Extremophiles (2017) 21:135–152
DOI 10.1007/s00792-016-0889-8

ORIGINAL PAPER

Microbial community analysis of pH 4 thermal springs in Yellowstone National Park

Xiaoben Jiang[1] · **Cristina D. Takacs-Vesbach**[1]

Received: 20 February 2016 / Accepted: 20 October 2016 / Published online: 2 November 2016

Abstract The pH of the majority of thermal springs in Yellowstone National Park (YNP) is from 1 to 3 and 6 to 10; relatively few springs (~5%) have a pH range of 4–5. We used 16S rRNA gene pyrosequencing to investigate microbial communities sampled from four pH 4 thermal springs collected from four regions of YNP that differed in their fluid temperature and geochemistry. Our results revealed that the composition of bacterial communities varied among the sites, despite sharing similar pH values. The taxonomic composition and metabolic functional potential of the site with the lowest temperature (55 °C), a thermal spring from the Seven Mile Hole (SMH) area, were further investigated using shotgun metagenome sequencing. The taxonomic classification, based on 372 Mbp of unassembled metagenomic reads, indicated that this community included a high proportion of Chloroflexi, Bacteroidetes, Proteobacteria, and Firmicutes. Functional comparison with other YNP thermal spring metagenomes indicated that the SMH metagenome was enriched in genes related to energy production and conversion, transcription, and carbohydrate transport. Analysis of genes involved in nitrogen metabolism revealed assimilatory and dissimilatory nitrate reduction pathways, whereas the majority of genes involved in sulfur metabolism were related to the reduction of sulfate to adenylylsulfate, sulfite, and H_2S. Given that pH 4 thermal springs are relatively less common in YNP and thermal areas worldwide, they may harbor novel microbiota and the communities that inhabit them deserve further investigation.

Keywords Yellowstone National Park · Thermal springs · pH 4 · Metagenomics · 16S rRNA · Microbial communities

Introduction

Yellowstone National Park (YNP) is one of the largest and most diverse hydrothermal areas on Earth, and it harbors more than 12,000 thermal springs that are characterized by a broad range of temperature (40–92 °C), pH (1–10), and geochemical properties (Fournier 1989; Rye and Truesdell 2007). YNP thermal springs often have abundant and diverse electron donors (e.g., H_2, sulfide, S^0, thiosulfate, and Fe^{2+}) and electron acceptors (e.g., dissolved O_2, S^0 and Fe^{3+}), which provide an abundance of potential niches. As a consequence, thermal springs support microbial communities that comprise a diverse array of metabolisms, including photoautotrophs, photoheterotrophs, chemolithoautotrophs, and chemoorganotrophs (Amend and Shock 2001).

pH is a primary environmental factor that directly influences microbial community composition in thermal springs at the regional and global scales (Boyd et al. 2010; Boyd et al. 2013; Dequiedt et al. 2009; Inskeep et al. 2013b; Song et al. 2013; Xie et al. 2015). While the range of pH in YNP thermal springs is broad (1–10), the majority of thermal springs in YNP can be classified into two categories by pH—acidic, vapor-dominated systems, and circumneutral

Communicated by A. Oren.

Electronic supplementary material The online version of this article (doi:10.1007/s00792-016-0889-8) contains supplementary material, which is available to authorized users.

✉ Cristina D. Takacs-Vesbach
cvesbach@unm.edu

Xiaoben Jiang
sdpapet@unm.edu

[1] Department of Biology, MSC03 2020 1UNM, University of New Mexico, Albuquerque, NM 87131, USA

