome. Dotted lines represent known nitrogen transformations for which no evidence was obtained from the Weebubbie cave metagenome.

Carbon fixation in *N. maritimus* SCM1 has been hypothesised to be carried out using the 3-hydroxy-propionate/4-hydrobutyrate pathway, with homologues of all the genes implicated in this pathway observed within this organism's genome (Walker *et al.,* 2010). Reads recruiting to all of these genes (biotin dependent acetyl-CoA/propionyl-CoA carboxylase, methylmalo-nyl-CoA epimerase and mutase and 4-hydroxybutyrate dehydratase) were observed in the Weebubbie meta-genome. Evidence for ribulose-1,5-bisphosphate carboxylase-oxygenase was also recovered from the bacterial binned component of metagenome reads.

Discussion

Weebubbie cave slime comprises a taxonomically diverse microbial community
The Nullarbor cave slime communities provide a fascinating ecosystem, in which microbes dominate a habitat which appears to be completely isolated from sunlight or photosynthetically derived carbon. Our examination of the phylogenetic composition of the Weebubbie microbial communities based on taxonomic binning of metagenome reads and 16S rRNA amplicon sequencing reveals a diverse community comprising both bacterial and a relatively high proportion of archaeal representatives spanning a range of described phyla, as well as a proportion of sequences which at this time are unable to be characterised. Both the Thaumarch-aeota and Nitrospirae have lower representation in the metagenomic binning compared with the 16S

rRNA amplicon phylogenetic analysis, this probably reflects the relative paucity of both taxa in genomic databases, leading to an under identification of sequences from these phyla. As the 16S rRNA gene can be present in different copy numbers in different genomes this potentially affects the estimation of the relative abundances of phylum. However, as the sequenced Thaumarchaeota generally have a single copy of the rRNA gene it seems unlikely that the 16S rRNA amplicon sequence data is overestimating the relative abundance of this phylum. Conversely, the proteobacteria and firmicutes have higher represen-tation in the metagenomic data compared with the 16S rRNA amplicon analysis, probably because these phyla are the two most heavily sampled by genome sequencing.

Discovery of a new, abundant Thaumarchaeote in the Webbubbie cave community
The most striking taxonomic observation from this work is the finding that Thaumarchaeota comprise a sizable proportion of assigned sequences, from 18–45% (Figure 2) based on metagenomic and 16S rRNA analyses, respectively. On the basis of the observation of an abundant filamentous organism in their s.e.m. examination of the microbial slimes whose morphology did not resemble that of the prevalent clone types, (Holmes *et al.,* 2001) hypothe-sised the potential for a 'missing link', an abundant organism not effectively amplified with their primer set. The observations made here confirm that an abundant phylum, namely Thaumarchaeota, was indeed missed in the cloning based approach, and

potentially comprise this "missing" filamentous morphotype. Examples of filamentous Thaumarchaeotes have been reported (Muller *et al.,* 2010) indicating the potential for the microbial slime filaments to similarly represent examples of novel thaumarchaeotal cell types.

A partial genome assembly of the Thaumarchaeota in the Weebubbie cave environment showed significant synteny, albeit with a number of large rearrangements, compared with the sequenced *N. maritimus* SCM1 genome. There are clearly clusters of genes within aligned Weebubbie contigs that are not observed in *N. maritimus* SCM1 indicating that cave environment likely contains novel sets of thaumarchaeotal genes. It is also likely that there are genomic regions in *N. maritimus* SCM1 which are not present in the cave environment, as indicated by a number of sizable *N. maritimus* SCM1 genome sections which have no matches at all in the metagenome reads. A recent analysis of metagenome reads most likely derived from environmental planktonic Thaumarchaeota in the Gulf of Maine, which when compared with *N. maritimus* also appeared to lack a number of genomic regions (Tully and Nelson, 2012). Combination of these three data sets putatively represents a core genome conserved amongst the group 1.1a Thaumarchaeota.

