www.nature.com/ismej

# ORIGINAL ARTICLE Archaea in Yellowstone Lake

Jinjun Kan<sup>1,6</sup>, Scott Clingenpeel<sup>2,6</sup>, Richard E Macur<sup>2</sup>, William P Inskeep<sup>2</sup>, Dave Lovalvo<sup>3</sup>, John Varley<sup>4</sup>, Yuri Gorby<sup>5</sup>, Timothy R McDermott<sup>2</sup> and Kenneth Nealson<sup>1,5</sup> <sup>1</sup>Department of Earth Sciences, University of Southern California, Los Angeles, CA, USA; <sup>2</sup>Thermal Biology Institute, Montana State University, Bozeman, MT, USA; <sup>3</sup>Eastern Oceanics, West Redding, CT, USA; <sup>4</sup>Big Sky Institute, Montana State University, Bozeman, MT, USA and <sup>5</sup>J Craig Venter Institute, San Diego, CA, USA

The Yellowstone geothermal complex has yielded foundational discoveries that have significantly enhanced our understanding of the Archaea. This study continues on this theme, examining Yellowstone Lake and its lake floor hydrothermal vents. Significant Archaea novelty and diversity were found associated with two near-surface photic zone environments and two vents that varied in their depth, temperature and geochemical profile. Phylogenetic diversity was assessed using 454-FLX sequencing (  $\sim$  51 000 pyrosequencing reads; V1 and V2 regions) and Sanger sequencing of 200 near-full-length polymerase chain reaction (PCR) clones. Automated classifiers (Ribosomal Database Project (RDP) and Greengenes) were problematic for the 454-FLX reads (wrong domain or phylum), although BLAST analysis of the 454-FLX reads against the phylogenetically placed fulllength Sanger sequenced PCR clones proved reliable. Most of the archaeal diversity was associated with vents, and as expected there were differences between the vents and the near-surface photic zone samples. Thaumarchaeota dominated all samples: vent-associated organisms corresponded to the largely uncharacterized Marine Group I, and in surface waters,  $\sim$  69–84% of the 454-FLX reads matched archaeal clones representing organisms that are Nitrosopumilus maritimus-like (96-97% identity). Importance of the lake nitrogen cycling was also suggested by >5% of the alkaline vent phylotypes being closely related to the nitrifier Candidatus Nitrosocaldus yellowstonii. The Euryarchaeota were primarily related to the uncharacterized environmental clones that make up the Deep Sea Euryarchaeal Group or Deep Sea Hydrothermal Vent Group-6. The phylogenetic parallels of Yellowstone Lake archaea to marine microorganisms provide opportunities to examine interesting evolutionary tracks between freshwater and marine lineages.

*The ISME Journal* (2011) **5**, 1784–1795; doi:10.1038/ismej.2011.56; published online 5 May 2011 **Subject Category:** microbial population and community ecology **Keywords:** archaea; Yellowstone Lake; pyrosequencing; N cycling; spatial distribution

## Introduction

Based primarily on the characterization of cultivated specimens, early work suggested that the *Archaea* were predominantly extremophiles, being found in high temperature, acidic, hypersaline or strictly anoxic habitats (Woese *et al.*, 1990; DeLong 1998). The 'extremophile stereotype' (Robertson *et al.*, 2005) now needs to be set aside as less biased molecular techniques have revealed the *Archaea* to be ubiquitous. A few examples of major ecosystems inhabited by *Archaea* include marine waters (DeLong, 1992; Fuhrman *et al.*, 1992) and sediments (Boetius *et al.*, 2000; Orphan *et al.*, 2002), soils (for example, Bintrim *et al.*, 1997; Jurgens *et al.*, 1997; Buckley

<sup>6</sup>These authors contributed equally to this work.

et al., 1998; Sandaa et al., 1999; Simon et al., 2000; Ochsenreiter et al., 2003) and freshwater bodies of various sizes (MacGregor et al., 1997, 2001; Schleper et al., 1997; Pernthaler et al., 1998; Glockner et al., 1999; Jurgens et al., 2000; Keough et al., 2003; Urbach et al., 2007; Lliros et al., 2008).

Still, extreme environments remain understudied with respect to Archaea diversity and discovery. The Yellowstone geothermal complex is an excellent example of an extreme environment known to be prime archaeal habitat. It is estimated to be comprised of an excess of 10 000 geysers, mud pots, springs and vents, as well as heated soils. The diverse combinations of chemistry and temperature provide numerous environments capable of supporting phylogenetically and functionally diverse microbial populations. Indeed, from a single pool in Yellowstone (Obsidian Pool), Barns et al. (1994) described phylogenetic diversity that essentially tripled the then known archaeal phylogeny. Since then, Yellowstone's geothermal features have continued to yield important discoveries that have documented novel Archaea and their viruses in both cultivation-based and -independent studies (Barns

Correspondence: TR McDermott, Thermal Biology Institute, Montana State University, Leon Johnson Hall, Bozeman, MT 59717, USA

or K Nealson, Department of Earth Sciences, University of Southern California, Los Angeles, CA 90089-0740, USA. E-mail: timmcder@montana.edu or KNealson@jcvi.org

Received 26 October 2010; revised 28 February 2011; accepted 24 March 2011; published online 5 May 2011

*et al.*, 1994, 1996; Meyer-Dombard *et al.*, 2005; Young *et al.*, 2005; de la Torre *et al.*, 2008; Elkins *et al.*, 2008).

We have extended work in Yellowstone to the study of the hydrothermal vents on the floor of Yellowstone Lake (Supplementary Figure S1). The northern half of the lake, spanning from the West Thumb region to Mary Bay (Supplementary Figure S1), accounts for  $\sim 10\%$  of the total geothermal flux in Yellowstone National Park (YNP) (Balistrieri et al., 2007). Recent surveys of Yellowstone Lake have employed bathymetric, seismic and submersible remotely operated vehicle (ROV) equipment to document hundreds of lake floor vent features (Morgan et al., 2003, 2007; Balistrieri et al., 2007). Efforts summarized herein focus on the characterization of the archaeal populations associated with two of these vents as well as in non-thermal, nearsurface waters. Geochemical profiling is included to provide environmental context to the microbial diversity data.

## Materials and methods

## Sampling locations and geochemical analysis

Vents and vent fields were located from global information system coordinates established from past USGS surveys (for example, Morgan *et al.*, 1977, 2007). Hydrothermal vents chosen for this study are located in the Inflated Plain and West Thumb regions of Yellowstone Lake, YNP. The coordinates of the vents are as follows: Inflated Plain Vent-329 (Inflated Plain region, 44° 32.111' N, 110° 21.240' W), Otter Vent-332 (West Thumb region, 44° 26.604' N, 110° 33.970' W), West Thumb Deep Vent-339 (West Thumb region, 44° 25.177' N, 110° 31.407' W) and West Thumb Vent-342 (44° 25.177' N, W110° 31.407' W). The relative locations for each site in the lake are shown in Supplementary Figure S1.