to alkaline, water-dominated systems (Fournier 1989). The vapor-dominated springs, which often discharge relatively little liquid water, contain H_2S that oxidizes to H_2SO_4 when it contacts air in perched pools of ground water. In contrast, the water-dominated systems discharge significant amounts of circumneutral or alkaline water enriched in chloride (Fournier 1989). Notably, there are few thermal features in YNP with intermediate pH in the range of 4–5. For example, of the more than 7000 thermal features inventoried by the US National Park Service (data available online at http://www.rcn.montana.edu/Default.aspx), only ~5% of the entries have a pH between 4 and 5. Microbial communities in the two pH systems common in YNP (acidic and circumneutral to alkaline) have extensively been studied and are known to each harbor distinct communities. Circumneutral to alkaline springs often include microbial communities that are dominated by members of the Aquificales, Chloroflexi, and Cyanobacteria (Inskeep et al. 2010, 2013b; Madigan 2003; Meyer-Dombard et al. 2005; Meyer-Dombard et al. 2011; Reysenbach et al. 1994, 2000; Spear et al. 2005; Ward et al. 1998b). Although the Archaea are present in circumneutral springs, they are estimated to be a lesser fraction of the total biomass (Inskeep et al. 2013b; Miller et al. 2009; Ward et al. 1998a). Conversely, Archaea predominate the microbial communities in acidic, vapor-dominated springs, particularly with elevated temperatures (Inskeep et al. 2013a), and often include members of the Crenarchaeota, Euryarchaeota, and Korarchaeota (Brock et al. 1972; Inskeep et al. 2013a; Jackson et al. 2001; Meyer-Dombard et al. 2005). A bimodal pH distribution among terrestrial thermal springs has been noted for thermal areas worldwide as well (Brock 1971). As a result, the previous surveys of microbial communities from other geothermal hotspots around the world, including El Tatio, Chile (Engel et al. 2013), Kamojang, Indonesia (Aditiawati et al. 2009), Nakabusa, Japan (Everroad et al. 2012), Odisha, India (Sen and Maiti 2014), and Tibet, China (Song et al. 2013; Wang et al. 2013), mainly focused on acidic springs with pH below 3 or circumneutral to alkaline springs with pH above 6. A recent metagenomic investigation of two intermediate-pH (e.g., ~4), high-temperature (e.g., >55 °C) sites found taxonomic profiles similar to those from acidic springs (Inskeep et al. 2013b). To date, no effort has been made to survey the microbial inhabitants and their functional roles in low-temperature (i.e., 45–55 °C), intermediate-pH (e.g., ~4) thermal springs (Inskeep et al. 2013b). Thus, little is known about the microbial ecology of intermediate-pH springs in YNP.

The goals of this study were to: (1) investigate taxonomic profiles of four YNP springs within an intermediate-pH range (4.05–4.35) using 16S rRNA gene amplicon pyrosequencing and (2) characterize the metabolic potential of one of these sites, a low-temperature (55 °C) thermal spring containing a distinct microbial community. To our knowledge, this study is the first survey of microbial taxonomic and functional diversity of pH 4 springs using a combined 16S rRNA gene amplicon and metagenomic sequencing approach.

Materials and methods

Site description and sample collection

Four geothermal springs were selected for field measurements and sample collection in the following YNP thermal areas (sample names are given in parentheses): Norris (NOR), Mary Bay Area (MRY), Mud Kettles (MKL), and Seven Mile Hole (SMH). Locations, descriptions, and photos are listed in Table 1 and Fig. 1. Water samples for geochemical analysis were collected at each site in conjunction with the biomass samples for sequencing analysis. Water samples were collected from the overflow channels of spring sources as proximate to the center of flow as possible, where the water was well mixed. Water for geochemical analysis was filtered through a 0.2 µm Sterivex filter using sterile 50 mL syringes and preserved as appropriate for the analysis to be performed (McCleskey et al. 2005). Briefly, syringes were rinsed three times with site water before collection and all samples were collected into acid pre-washed polyethylene bottles (soaked in 5% HCl for 3 h and rinsed three times with deionized water), except the anion samples, which were collected into deionized water pre-washed polyethylene bottles (soaked for 3 h and rinsed three times with deionized water). Water (30 mL) for cation analysis was preserved by acidification using 0.3 mL 3 N nitric acid, whereas water (125 mL) for anion analysis did not receive any protective reagents. Water (100 mL) for As and Fe species was collected into opaque polyethylene bottles and preserved with 1 mL of 6 N HCl, and water (30 mL) for the ammonium analysis was preserved with the addition of 0.3 mL of 4.5 N H_2SO_4. For the SiO_2 analysis, 1 mL of water was immediately diluted by 9 mL deionized water to prevent precipitation. For the sulfate analysis, 30 mL of water was preserved with 0.5 mL of $ZnCl_2$, followed by 0.5 mL of NaOH. Samples for the determination of dissolved organic carbon (DOC, 100 mL) were collected into heat-combusted (475 °C for 4 h) amber glass bottles and preserved with 1 ml 6 N HCl. All water samples were transported and stored at 4 °C until analysis which was no more than 2 weeks after sample collection.