Phylogenetic analysis of a partial 16S rRNA gene sequence recovered from the Weebubbie metagenome provides further confirmation of the presence of at least one thaumarchaeotal representative in this environment. While two contigs containing partial thaumarchaeotal 16S rRNA genes were recovered, one was of insufficient length to carry out meaningful phylogenetic analysis and did not overlap with the other, making it difficult to say whether multiple lineages of Thaumarchaeota are present within the cave community. Analysis of the longer sequence indicates this organism grouped most closely with the marine sponge symbiont *Ca.* C. symbiosum (Preston *et al.,* 1996) and a presumably planktonic metagenome representative found in deep-sea hydrothermal vent water (Martin-Cuadrado *et al.,* 2008). It is of interest that the Weebubbie sequence groups together with representatives from marine environments, group 1.1a, rather than 1.1b, the group predominantly detected in terrestrial environments. Another recent metagenomic analysis of AOA in a radioactive thermal cave in the Austrian Alps found that five of six 16S rRNA gene fragments recovered clustered with the 1.1b group (Bartossek *et al.,* 2012). The Weebubbie thaumarchaeotal representative may indeed have had a marine origin, as periodic inundations of the Nullarbor caves by the sea are thought to have occurred a number of times in the geological past, during periods of high sea levels, however, it is thought that the final retreat of the sea occurred ca. 14 Ma after which the region experienced uplifting and has remained relatively geologically stable (Webb and James, 2006).

An unusual chemolithotrophic community utilising multiple nitrogen forms?

It has been hypothesised previously that the Nullarbor cave slime represents a chemolithotrophic community supported by nitrite oxidation, with water analysis indicating very high nitrite (5–10 p.p.m.) and relatively high nitrate levels (98 p.p.m.) around Weebubbie communities (Holmes *et al.,* 2001). Metagenomic sequencing has allowed us to look directly for genes relating to nitrogen cycling in this community. This analysis indicated that both stages of nitrification are likely to be key to energy generation in this community. Examination of the detected ammonia monooxygenase (*amo*) gene copies indicates that the thaumarchaeotal AOA dominate in the Weebubbie community, as they have been shown to do in the majority of soil, estuarine and hot spring sediments, coastal and marine environments for which relative abundances have been investigated (reviewed by Erguder *et al.* (2009)). Ratios of archaeal to bacterial *amoA* have been shown to be quite variable, determined to be around 10–100 in coastal waters, 10–1000 in the open ocean (Wuchter *et al.,* 2006) and as much as 80 times greater in estuarine environments (Caffrey *et al.,* 2007). In the Weebubbie metagenome reads containing archaeal *amo* gene fragments were roughly five times more abundant than bacterial copies, suggesting that while AOA are predominant, AOB activity may still contribute to the first stage of nitrification in this community.

The second stage of nitrification, oxidation of nitrite to nitrate also appears to rely on both archaeal and bacterial activity, however in this instance it appears the nitrite oxidising bacteria may be predominantly responsible. Our metagenome analyses found no evidence of nitrogen fixing pathways in this environment.

Conclusion

In this work metagenomic and 16S rRNA amplicon sequencing has provided a more complete view of the taxonomic composition of the Weebubbie microbial community, revealing that archaeal, as well as bacterial representatives live together within these unusual slime curtain structures. In 2001, Holmes and colleagues applied up-to-date molecular methods to examine the phylotypic structure of this community, providing the first glimpse into the ecology of this rare system (Holmes *et al.,* 2001). A little more than 10 years later microbiologists can make use of a range of new technologies, including next-generation sequencing of environmental DNA for metagenomics, to take an increasingly in-depth look at the composition, both taxonomic and metabolic, of such communities. Our analyses have significantly broadened our understanding of what organisms make up the Weebubbie cave slime community and what they may be doing to survive

in such conditions. On the basis of metabolic profiling we hypothesise that the Weebubbie 'slime curtains' represent novel chemolithotrophic communities driven by nitrification.

References

Bartossek R, Spang A, Weidler G, Lanzen A, Schleper C. (2012). Metagenomic analysis of ammonia-oxidizing archaea affiliated with the soil group. *Front Microbiol* **3**: 208.

Bates ST, Berg-Lyons D, Caporaso JG, Walters WA, Knight R, Fierer N. (2011). Examining the global distribution of dominant archaeal populations in soil. *ISME J* **5**: 908–917.

Blainey PC, Mosier AC, Potanina A, Francis CA, Quake SR. (2011). Genome of a low-salinity ammonia-oxidizing archaeon determined by single-cell and metagenomic analysis. *PLoS One* **6**: e16626.

Brady A, Salzberg SL. (2009). Phymm and PhymmBL: metagenomic phylogenetic classification with interpolated Markov models. *Nat Methods* **6**: 673–676.