#### Water sampling and analysis

The ROV was described previously (Lovalvo *et al.*, 2010) and microbial and geochemical sampling methods were as described by Clingenpeel *et al.* (2011). Briefly for microbial samples, 100–3001 of lake/vent water was pumped through a 20  $\mu$ m prefilter into 50 l carboys on the boat deck. Before use, the carboys were sterilized by either autoclaving or by soaking with 10% bleach, followed by rinsing with autoclaved deionized water. Using techniques previously described for the global ocean survey (Rusch *et al.*, 2007), lake and vent water was size fractionated by serial filtration through 3.0, 0.8 and 0.1  $\mu$ m membrane filters. Filters were sealed in plastic bags and frozen at -20 °C for transport to the laboratory, where they were stored at -80 °C.

#### DNA extraction

Filters were aseptically cut into quarters with sterile scalpels, with one quarter placed in a sterile 50 ml tube for DNA extraction. Unused filter samples were

refrozen and stored at -80 °C. The filter sample was cut into small pieces and 15 ml of buffer (Tris-HCl, 0.1 M; EDTA, 0.1 M; sucrose, 0.8 M; pH 8) was added. Lysozyme was added to a final concentration of 1 mg ml<sup>-1</sup> and the solution was incubated at 37 °C for 30 min. Proteinase K was added to a final concentration of  $0.1 \,\mathrm{mg}\,\mathrm{ml}^{-1}$  and sodium dodecyl sulfate was added to a final concentration of 1% (w  $v^{-1}$ ). This mixture was incubated at 37 °C for 4 h. Polysaccharides and residual proteins were aggregated by the addition of hexadecyltrimethyl ammonium bromide to a final concentration of  $1\% (w v^{-1})$ and sodium chloride at a final concentration of 0.14 M, and the mixture then incubated at 65  $^\circ \mathrm{C}$  for 30 min. DNA was purified by two rounds of extraction with phenol-chloroform-isoamyl alcohol (25:24:1), followed by two rounds of chloroformisoamyl alcohol (24:1). DNA was precipitated by the

addition of an equal volume of isopropanol with incubation at -20 °C for 2 h, followed by centrifugation (13 000 g, 15 min). The DNA pellet was washed twice with 70% ethanol and resuspended in TE buffer (Tris-HCl, 10 mM; EDTA, 1 mM; pH 8).

#### Pyrosequencing

The V1 and V2 regions of the archaeal 16S rRNA encoding gene were targeted using two forward primers A2Fa: 5'-TTCCGGTTGATCCYGCCGGA-3' and N3F: 5'-TCCCGTTGATCCTGCG-3' coupled with one reverse primer A571R: 5'-GCTACGGÑYSCTTT ARGC-3' (Baker et al., 2003) with sequencing occurring from the 3' end of the amplicons. The polymerase chain reaction (PCR) mix was 50 µl, containing 1.5 mM MgCl<sub>2</sub>, 20 µg bovine serum albumin, 0.2 mM each dNTP,  $1 \mu$ M each primer and 1.25 U Taq polymerase. The PCR program was 94 °C for 5 min, 25 cycles of 94 °C for 1 min, 55 °C for 1 min, and 72  $^\circ \! C$  for 1 min, 72  $^\circ \! C$  for 7 min and 4 °C hold. After 25 cycles of amplification, five more cycles were used to add the sample-specific barcodes and the adaptor sequences required for 454-FLX pyrosequencing. The barcode sequences used were selected from the list provided by Hamady et al. (2008). The barcoded 16S rRNA gene PCR amplicons obtained from the different environments were pooled according to their relative amplicon abundance (determined under standardized PCR conditions) so that the different environments were proportionally represented in the pooled amplicon that was then pyrosequenced using 454-GS FLX sequencer version of 454 Life Sciences (Branford, CT, USA) (Margulies et al., 2005) at the JC Venter Institute sequencing center.

#### PCR clone libraries

Amplification of the 16S rRNA gene for the production of clone libraries was performed using two pairs of primers, 8aF and 1513R (Eder *et al.*, 1999), and N3F and N1406R (Huber *et al.*, 2002). The PCR mix

1786

was 50 µl containing 1.5 mM MgCl<sub>2</sub>, 20 µg bovine serum albumin, 0.2 mM each dNTP, 1 µM each primer and 1.25 U Taq polymerase. The PCR program was 94 °C for 5 min, 30 cycles of 94 °C for 1 min, 54 °C for 1 min, and 72 °C for 1 min, 72 °C for 7 min and 4 °C hold. PCR products were cloned using the TOPO TA Cloning Kit (Invitrogen Corp., Carlsbad, CA, USA), following the manufacturer's instructions. Individual clones were cultured and their plasmids purified using the QIAprep Spin Miniprep Kit (Qiagen, Valencia, CA, USA). Inserts in each were sequenced from two ends using Big Dye Terminator chemistry (Applied Biosystems, Foster City, CA, USA) and the Applied Biosystems 3100 Genetic Analyzer. Chimeric sequences were screened by the 'CHIMERA DETECTION' program of the Ribosomal Database Project (RDP) (Maidak et al., 1997) and removed from further analysis.

#### Data analysis

Near-full-length clone libraries were aligned, trimmed and then initially classified using BLAST (Altschul *et al.*, 1990). Neighbor-joining distance trees were constructed using the MacVector 10.0 software package (GCG) and bootstrapping was based on 1000 resampling data sets. The grouping of full-length clone sequences was accomplished using the ARB software (Ludwig et al., 2004) and the latest released Silva 102 database (Pruesse et al., 2007). For pyrosequencing libraries, the pyrosequencing reads were quality trimmed according to Kunin *et al.* (2010), followed by clustering using an abundance-sorted preclustering as per Huse et al. (2010) and a final complete linkage (furthest neighbor) clustering using the mothur software (Schloss et al., 2009). Statistical analysis of operation of taxonomical units (OTU) richness via rarefaction, Chao1 and ACE (abundance-based coverage) estimates were performed in mothur, with the pyrosequencing data sets all normalized to the same number of reads. Statistical analysis of OTU richness via rarefaction, Chao1 and ACE estimates were performed as described in Roesch et al. (2007). Pyrosequencing reads were classified using three methods: (i) the Ribosomal Database Project (RDP) classifier (Wang et al., 2007; Cole et al., 2009); (ii)

Greengenes (DeSantis *et al.*, 2006); and (iii) by comparing the 454-FLX reads to the near-full-length clone sequences using BLAST (Altschul *et al.*, 1990). For the RDP classifier, default classifier conditions were applied, and for Greengenes, the sequences were batch aligned using the Compare Align tool, setting the match requirements at a minimum length of 50 bases and minimum identity set at 75%. For BLAST matching the 454-FLX reads to the full-length clones, match criteria required  $\geq$ 97% identity for  $\geq$ 95% of the read length. DNA pyrosequences are available in the following GenBank SRA Accession No. SRA027147.1.