Sediment, mat, or filament samples were collected aseptically into 2 mL microcentrifuge tubes and preserved in sucrose lysis buffer (SLB; 20 mM EDTA, 200 mM NaCl, 0.75 M sucrose, 50 mM Tris–HCl, pH 9.0). Sediment samples were collected from the top 3 cm of the spring bed,

Table 1 Geographic and geochemical parameters for the four sampling sites in YNP

Sample area	Norris	Mary Bay Area	Mud Kettles	Seven Mile Hole
Sample name	NOR	MRY	MKL	SMH
Sample ID	03YNOR021	03YMRY047	03YMKL049	04YSMH020
GPS location (N/W)	44.732423/−110.709777	44.553361/−110.3045805	44.634664/−110.6045546	44.754916/−110.4158659
Physical context	Sediment, pool edge	Sediment, pool edge	Sediment, gas turbulent pool bottom	Clumpy mat, thermal creek
Temperature (°C)	84	80	72	55
pH	4.34	4.32	4.35	4.05
Alkalinity as HCO_3 (1.0 mg L^{-1})	BD	BD	BD	BD
Conductivity (μS/cm)	2130	1260	1170	1538
DOC (0.1 mg L^{-1})	0.3	BD	1.7	BD
Ca (0.1 mg L^{-1})	4.3	6.0	46.0	88.8
K (0.04 mg L^{-1})	44.00	4.20	23.00	30.00
Na (0.1 mg L^{-1})	382.2	4.6	137.6	152.5
Mg (0.001 mg L^{-1})	0.020	4.000	19.000	29.100
SO_4 (1.0 mg L^{-1})	73	542	571	306
Cl (0.05 mg L^{-1})	599.00	3.67	1.82	272.00
Al (0.07 mg L^{-1})	0.91	BD	0.43	0.26
As (total) (0.0001 mg L^{-1})	2.1600	0.0380	0.0020	0.2000
As (III) (0.0005 mg L^{-1})	2.1600	0.0250	0.0010	0.0050
Ba (0.001 mg L^{-1})	0.019	0.040	0.083	0.073
Be (0.001 mg L^{-1})	0.004	BD	BD	BD
Cd (0.0001 mg L^{-1})	BD	0.0001	BD	0.0001
Co (0.0007 mg L^{-1})	BD	BD	BD	0.003
Cr (0.0005 mg L^{-1})	BD	BD	BD	0.0029
Cu (0.0005 mg L^{-1})	0.0014	0.0010	0.0006	0.0009
Fe (total) (0.002 mg L^{-1})	0.033	1.400	0.264	0.009
Fe(II) (0.002 mg L^{-1})	0.031	0.359	0.252	0.003
Li (0.003 mg L^{-1})	3.800	0.005	0.072	0.611
Mn (0.001 mg L^{-1})	0.003	0.130	0.290	1.070
Ni (0.002 mg L^{-1})	BD	BD	BD	0.010
P (0.02 mg L^{-1})	BD	0.02	0.13	BD
Pb (0.0008 mg L^{-1})	BD	BD	BD	BD
Sb (0.001 mg L^{-1})	0.060	0.002	BD	0.001
Se (0.001 mg L^{-1})	0.007	BD	BD	0.004
SiO_2 (0.05 mg L^{-1})	396.00	237.00	282.00	190.00
Sr (0.001 mg L^{-1})	0.014	0.019	1.100	1.010
V (0.002 mg L^{-1})	BD	BD	BD	BD
Zn (0.001 mg L^{-1})	0.016	0.067	0.078	0.031
Br (0.1 mg L^{-1})	2.0	BD	BD	0.9
F (0.1 mg L^{-1})	8.7	0.3	0.5	0.5
H_2S (0.002 mg L^{-1})	0.327	1.175	1.625	BD
NH_4 (0.3 mg L^{-1})	7.2	BD	33.1	16.9
NO_2 (0.0003 mg L^{-1})	0.0029	BD	0.0013	0.0530
NO_3 (0.1 mg L^{-1})	BD	0.2	BD	37.0