Caffrey JM, Bano N, Kalanetra K, Hollibaugh JT. (2007). Ammonia oxidation and ammonia-oxidizing bacteria and archaea from estuaries with differing histories of hypoxia. *ISME J* **1**: 660–662.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK *et al.* (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**: 335–336.

Cole JR, Chai B, Farris RJ, Wang Q, Kulam SA, McGarrell DM *et al.* (2005). The Ribosomal Database Project (RDP-II): sequences and tools for high-throughput rRNA analysis. *Nucleic Acids Res* **33**: D294–D296.

Darling AE, Mau B, Perna NT. (2010). progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* **5**: e11147.

Darriba D, Taboada GL, Doallo R, Posada D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods* **9**: 772.

Erguder TH, Boon N, Wittebolle L, Marzorati M, Verstraete W. (2009). Environmental factors shaping the ecological niches of ammonia-oxidizing archaea. *FEMS Microbiol Rev* **33**: 855–869.

Gillieson DS, Spate AS. (1992). The nullarbor karst. In: Gillieson DS (ed). Geology, climate, hydrology and karst formation: IGCP Project 299, Field symposium. Department of Geography and Oceanography, Australian Defense Force Academy Special Publication edn: Canberra, pp 65–99.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Botany* **5**: 307–321.

Hallam SJ, Konstantinidis KT, Putnam N, Schleper C, Watanabe Y, Sugahara J *et al.* (2006). Genomic analysis of the uncultivated marine crenarchaeote Cenarchaeum symbiosum. *Proc Natl Acad Sci USA* **103**: 18296–18301.

Holmes AJ, Tujula NA, Holley M, Contos A, James JM, Rogers P *et al.* (2001). Phylogenetic structure of unusual aquatic microbial formations in Nullarbor caves, Australia. *Env Microbiol* **3**: 256–264.

James J, Rogers P. (1994). The 'mysterious' calcite precipitating organism of the Nullarbor caves, Australia. In: Sasowsky ID, Palmer M (eds). Breakthroughs in karst geomicrobiology and redox geochemistry. Karst Waters Institute Special Publication 1: Charlestown, pp 34–35.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA *et al.* (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**: 376–380.

Martin-Cuadrado AB, Rodriguez-Valera F, Moreira D, Alba JC, Ivars-Martinez E, Henn MR *et al.* (2008). Hindsight in the relative abundance, metabolic potential and genome dynamics of uncultivated marine archaea from comparative metagenomic analyses of bathypelagic plankton of different oceanic regions. *ISME J* **2**: 865–886.

Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M *et al.* (2008). The metagenomics RAST server—a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* **9**: 386.

Muller F, Brissac T, Le Bris N, Felbeck H, Gros O. (2010). First description of giant Archaea (Thaumarchaeota) associated with putative bacterial ectosymbionts in a sulfidic marine habitat. *Env Microbiol* **12**: 2371–2383.

Preston CM, Wu KY, Molinski TF, DeLong EF. (1996). A psychrophilic crenarchaeon inhabits a marine sponge: Cenarchaeum symbiosum gen. nov., sp. nov. *Proc Natl Acad Sci USA* **93**: 6241–6246.

Richards AM. (1971). An ecological study of the cavernicolous fauna of the Nullarbor Plain Southern Australia. *J Zool, Lond* **164**: 1–60.

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB *et al.* (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75**: 7537–7541.

Tully BJ, Nelson WC, Heidelberg JF. (2012). Metagenomic analysis of a complex marine planktonic thaumarchaeal community from the Gulf of Maine. *Environ Microbiol* **14**: 254–267.

Walker CB, de la Torre JR, Klotz MG, Urakawa H, Pinel N, Arp DJ *et al.* (2010). Nitrosopumilus maritimus genome reveals unique mechanisms for nitrification and autotrophy in globally distributed marine crenarchaea. *Proc Natl Acad Sci USA* **107**: 8818–8823.

Webb JA, James JM. (2006). Karst evolution of the Nullarbor Plain, Australia. In: Harmon RS, Wicks CM (eds). *Karst Geomorphology, Hydrology And Geochemistry*. Geological Society of America Special Paper 404: Boulder, pp 65–78.

Wuchter C, Abbas B, Coolen MJ, Herfort L, van Bleijswijk J, Timmers P *et al.* (2006). Archaeal nitrification in the ocean. *Proc Natl Acad Sci USA* **103**: 12317–12322.

Supplementary Information accompanies the paper on The ISME Journal website (http://www.nature.com/ismej)