## Results

#### Geochemical analysis

Sampling focused on regions of the lake where significant hydrothermal vent fields are known to occur (Morgan et al., 2003, 2007; Supplementary Figure S1). And as a control, samples were also taken from the Southeast Arm (Supplementary Figure S1), which is located well outside the caldera boundary, is most proximal to the major lake input (Yellowstone River) and where extensive USGS surveys suggest that geothermal vents are absent (Morgan et al., 1977, 2007). We recently reported a comprehensive geochemical analysis description of the four sites considered in this report as well as several other lake sites (Clingenpeel et al., 2011). These geochemical descriptions for the four sites are also provided here in Table 1 and Supplementary Table S1. In general, chemistry of the different vent waters (Table 1) was within the range of previous reports (Aguilar et al., 2002; Cuhel et al., 2002; Remsen et al., 2002). The two vents highlighted here were similar in temperature, but differed in pH and depth in relation to the lake surface. The vents also differed with respect to various important gases such as  $CO_2$ ,  $O_2$ and CH<sub>4</sub> (Table 1), but contained similar concentrations of H<sub>2</sub>. The Inflated Plain surface water samples were obtained directly within a significant gas plume that was visually evident by profuse bubbles rising to the surface.  $H_2$ ,  $CH_4$  and  $CO_2$  levels in this sample were significantly greater than in surface waters at the Southeast Arm location (Table 1).

Table 1 Summary of prominent geochemical characteristics of the Yellowstone Lake environments examined for archaeal diversity

| Sampling site         | Select physicochemical measurements |            |               |                  |                              |                 |                              |                             |             |  |  |
|-----------------------|-------------------------------------|------------|---------------|------------------|------------------------------|-----------------|------------------------------|-----------------------------|-------------|--|--|
|                       | Depth<br>(m)                        | рН<br>(°С) | Temp.<br>(mM) | СО2<br>(aq) (тм) | S <sup>2–</sup><br>(aq) (mM) | O2<br>(aq) (mM) | CH <sub>4</sub><br>(aq) (nM) | H <sub>2</sub><br>(aq) (mM) | NH₄<br>(μM) |  |  |
| Inflated Plain Photic | 2.7                                 | 6.9        | 10.9          | 0.02             | bdª                          | 234             | 0.2                          | 114                         | 2.8         |  |  |
| Southeast Arm Photic  | 2.5                                 | 7.0        | 10.6          | 0.01             | bd                           | 234             | 0.1                          | 10                          | 2.8         |  |  |
| West Thumb Otter Vent | 4.3                                 | 8.4        | 63-68         | 0.01             | bd                           | 26              | 0.1                          | 43                          | 2.8         |  |  |
| West Thumb Deep Vent  | 52.0                                | 6.2        | 60–66         | 1.65             | 2.1                          | 113             | 6.4                          | 49                          | 12.1        |  |  |

Data shown is a subset of a more complete description in Clingenpeel et al. (2011).

<sup>a</sup>bd = below detection, detection limits: 0.3 μM S<sup>2-</sup>, 2.7 μM S<sub>2</sub>O<sub>3</sub><sup>2-</sup>, 4.2 μM PO<sub>4</sub><sup>3-</sup>, 0.3 μM Al, 0.1 μM Mn, 0.1 μM Fe, 0.1 μM Se, 0.03 μM Mo, 0.01 μM W and 0.02 μM Pb.

Diversity analysis in this study focused on only the >0.1-0.8 < um size class. For a single 1/2 plate 454-FLX pyrosequencing run, a total of 333272 reads were generated (including amplicons generated with *Bacteria*-specific primers; see Clingenpeel et al., 2011). Of these, 51017 were generated the archaeal primers (Supplementary with Table S2). The lower proportion of the total archaeal 454-FLX reads reflects their approximate contribution to total lake microbial community composition as judged by archaeal amplicon strength from all lake samples and domain-specific primer sets; that is, all PCRs were optimized, however the archaeal amplicon strength was considerably less than that obtained for the *Bacteria* primer sets. Hence, it was concluded that the archaeal amplicons reflected minority abundance relative to Bacteria, and consequently comprised a commensurate minority proportion of the total amplicon pool submitted for pyrosequencing.

pyrosequencing statistics reflected the The expected proportional read number composition and suggested that for all but one amplicon pool component the sequencing capacity of the 454-FLX technology employed was attained (average read length = 232 bp; Supplementary Table S2). Following quality trimming protocols suggested by Kunin et al. (2010), 37 361 reads were then prepared for analysis. Collector's curve analysis (Figure 1), and Chao1 and ACE richness estimates (Table 2) suggested that there was considerable archaeal diversity associated with each of these lake environments. The greatest diversity was found associated with the heavier sampled vent waters, followed by significantly lower estimates for the near-surface photic zone waters (Figure 1 and Table 2). In examining the unique distribution of the 454-FLX read OTUs (set at 97% identity), 28% were only found in the West Thumb Deep Vent and 41% were unique to the alkaline pH Otter Vent (Figure 2). The majority of archaeal phylotypes were found in the extreme environments of the vent waters, with OTUs present in one or both vents making up 76% of the total, whereas only 1.4% were found in all four lake environments (Figure 2).

For taxonomic classification of the pyrosequencing reads, the automated classifiers at the RDP and at Greengenes proved problematic. Neither classifier has the category of Thaumarchaeota and so for this comparison the Thaumarchaeal sequences are considered to be Crenarchaea. To illustrate the problems encountered, Table 3 compares classification results obtained for the following: (i) the entire 454-FLX data set for the Inflated Plain (10 580 reads) assessed using the RDP classifier; (ii) a random 500 read sample from the Inflated Plain data set examined using the more time-consuming Greengenes 'Compare' tool that first aligns and then classifies using RDP, Hugenholtz and NCBI taxonomies; and (iii) BLAST comparison of the 454-FLX reads against



Figure 1 Collector's curves estimating numbers of *Archaea* OTUs identified for all samples, and as a function of sequence identity set at 97% and quality screened as described by Kunin *et al.* (2010).

 Table 2
 Richness estimates for total archaeal 16S rRNA gene sequences derived from Yellowstone Lake

| Sample   | Richness estimator                                       |  |   |  |  |  |  |
|--|--|--|---|--|--|--|--|
|  | ACE  | Chao1  | Rarefaction                                       |  |  |  |  |
| West Thumb Deep Vent<br>West Thumb Otter Vent<br>Inflated Plain Surface<br>Southeast Arm Surface | $255 \pm 37^{a} \\ 245 \pm 27 \\ 74 \pm 4 \\ 158 \pm 17$ | $197 \pm 30 \\ 224 \pm 23 \\ 74 \pm 5 \\ 164 \pm 23$ | $233 \pm 2 \\ 273 \pm 3 \\ 83 \pm 1 \\ 166 \pm 2$ |  |  |  |  |

Abbreviations: ACE, abundance-based coverage; OUT, operational taxonomic unit.

Estimates were derived from OTUs assigned based on 97% sequence identity.