Detection limits and units of chemical species measured are given in parentheses

BD below the detection limit

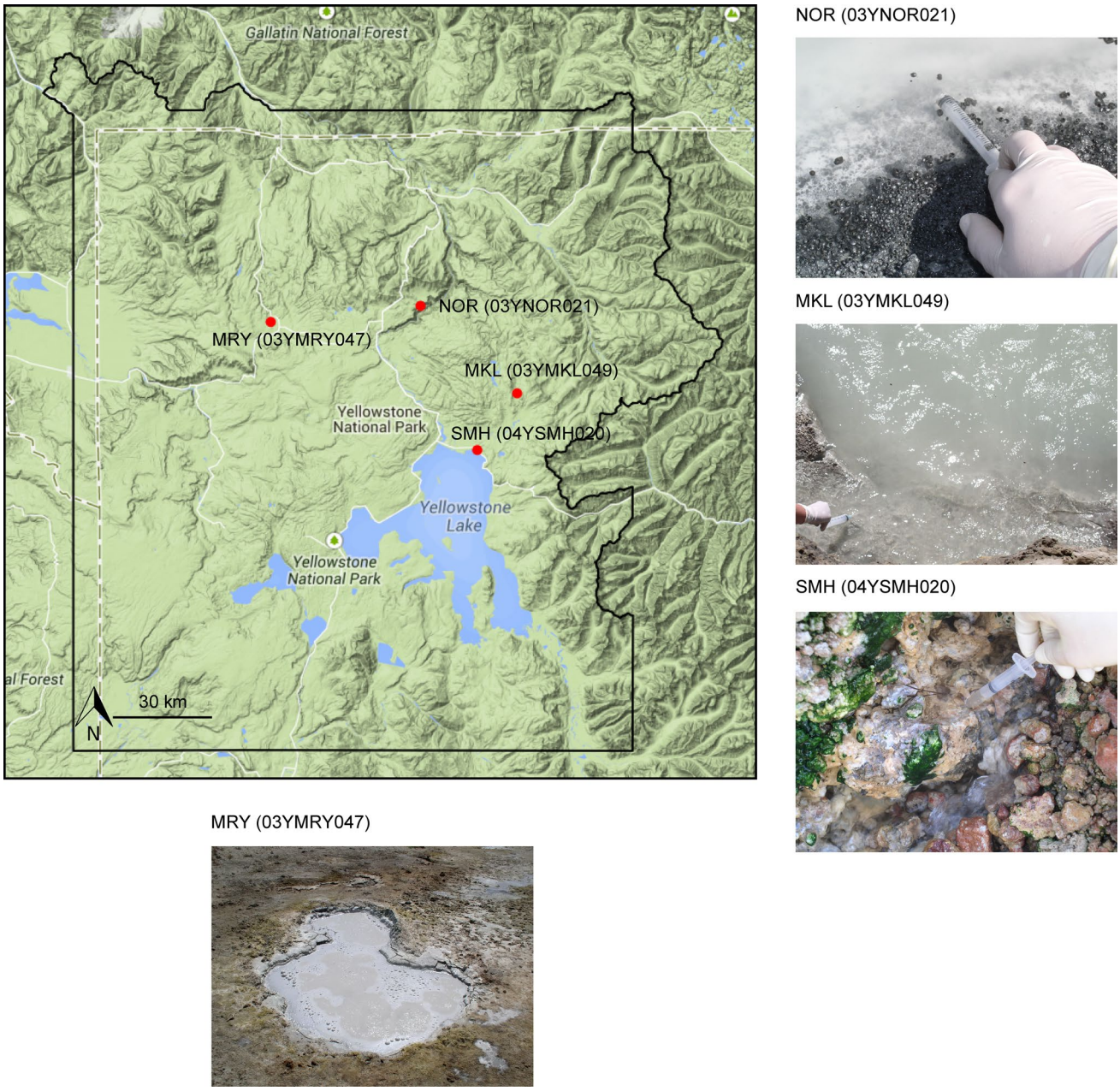

Fig. 1 Geographic map and photos showing sampling locations

and biofilm samples (mats or filaments) were collected right below the air–water interface (<1 cm). Samples were stored at ambient temperature (~10–26 °C) for up to 10 days before they were stored at −80 °C. The previous experiments indicated that storage of samples in SLB without freezing did not lead to a loss of DNA or microbial diversity relative to samples immediately frozen in liquid nitrogen (Mitchell and Takacs-Vesbach 2008). Once in the laboratory, samples were stored at −80 °C until DNA extraction.