<sup>a</sup>Mean±s.d



**Figure 2** Proportional occurrence of 454-FLX *Archaea* OTUs unique to each of the designated lake locations. OTUs were assigned based on 97% identity match. DV, West Thumb Deep Vent; OV, Otter Vent; IP, Inflated Plain Photic Zone; SEA, Southeast Arm Photic Zone.

| Table 3         Classification (%) | ) of t | he archaea | l 454-FLX | reads | using | different | methods | of | classifica | ation |
|------------------------------------|--------|------------|-----------|-------|-------|-----------|---------|----|------------|-------|
|------------------------------------|--------|------------|-----------|-------|-------|-----------|---------|----|------------|-------|

| Taxon  | Classification method                          |   |                   |  |                         |                          |                                      |                                  |                              |  |
|--|--|---|-------------------|--|-------------------------|--------------------------|--------------------------------------|----------------------------------|------------------------------|--|
|  |  |   | Gre               | Greengenes classifier<br>(taxonomy)        |                         |                          |                                      |                                  |                              |  |
|  | RDP classifier                                 | ssifier BLAST comparison to near-full-length clones |                   |  |                         |                          |                                      |                                  | NCBI                         |  |
|  | Complete<br>Inflated Plain<br>454-FLX read set | 500 Read<br>Inflated Plain<br>subsample             | Inflated<br>Plain | Southeast<br>Arm                           | West Thumb<br>Deep Vent | West Thumb<br>Otter Vent | 500 Read Inflated<br>Plain subsample |                                  |                              |  |
| Unclassified<br>Bacteria<br>Archaea  | 12.1<br>14.2<br>73.7                           | 16.2ª   | 15.2ª             | 27.3ª                                      | 10.9 <sup>a</sup>       | 28.0 <sup>a</sup>        | $9.8^{ m b}$<br>88.5<br>11.5         | 9.8 <sup>b</sup><br>87.0<br>13.0 | $9.8^{ m b}$<br>15.9<br>84.1 |  |
| <i>Unclassified</i> Archaea <sup>c</sup><br>Crenarchaeota <sup>c</sup><br>Euryarchaeota <sup>c</sup> | $45.0 \\ 1.5 \\ 52.6$                          | 82.8<br>1.0   | 84<br>0.8         | $\begin{array}{c} 68.7 \\ 4.0 \end{array}$ | 87.6<br>1.6             | 37.7<br>34.3             | 0.0<br>100.0                         | 0.0<br>100.0                     | 1 read<br>>99.9              |  |

Abbreviations: NCBI, National Center for Biotechnology Information; RDP, Ribosomal Database Project.

For the RDP classifier, the entire Inflated Plain data set (10 580 reads after quality trimming) was employed. For the Greengenes classifier and BLAST comparison with full-length PCR clones, a 500 read subsample from the Inflated Plain data set was used to show contrasting results. <sup>a</sup>Failed to match any of the 200 near-full-length clones based on the criterion of 97% match over 95% length of the 454-FLX read. <sup>b</sup>Percent of reads that failed the initial alignment criteria of >75% identity over a minimum of 50 bases.

°% of all reads classified as Archaea.

the Sanger-sequenced near-full-length PCR clones (discussed more fully below). In the complete Inflated Plain data set, the RDP classifier could not classify approximately 12% of the reads and determined that roughly 14% belonged to the domain Bacteria (Table 3). Of the 73.7% of the reads classified in the domain Archaea, nearly half could not be placed at the phylum level, with roughly 53% incorrectly classified as Euryarchaeota being (Table 3). These results are in variable contrast to that obtained from Greengenes. Close to 10% did not pass the low stringency alignment constraint imposed (>75% identity across at least 50 bases). Of those that were aligned, the RDP and Hugenholtz taxonomies within Greengenes both incorrectly classified the reads as approximately 88% Bacteria, with the balance further incorrectly classified as Eurvarchaeota (Table 3). By contrast, the Greengenes NCBI taxonomy determined that  $\sim 84\%$  of the aligned reads belonged to the domain Archaea, with all but one read being Euryarchaeota (Table 3). These results were in sharp disagreement with a separate analysis, where the same 500 read Inflated Plain subsample was compared to the full-length PCR clones that had been determined to be all Archaea (via manual BLAST analysis against the GenBank database and phylogenetic analysis) and to represent the Euryarchaeota and the Crenarchaeota (see below). In this analysis, 16.2% of the reads did not match any of the full-length clones using more stringent criterion ( $\geq 97\%$  identity for  $\geq 95\%$  of the pyrosequencing read length), with the majority of the classifiable reads being determined to belong to the Crenarchaeota (Table 3).

To further examine these discrepancies, individual 454-FLX reads that had been incorrectly classified by Greengenes were individually BLAST searched against the Greengenes database. All such searches yielded correct classifications and in agreement with the results obtained when the reads had been compared to the full-length clones. That is, Greengene BLAST results were internally inconsistent with the automated classifier. BLAST searches of the same individual read samples at GenBank again generated results consistent with that obtained via full-length comparisons.

#### Sanger sequencing of near-full-length PCR clones

A total of 384 near-full-length amplicons were screened with a single sequencing read for confirmation of valid archaeal sequence and for selectclones for full-length sequencing ing and phylogenetic analysis (200 clones total). These PCR clones allowed for longer taxonomic strings and were important for examining archaeal diversity at finer taxonomic resolution than allowed by the 454-FLX sequences. The majority of clones retrieved from the lake were vent associated and related to Marine Group 1 (recently proposed as phylum Thaumarchaeota), Candidatus Nitrosocaldus, and unclassified and miscellaneous crenarchaeal groups (Figure 3 and Supplementary Figures S2-S5). Owing to the relative lack of cultured and characterized Archaea, putative functional information was infrequent, although there were a few important exceptions. Clone group YLCG-1.1 was 96-97% identical to the cultured organism *Nitrosopumilus maritimus*, a Thaumarchaeal marine nitrifier (Könneke et al., 2005). The organism(s) represented by this phylotype dominated the Inflated Plain and Southeast Arm photic zones, accounting for 83.8 and 68.7% of



**Figure 3** Phylogenetic associations of the near-full-length PCR clones of Yellowstone Lake Crenarchaeal Groups (YLCGs). Approximate % representation of 454-FLX sequences for West Thumb Deep Vent, Otter Vent, Inflated Plain Photic and Southeast Arm Photic are shown within parentheses. Clones lacking parenthetical data were not found in the pyrosequencing data set. Clones highlighted in gray boxes represent phylotypes that represented at least 10% of the pyrosequencing reads from at least one location.

the 454-FLX reads in these samples, respectively (Figure 3 and Supplementary Figure S2). Another interesting near match to a cultured archaeaon and that again functionally linked with nitrogen cycling were clone groups YLCG-2.1 and YLCG-2.2, which were closely related (96–99% identical) to the Crenarchaeal *Candidatus* Nitrosocaldus yellowstonii (Figure 3 and Supplementary Figure S5). Clone group YLCG-2.1 was found exclusively in the Otter Vent waters, whereas YLCG-2.2 was primarily associated with Otter Vent, but also found in fluids obtained from the West Thumb Deep Vent (Figure 3 1789



**Figure 4** Phylogenetic associations of the near-full-length PCR clones of Yellowstone Lake Euryarchaeal Groups (YLEGs). Approximate % representation of 454-FLX sequences represented for West Thumb Deep Vent, Otter Vent, Inflated Plain Photic and Southeast Arm Photic are shown within parentheses. Clones lacking parenthetical data were not found in the pyrosequencing data set. Clones highlighted in gray boxes represent phylotypes that represented at least 10% of the pyrosequencing reads from at least one location.

and Supplementary Figure S5). Other Crenarcheaotes were also vent associated and are annotated as 'Unclassified' or 'Miscellaneous' (Figure 3).

The Euryarchaeote signatures were most closely related to other environmental clones and were most prevalent in vent fluids (Figure 4 and Supplementary Figures S6 and S7), particularly in Otter Vent (Table 3). For example, clone YLA099, YLEG-1.4b and YLEG-2 were abundant in Otter Vent, but essentially absent, or nearly so, in all other locations (Figure 4 and Supplementary Figure S5 and S6). This implies that the organism(s) represented by phylotype YLA099 is a thermophile even though the closest phylogenetic neighbor was an environmental

1790

clone from a hypersaline mat (Figure 4). The lone functionality that potentially may be inferred for the Yellowstone Lake Euryarchaeotes might be the association between Otter Vent clones YLA020, YLA015 and YLA025 with *Methanoregula boonei*, an acidophilic methanogen (Bräuer *et al.*, 2011), and which also shared a clade structure with other methanogens (Figure 4). We note, however, that these specific methanogen sequences were not found in the 454-FLX pyrosequencing library. This discrepancy is likely due to different biases in the primer sets used for the pyrosequencing and the clone libraries.

## Discussion

Contemporary international research efforts have illustrated the ubiquitous nature of the Archaea, and thus their presence in the hydrothermally active Yellowstone Lake was anticipated. The two vents were selected for microbial community analysis based on their different aqueous chemistry profiles, and two near-surface photic zones were selected because they differed with respect to their relative proximity to lake floor vent activity and to the lake's largest tributary, the Yellowstone River (Table 1 and Supplementary Table S1). Vent chemistry differed significantly in pH and in potentially important microbial nutrients such as CO<sub>2</sub>, CH<sub>4</sub> and NH<sub>4</sub> (Table 1 and Supplementary Table S1), all of which would be expected to exert selective effects on the associated microbial physiology types. Gas compositions of vent waters studied in the current project are consistent with previous reports (Remsen et al., 1990; Cuhel et al., 2002; Remsen et al., 2002; Spear et al., 2005), and significantly greater than that reported for Yellowstone's terrestrial hot springs (for example, Langner et al., 2001; Macur et al., 2004; Spear et al., 2005; D'Imperio et al., 2008). The  $H_2$ measurements (Table 1) represent a new addition to our understanding of the chemistry in this lake (Table 1 and Supplementary Table S1), which is important as H<sub>2</sub> is an important energy source for microbial metabolisms in Yellowstone high-temperature ecosystems (Spear et al., 2005). A more comprehensive assessment and discussion of the lake chemistry can be found in Clingenpeel et al. (2011).

As judged by the strength of normalized and optimized PCRs, the archaeal amplicons for each lake sample were noticeably weaker than those generated with bacterial primers, and thus while not quantitative, the *Archaea* were concluded to represent a smaller component of the microbial communities in the various lake environments. Consequently, the archaeal amplicons were combined as a minority component (~15%) of an amplicon pool submitted for pyrosequencing. The minority estimate for *Archaea* in these lake environments is consistent with reports of other lakes (although with no thermal inputs) in which a variety

of more quantitative techniques were used (Pernthaler et al., 1998; Glockner et al., 1999; Keough et al., 2003; Urbach et al., 2007). And while PCR biases (Suzuki and Giovannoni, 1996; von Wintzingerode et al., 1997) presumably occurred to some extent, we assumed that amplicon strength provides a reasonable comparison of relative occurrence and abundance of specific phylotypes in the four lake environments sampled. Based on the shape of the collector's curves for each site (Figure 1), coverage appeared somewhat similar for the four environmental samples, although sampling intensity was much greater for the vent samples. This suggests that the estimates of archaeal proportional abundance in the microbial communities associated with different lake environments were reasonable (although still only approximate) and that archaeal diversity was greater in the vent waters. Future efforts will focus on quantitative assessments of targeted taxa to more precisely estimate their abundance, population dynamics and functional roles.

Quality trimming (as per Kunin et al. (2010)) of the pyrosequences resulted in a 26.8% cull rate. Among the 37 361 remaining reads, classification was complicated by lack of agreement among the automated classifiers. The difficulty experienced in this particular study may stem from the smaller Archaea training set (compared with that for Bacteria) available for classifier use. More specifically, the RDP classifier taxonomy base does not contain representative sequences for the Crenarchaeota and Thaumarchaeota clades recovered in this study and hence the difficulty in their classification (James Cole, Ribosome Database Project, personal communication). We were unable to ascertain the basis for the severe and inconsistent classifier problems encountered with the Greengenes classifier.

Although BLAST match comparisons of the 454-FLX reads against the near-full-length clones proved to be the most reliable way of classifying the pyrosequencing data, not all full-length clones could be accounted for in the 454-FLX data sets. Examples include YLA044, YLA030 and YLA026 among several Crenarchaeota phylotypes (Figure 3), and YLA087, YLA073 and YLA104 among a number of Eurvarchaeota full-length clones (Figure 4). This implies that additional pyrosequencing would likely reveal additional diversity, which is not surprising. Where perfect matches were not found, there were closely related near-full-length clones that could be matched with the pyrosequencing data, which allowed for general assessments of the represented organism's distribution within the lake. We note that not all 454-FLX reads could be matched with the full-length clones (Table 3), resulting in 10.9–28.0% of the pyrosequencing reads remaining unclassified using the matching criteria imposed (Table 3).

Combining clone distribution patterns with the geochemical data (Table 1 and Supplementary Table S1) allows for speculation about ecological 1792

context. In spite of the near daily strong winds that generate surface currents contributing to lake mixing (Benson, 1961), roughly 78% of the OTUs (97%) identity) were only found associated with vent emissions (Figure 2). Exclusive vent associations that could be linked to characterized thermophiles involved the presence of organisms represented by the Crenarchaeota group YLCG-2.2 that are closely related to the thermophilic nitrifier Candidatus Nitrosocaldus vellowstonii (Figure 3 and Supplementary Figure S5) found exclusively in the Otter Vent. Distribution patterns and close phylogenetic relatedness to another known nitrifier also putatively identify apparent novel thermophiles. Within the YLCG-1.1 group (Figure 3 and Supplementary Figure S3), the consistent distribution pattern of the N. maritimus-like reads suggests that there are YLCG-1.1 organisms (in particular YLA046; Supplementary Figure S2) associated with the West Thumb Deep Vent as well as the near-surface photic zone samples taken in the Inflated Plain and Southeast Arm (Figure 3 and Supplementary Figure S2). This implies that phylogenetically closely related archaeal organisms have adapted to very different environments, and also reflects the broad environmental distribution of the Marine Group 1 Thaumarchaeota that includes mines, freshwater, saltwater, drinking water plants, soils and sponge symbionts (Figure 3), an observation noted previously (Nicol and Schleper, 2006; Schleper, 2010). Inferring dual thermophilic and mesophilic habitats for the YLCG-1.1 organisms is strengthened by contrasting distribution patterns, for example, YLCG-1.2d, another dominant 454-FLX phylotype (Figure 3 and Supplementary Figure S4), was abundant in the West Thumb Deep Vent, yet was undetected in the near-surface waters. Additional comparative contrasts add further momentum to such inferences; the near equal distribution of minor signatures such as YLA067, YLA097 and YLA098 (Figure 4) throughout the lake implies that more rare phylotypes could be detected at the level of pyrosequencing used and thus lack of detection of YLCG-1.2d is not necessarily due to shallow sampling. When viewed in sum and in combination, 454-FLX read distribution patterns likely represent real spatial distribution of the Archaea within this lake.