Geochemical analyses

At each sampling location, temperature and pH were measured using a Thermo Orion 290A+ meter, and electrical conductivity was measured with a WTW meter with temperature correction. Dissolved H_2S was measured in the field using a portable colorimeter (Hach DR/850) by the methylene blue method (APHA 1985). Because dissolved H_2S is volatile and oxidized quickly, spring water was directly drawn into a plastic syringe and filtered through a

0.2 µm filter into a measuring cuvette. Methylene blue reagents were added immediately and the absorbance and temperature of the solution were measured after color development. The temperature dependence of the H_2S-methylene blue color complex was corrected using the method detailed in McCleskey et al. (2005). DOC concentrations were measured using the wet oxidation method (Aiken 1992) with a TOC Analyzer (Oceanography International Model 700). Major anions were measured using ion chromatography (IC), and cations and trace metals were measured using inductively coupled plasma-optical emission spectrometry (ICP-OES). All geochemical analyses, including anions and cations, were conducted using standard USGS methods, and typical measurement uncertainties were <5% (McCleskey et al. 2005). Major ion composition of these four sites was compared with 97 other YNP inventory sites that were sampled and analyzed using the same methods (data available online at http://www.vesbachlab.org/data.html) as part of a larger microbial inventory conducted in YNP.

DNA extraction

DNA was extracted from 0.2 g of each sample following bead-beating in a CTAB buffer (1% CTAB, 0.75 M NaCl, 50 mM Tris pH 8, 10 mM EDTA) and subsequent phenol–chloroform purification steps as described in (Mitchell and Takacs-Vesbach 2008). Briefly, 2 volumes of 1% CTAB buffer and proteinase K (final concentration 100 µg mL^{-1}) were added to the samples, which were then incubated for 1 h at 60 °C. SDS (final concentration 2%) and 0.1 mm diameter Zirconia/Silica beads were added. Samples were bead beaten for 45 s at 50 strokes per second. After incubating for 1 h at 60 °C, DNA was extracted once with an equal volume phenol:chloroform:isoamyl alcohol (25:24:1), followed by two extractions with an equal volume of chloroform. Finally, the DNA was precipitated with two volumes of 95% ethanol, washed with 70% ethanol, dried by speed-vac, and reconstituted with 50 µL of filter-sterilized, autoclaved 10 mM Tris pH 8.0. DNA extracts were quantified using a Nanodrop ND-2000c spectrophotometer.

16S rRNA gene pyrosequencing

Barcoded amplicon pyrosequencing of 16S rRNA genes was performed as described previously (Van Horn et al. 2013). Briefly, DNA isolated from each sample was amplified using the universal bacterial primers 28F (5′-GAGTTTGATC NTGGCTCAG-3′) and 519R (5′-GTNTTACNGCGGCKG CTG-3′), and archaeal primers Arch349F (5′-GYGCASC AGKCGMGAAW-3′) and Arch806R (5′-GGACTACVSG GGTATCTAAT-3′) targeting the 16S rRNA genes as described previously (Colman et al. 2015; Rhoads et al. 2012). PCR was performed as follows: an initial cycle of 95 °C for 5 min, followed by 30 cycles of 95 °C for 30 s, 54 °C for 40 s, and 72 °C for 1 min, and a final elongation step for 10 min at 72 °C. Successful amplification was confirmed by agarose gel electrophoresis. Triplicate PCR mixtures per sample were combined and subsequently purified with an UltraClean™ GelSpin™ DNA Extraction Kit (MoBio Laboratories, Carlsbad, CA, USA). The purified DNA was quantified using a Nanodrop ND-2000c spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, USA). Amplicons from all samples with different barcodes were pooled at equimolar concentrations for pyrosequencing on a 454 GS FLX (454 Life Sciences, Branford, CT, USA) using Titanium reagents according to the manufacturer's protocol.

16S rRNA gene pyrosequencing data processing

Raw sequences obtained from pyrosequencing were denoised to correct for sequencing errors and remove low-quality sequences and potential sequencing chimeras using AmpliconNoise (Quince et al. 2011) integrated into QIIME (Ver. 1.8.0, Caporaso et al. 2010b). Adapters, multiplex identifiers, and primers were trimmed from denoised data. Operational taxonomic units (OTUs) were identified at the 97% DNA similarity level using UCLUST (Edgar 2010) in QIIME. The most abundant sequence from each OTU was picked as a representative sequence and aligned using the PyNAST aligner (Caporaso et al. 2010a) and the Greengenes database (GG 13_5, DeSantis et al. 2006). Taxonomic assignments were made using the Ribosomal Database Classifier program (Wang et al. 2007). Good's coverage estimates for the data sets were performed with randomly drawn subsets of 800 sequences per sample to standardize for varying sequencing efforts across samples.