The above notwithstanding, we also note that low abundance of some predominantly photic zone clones in vent emissions could nevertheless have resulted from low-level proportional mixing of lake water with vent water. Rocks surrounding the various vents made it difficult (or impossible in some cases) for the ROV sampling cup to form a tight seal around the vent orifice (Clingenpeel *et al.*, 2011). This provides opportunity for surrounding lake water to mix to some degree with vent water during sampling. Geochemical evidence of this comes from the oxygen content in the vent water samples (Table 1), which would otherwise be expected to be anaerobic.

Crenarchaeota (Thaumarchaeota) dominated the 454-FLX data, in line with other reports of archaeal diversity for several of the great lakes around the world (Keough et al., 2003). However, we note with interest the complete absence of Thermoprotei, a common crenarchaeal class found in thermophilic environments, including Yellowstone's hot springs (Barns et al., 1994; Mever-Dombard et al., 2005). One of the primer sets used was the same as in Meyer-Dombard et al. (2005), and thus a lack of detection due to primer bias cannot explain the absence of the Thermoprotei in the data. The majority of the Crenarchaeota 454-FLX reads grouped with Marine Group 1 phylotypes (Figure 3), which was recently proposed as a newly designated phylum, and Thaumarchaeota, which contains the orders Nirosopumilales and Cenarchaeales (Brochier-Armanet et al., 2008). Molecular surveys have revealed several lineages that are related to this mesophilic archaeal phylum, such as SAGMCG-1, FFS, marine benthic groups B and C, YNPFFA and THSC1 (Schleper et al., 2005), suggesting the potential expansion of this new phylum. However, given the lack of genomic information and poor resolution of archaeal phylogeny/taxonomy, in this report we cautiously only included Marine Group 1 in the Thaumarchaeota clade. Within this clade, phylotype groups YLCG-1.2b, YLCG-1.2c and YLCG-1.2d were most closely related to a 16S rRNA gene sequence observed in a marine metagenome contig (Rusch et al., 2007). They were absent in the photic zone waters of Inflated Plain and the Southeast Arm, but comprised 27 to  $\sim 45\%$  of the reads in the West Thumb Deep Vent emissions (Figure 3 and Supplementary Figures S3 and S4), suggesting that the represented organisms are thermophiles. Also, group YLCG-1.2a represented approximately a third of the clones in the Otter Vent waters, again implying thermophily.

Although physiological inference was limited for most of the full-length and 454-FLX sequences, there were two group designations for which specific lake function might be inferred. The relatedness of phylogroup YLCG-1.1 to the marine nitrifier N. maritimus (Könneke et al., 2005) is a potentially significant observation in this regard. The represented organism(s) appeared to be an important lake archaeal picoplankton component, comprising  $\sim 69$  and 84% of the archaeal pyrosequencing reads in the Southeast Arm and Inflated Plain surface photic zone waters, respectively (Figure 3). The association of this organism(s) with the surface waters is consistent with a recent study that examined high mountain lakes in the Pyrenees, where archaeal nitrifier signatures (N. maritimus accC gene) were found in the lake neuston environment (Auguet et al., 2008).

Another strong nitrifier signature was evident in Yellowstone Lake crenarachaeota groups YLCG-2.1 and YLCG-2.2, found almost exclusively in the Otter Vent waters (Figure 3 and Supplementary Figure S5), sharing 96–99% identity with *Cadidatus* Nitrosocaldus vellowstonii. first isolated from alkaline (pH 8.3) sediments from hot springs in the Heart Lake region in YNP (de la Torre et al., 2008). Among the lake sites studied, the pH of the Otter Vent (pH 8.4) most closely matched that of the Heart Lake location, even though nitrifier-relevant concentrations of carbon  $(CO_2)$  and energy source  $(NH_3)$ were greater in the West Thumb Deep Vent (Table 1 and Supplementary Table S1).

Although the Euryarchaeota were a minor component of the pyrosequencing libraries, there were two phylotypes that appeared particularly abundant in specific environments. Clone groups YLA099 and YLEG-1.4b comprised  $\sim 8-12\%$  of the 454-FLX clones derived from the West Thumb Deep Vent (Figure 4 and Supplementary Figure S6). Their association with the vent waters is likely not coincidental as they grouped with the marine Deep Sea Hydrothermal Euryarchaeal Group 6 (Figure 4). Euryarchaeota phylotype group YLEG-2 was also almost exclusively found in the West Thumb Deep Vent emissions (Figure 4 and Supplementary Figure S7).

Finally, we comment on the striking phylogenetic similarity of Yellowstone Lake's microbial community to that described for marine environments and draw attention to a recent review by Logares *et al.* (2009). A majority of Thaumarchaeota appear related to the marine nitrifier *N. maritimus* (Figure 3), and most of the Euryarchaeota clones grouped within the Deep Sea Hydrothermal Vent Euryarchaeal Group 6 (Figure 4) or the Deep Sea Euryarchaeal Group (Figure 4). This pattern is consistent with our finding a Prochlorococcus phylotype in this lake (Clingenpeel et al., 2011), and presents opportunities to assess interesting evolutionary relationships between marine and freshwater microorganisms.

## Acknowledgements

This research was supported primarily by a grant from the Gordon and Betty Moore Foundation (Grant No. 1555), with additional funding from the National Park Service Centennial Challenge Match Program (PMIS No. 137808). We particularly thank James Cole and Qiong Wang of the Ribosome Database Project for their assistance in examining sample data sets to better understand variations in classifier annotations. Work was conducted under NPS research permit no. 5700.

# References

Aguilar C, Cuhel RL, Klump JV. Porewater and hydrothermal water inputs to Yellowstone vent Lake, Wyoming. In: Anderson RJ, Harmon D (eds). Yellowstone Lake: Hotbed of Chaos or Reservoir of Resilience? Proceedings of the Sixth Biennial Scientific Conference on the Greater Yellowstone Ecosystem; 8–10 October 2001; Mammoth Hot Springs Hotel, Yellowstone National Park. Yellowstone Center for Resources and George Wright Society: Hancock,

- MI, 2002, pp 1–18. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. (1990). Basic local alignment search tool. J Mol Biol **215**: 403-410.
- Auguet J-C, Borrego CM, Bañeras L, Casamayor EO. (2008). Fingerprinting the genetic diversity of the biotin carboxylase gene (accC) in aquatic ecosystems as a potential marker for studies of carbon dioxide assimilation in the dark. Environ Microbiol 10: 2527–2536.
- Baker GC, Smith JJ, Cowan DA. (2003). Review and reanalysis of domain-specific 16S primers. J Microbiol Methods 55: 541-555.
- Balistrieri LS, Shanks III WC, Cuhel RL, Aguilar C, Klump JV. The influence of sub-lacustrine hydrothermal vents on the geochemistry of Yellowstone Lake. In: Anderson RJ, Harmon D (eds). Yellowstone Lake: Hotbed of Chaos or Reservoir of Resilience? Proceedings of the Sixth Biennial Scientific Conference on the Greater Yellowstone Ecosystem; 8-10 October; Mammoth Hot Springs Hotel, Yellowstone National Park. Yellowstone Center for Resources and George Wright Society: Hancock, MI, 2007, pp 169-199.
- Barns SM, Delwiche CF, Palmer JD, Pace NR. (1996). Perspectives on archaeal diversity, thermophily and monophily from environmental rRNA sequences. Proc Natl Acad Sci USA 93: 9188-9193.
- Barns SM, Fundyga RE, Jeffries MW, Pace NR. (1994). Remarkable archaeal diversity in a Yellowstone hot spring environment. Proc Natl Acad Sci USA 91: 1609-1613.
- Benson NG. (1961). Limnology of Yellowstone Lake in relation to the cutthroat trout. United States Department of the Interior Fish and Wildlife Service, Bureau of Sport Fisheries and Wildlife. Research report no. 56.
- Bintrim SB, Donohue TJ, Handelsman J, Roberts GP, Goodman RM. (1997). Molecular phylogeny of Archaea from soil. Proc Natl Acad Sci USA 94: 277-282.
- Boetius A, Ravenschlag K, Schubert CJ, Rickert D, Widdel F, Gieseke A et al. (2000). A marine microbial consortium apparently mediating anaerobic oxidation of methane. Nature 407: 623-626.
- Bräuer S, Cadillo-Quiroz H, Ward RJ, Yavitt J, Zinder S. (2011). Methanoregula boonei gen. nov., sp. nov., an acidiphilic methanogen isolated from an acidic peat bog. Int J Syst Evol Microbiol 61(Part 1): 45-52.
- Brochier-Armanet C, Boussau B, Gribaldo S, Forterre P. (2008). Mesophilic crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. Nat Rev Microbiol 6: 245–252.
- Buckley DH, Graber JR, Schmidt TM. (1998). Phylogenetic analysis of nonthermophilic members of the kingdom Crenarchaeota and their diversity and abundance in soils. Appl Environ Microbiol 64: 4333-4339.
- Clingenpeel S, Macur RE, Kan J, Inskeep WP, Lovalvo D, Varley J et al. (2011). Yellowstone Lake: high energy geochemistry and rich bacterial diversity. Environ Microbiol (in press). doi:10.1111/j.1462-2920.2011.02466.x.
- Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ et al. (2009). The Ribosomal database project: improved alignments and new tools for rRNA analysis. Nucleic Acids Res 37(Database issue): D141–D145.
- Cuhel RL, Aguilar C, Anderson PD, Maki JS, Paddock RW, Remsen CC et al. Underwater domains in Yellowstone Lake hydrothermal vent geochemistry and bacterial chemosynthesis. In: Anderson RJ, Harmon D (eds).

Archaea in Yellowstone Lake J Kan et al

- Yellowstone Lake: Hotbed of Chaos or Reservoir of Resilience? Proceedings of the 6th Biennial Scientific Conference on the Greater Yellowstone Ecosystem; 8–10 October 2001; Mammoth Hot Springs Hotel, Yellowstone National Park. Yellowstone Center for Resources and George Wright Society: Hancock, MI, 2002, pp 27–53.
- D'Imperio S, Lehr CR, Oduro H, Druschel G, Kühl M, McDermott TR. (2008). Relative importance of  $H_2$  and  $H_2S$  as energy sources for primary production in geothermal springs. *Appl Environ Microbiol* **74**: 5802–5808.
- de la Torre JR, Walker CB, Ingalls AE, Könneke M, Stahl DA. (2008). Cultivation of a thermophilic ammonia oxidizing archaeon synthesizing crenarchaeol.. *Environ Microbiol* **10**: 810–818.
- DeLong EF. (1992). Archaea in coastal marine environments. *Proc Natl Acad Sci USA* **89**: 5685–5689.
- DeLong EF. (1998). Everything in moderation: Archaea as 'non-extremophiles. *Curr Opin Genet Dev* 8: 649–654.
- DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K *et al.* (2006). Greengenes, a chimerachecked rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* **72**: 5069–5072.
- Eder W, Ludwig W, Huber R. (1999). Novel 16S rRNA gene sequences retrieved from highly saline brine sediments of Kebrit Deep, Red Sea. *Arch Microbiol* **172**: 213–218.
- Elkins JG, Podar M, Graham DE, Makarova KS, Wolf Y, Randau L *et al.* (2008). A korarchaeal genome reveals insights into the evolution of the Archaea. *Proc Natl Acad Sci USA* **105**: 8102–8107.
- Fuhrman JA, McCallum K, Davis AA. (1992). Novel major archaebacterial group from marine plankton. *Nature* 356: 148–149.
- Glockner FO, Fuchs BM, Amann R. (1999). Bacterioplankton compositions of lakes and oceans: a first comparison based on fluorescence *in situ* hybridization. *Appl Environ Microbiol* **65**: 3721–3726.
- Hamady M, Walker JJ, Harris JK, Gold NJ, Knight R. (2008). Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. *Nat Methods* 5: 235–237.
- Huber H, Hohn MJ, Rachel R, Fuchs T, Wimmer VC, Stetter KO. (2002). A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. *Nature* **417**: 63–67.
- Huse SM, Welch DM, Morrison HG, Sogin ML. (2010). Ironing out the wrinkles in the rare biosphere through improved OTU clustering. *Environ Microbiol* **12**: 1889–1898.
- Jurgens G, Glockner FO, Amann R, Saano A, Montonen L, Likolammi M, Munster U. (2000). Identification of novel Archaea in bacterioplankton of a boreal forest lake by phylogenetic analysis and fluorescent *in situ* hybridization. *FEMS Microbiol Ecol* **34**: 45–56.
- Jurgens G, Lindstorm K, Saano A. (1997). Novel group within the kingdom Crenarchaeota from boreal forest soil. *Appl Environ Microbiol* **63**: 803–805.
- Keough BP, Schmidt TM, Hicks RE. (2003). Archaeal nucleic acids in picoplankton from great lakes on three continents. *Microb Ecol* **46**: 238–248.
- Könneke M, Bernhard AE, de la Torre JR, Walker CB, Waterbury JB, Stahl DA. (2005). Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* **437**: 543–546.

- Kunin V, Engelbrektson A, Ochman H, Hugenholtz P. (2010). Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environ Microbiol* **12**: 118–123.
- Langner H, Jackson CR, McDermott TR, Inskeep WP. (2001). Rapid oxidation of arsenite in a hot spring ecosystem, Yellowstone National Park. *Environ Sci Technol* **35**: 3302–3309.
- Lliro's M, Casamayor EO, Borrego CM. (2008). High archaeal richness in the water column of a freshwater sulphurous karstic lake along an inter-annual study. *FEMS Microbiol Ecol* **66**: 331–342.
- Logares R, Bråte J, Bertilsson S, Clasen JL, Shalchian-Tabrizi K, Rengefors K. (2009). Infrequent marinefreshwater transitions in the microbial world. *Trends Microbiol* **17**: 414–422.
- Lovalvo D, Clingenpeel SR, McGinnis S, Macur RE, Varley JD, Inskeep WP *et al.* (2010). A geothermallinked biological oasis in Yellowstone Lake, Yellowstone National Park, Wyoming. *Geobiology* **8**: 327–336.
- Ludwig W, Strunk O, Westram R, Richter L, Meier H, Kumar Y *et al.* (2004). ARB: a software environment for sequence data. *Nucleic Acids Res* **32**: 1363–1371.
- MacGregor BJ, Moser DP, Alm EW, Nealson KH, Stahl DA. (1997). Crenarchaeota in Lake Michigan sediment. *Appl Environ Microbiol* **63**: 1178–1181.
- MacGregor BJ, Moser DP, Baker BJ, Alm EW, Maurer M, Nealson KH, Stahl DA. (2001). Seasonal and spatial variability in Lake Michigan sediment small-subunit rRNA concentrations. *Appl Environ Microbiol* **67**: 3908–3922.
- Macur RE, Langner HW, Kocar BD, Inskeep WP. (2004). Linking geochemical processes with microbial community analysis: successional dynamics in an arsenic-rich, acid-sulphate-chloride geothermal spring. *Geobiology* **2**: 163–177.
- Maidak BL, Olsen GJ, Larsen N, Overbeek R, McCaughey MJ, Woese CR. (1997). The RDP (Ribosomal Database Project). *Nucleic Acids Res* **25**: 109–111.
- 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.
- Meyer-Dombard, DR, Shock EL, Amend JP. (2005). Archaeal and bacterial communities in geochemically diverse hot springs of Yellowstone National Park, USA. *Geobiology* **3**: 211–227.
- Morgan LA, Blackwell DD, Spafford RE, Smith RB. (1977). Heat flow measurements in Yellowstone Lake and the thermal structure of the Yellowstone caldera. *J Geophys Res* **82**: 3719–3732.
- Morgan LA, Shanks III WC, Lovalvo D, Johnson SY, Stephenson W, Pierce KL *et al.* (2003). Exploration and discovery in Yellowstone Lake: results from highresolution sonar imaging, seismic reflection profiling, and submersible studies. *J Volcanol Geotherm Res* **122**: 221–242.
- Morgan LA, Shanks III WC, Pierce KL, Lovalvo DA, Lee GK, Webring MW et al. (2007). The floor of Yellowstone Lake is anything but quiet—new discoveries from high resolution sonar imaging, seismicreflection profiling, and submersible studies. In: Morgan LA (ed.). Integrated Geoscience Studies in the Greater Yellowstone Area—Volcanic, Tectonic, and Hydrothermal Processes in the Yellowstone Geoecosystem, US Geological Survey Professional Paper No. 1717, pp 95–126.

1794

- Nicol GW, Schleper C. (2006). Ammonia-oxidizing *Crenarchaeota*: important players in the nitrogen cycle? *Trends Microbiol* **14**: 207–212.
- Ochsenreiter T, Selezi D, Quaiser A, Bonch-Osmolovskaya L, Schleper C. (2003). Diversity and abundance of Crenarchaeota in terrestrial habitats studied by 16S RNA surveys and real time PCR. *Environ Microbiol* 5: 787–797.
- Orphan VJ, House CH, Hinrichs KU, McKeegan KD, DeLong EF. (2002). Multiple archaeal groups mediate methane oxidation in anoxic cold seep sediments. *Proc Natl Acad Sci USA* **99**: 7663–7668.
- Pernthaler J, Glockner FO, Unterholzer S, Alfreider A, Psenner R, Amann R. (1998). Seasonal community and population dynamics of pelagic bacteria and archaea in a high mountain lake. *Appl Environ Microbiol* **64**: 4299–4306.
- Pruesse E, Quast C, Knittel K, Fuchs B, Ludwig W, Peplies J, Glöckner FO. (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* 35: 7188–7196.
- Remsen CC, Klump JV, Kaster J, Padock R, Anderson P, Maki JS. (1990). Hydrothermal springs and gas fumuroles in Yellowstone Lake, Yellowstone National Park, Wyoming. Natl Geograph Res 6: 509–515.
- Remsen CC, Maki JS, Val Kump J, Auilar C, Anderson PD, Buchholz L et al. Sublacrustrine geothermal activity in Yellowstone Lake: studies past and present. In: Anderson RJ, Harmon D (eds). Yellowstone Lake: Hotbed of Chaos or Reservoir of Resilience? Proceedings of the Sixth Biennial Scientific Conference on the Greater Yellowstone Ecosystem; 8–10 October 2001; Mammoth Hot Springs Hotel, Yellowstone National Park. Yellowstone Center for Resources and George Wright Society: Hancock, MI, 2002, pp 192–212.
- Robertson CE, Harris JK, Spear JR, Pace NR. (2005). Phylogenetic diversity and ecology of environmental Archaea. *Curr Opin Microbiol* **8**: 638–642.
- Roesch LFW, Fulthorpe RR, Riva A, Casella G, Hadwin AKM, Kent AD *et al.* (2007). Pyrosequencing enumerates and contrasts soil microbial diversity. *ISME J* 1: 283–290.
- Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S, Yooseph S *et al.* (2007). The Sorcerer II Global Ocean Sampling Expedition: Northwest Atlantic through Eastern Tropical Pacific. *PLoS Biol* **5**: e77.
- Sandaa RA, Enger O, Torsvik V. (1999). Abundance and diversity of Archaea in heavy-metal-contaminated soils. *Appl Environ Microbiol* **65**: 3293–3297.

- Schleper C. (2010). Ammonia oxidation: different niches for bacteria and archaea? *ISME J* **4**: 1092–1094.
- Schleper C, Holben W, Klenk HP. (1997). Recovery of crenarchaeotal ribosomal DNA sequences from freshwater lake sediments. Appl Environ Microbiol 63: 321–323.
- Schleper C, Jurgens G, Jonuscheit M. (2005). Genomic studies of uncultivated archaea. Nat Rev Microbiol 3: 479–488.
- Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75**: 7537–7541.
- Simon HM, Dodsworth JA, Goodman RM. (2000). Crenarchaeota colonize terrestrial plant roots. *Environ Microbiol* **2**: 495–505.
- Spear JR, Walker JJ, McCollom TM, Pace NR. (2005). Hydrogen and bioenergetics in the Yellowstone geothermal ecosystem. *Proc Natl Acad Sci USA* **102**: 2555–2560.
- Suzuki MT, Giovannoni SJ. (1996). Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. *Appl Environ Microbiol* **62**: 625–630.
- Urbach E, Vergin K, Larson G, Giovannoni S. (2007). Bacterioplankton communities of Crater Lake, OR: dynamic changes with euphotic zone food web structure and stable deep water populations. *Hydrobiologia* 574: 161–177.
- Von Wintzingerode F, Gobel UB, Stachebrandt E. (1997). Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiol Rev* **21**: 213–229.
- Wang QG, Garrity M, Tiedje JM, Cole JR. (2007). Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* **73**: 5261–5267.
- Woese CR, Kandler O, Wheelis ML. (1990). Towards a natural system of organisms-proposal for the domain Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci USA 87: 4576–4579.
- Young M, Wiedenheft B, Snyder J, Spuhler J, Roberto F, Douglas T. (2005). Archaeal viruses from Yellowstone's high temperature environments. In: Inskeep WP, McDermott TR (eds). Geothermal Biology and Geochemistry in Yellowstone National Park. Proceedings of the Thermal Biology Workshop, Yellowstone National Park, WY, october 2003. Montana State University: Bozeman, MT, pp 289–304.

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