Shotgun metagenome sequencing

The Seven Mile Hole (04YSMH020, SMH, Table 1) sample was selected for further characterization by metagenomic sequencing, because it contained a relatively unique microbial community compared with the other samples described here, and the previous reports of pH 4 sites were focused on sites with much higher temperatures (>55 °C, Inskeep et al. 2013b). Approximately 500 ng of SMH DNA was used for library construction. Metagenome library preparation and sequencing were performed on one-half of a picotiter plate according to manufacturer's protocol on a 454 GS FLX Titanium platform (454 Life Sciences, Branford, CT, USA).

COG function enrichment analysis

Metagenomic sequencing reads were quality-filtered and assembled using Newbler 2.6 (Margulies et al. 2005) using default settings. Contigs and singleton reads were submitted to the JGI IMG/M annotation pipeline (Markowitz et al. 2008) and annotated using the Clusters of Orthologous Group (COG, Tatusov et al. 2000) database.

To provide an assessment of the microbial community type of this site, COG functions from the SMH metagenome assembly were compared with functions from other published YNP metagenomes (Table S1, Inskeep et al. 2013b). For each metagenome, data were normalized by the total number of COG functions detected and weighted by contig depth if assembly information was available. For unassembled singleton reads, a contig depth of one was assumed. COG functions were classified into COG categories on IMG/M and a Bray–Curtis dissimilarity matrix based on the COG category abundance table was subsequently constructed. Principal Coordinates Analysis (PCoA) was performed with COG categories and a two-way hierarchical clustering was also done on COG category abundance to confirm the grouping pattern observed from the PCoA analysis. All multivariate comparisons and ordinations were performed using the R (Team 2011) statistical software with the 'vegan' (Oksanen et al. 2012) and 'cluster' (Maechler et al. 2012) packages.

The "Function Comparison" tool on the IMG/M server was used to determine which COG functions were statistically overrepresented in the SMH data set compared with other publically available YNP metagenomic data sets that were most similar to SMH (7 phototrophic samples identified by the PCoA and cluster analyses described above). The relative abundances of COG functions were calculated based on normalized gene counts and expressed as *d* scores (Markowitz et al. 2008). *d* score is equivalent to the standard variation from the null hypothesis (i.e., relative gene counts in metagenome A = relative gene counts in metagenome B). For each comparison, the *P* value cutoff for statistical significant *d* scores was assessed using a false discovery rate of 0.05.

Community composition and metabolic mapping

Unassembled raw reads were also annotated on the metagenomic analysis server, MG-RAST (Meta Genome Rapid Annotation using Subsystem Technology, v3.3 Glass et al. 2010), using the default quality control pipeline. Microbial composition and functional analyses were conducted via the MG-RAST best-hit classification tool against the GenBank (NCBI-nr), M5NR (M5 non-redundant protein), and RefSeq (NCBI Reference Sequences) databases using a minimum identity of 60%, e-value cutoff of 10^{-5} and a minimum alignment length of 50 bp.

BLASTX results from the NCBI-nr database were imported into the MEtaGenome ANalyzer software (MEGAN v4.70.4, Huson et al. 2007). Taxonomic classifications were made using the least common ancestor (LCA) algorithm based on the top 10 BLAST alignments for each read, and metabolic pathways were classified by KEGG database (Ogata et al. 1999). The sequences in each pathway (oxidative phosphorylation, methane metabolism, nitrogen metabolism, carbon fixation pathways in prokaryotes, carbon fixation in photosynthetic organisms, sulfur metabolism, and photosynthesis) were given taxonomic assignments at the phylum level. The pathways involved in nitrogen metabolism and sulfur metabolism were mapped and were reconstructed using KEGG identifiers.

Sequence data submission

Raw 16S rRNA gene amplicon sequencing data reported here are available through the NCBI Sequence Read Archive. The individual sff files were assigned the accession numbers SRX1031281–SRX1031284 under Bioproject PRJNA284196. The SMH metagenome is publicly available on IMG/M (SMH: IMG submission ID 13526) and MG-RAST (SMH: ID 4523620.3).

Results

Geochemistry

The physical and geochemical parameters measured for the four springs studied here are reported in Table 1. The pH values of the springs we sampled ranged from 4.05 to 4.35, whereas temperature ranged from 55 to 84 °C. Site SMH was the site with the lowest temperature (55 °C), compared with the other three sampling sites, which had temperatures that were 72 °C or higher (Table 1). Chloride and sulfate concentrations ranged among the sites from 1.82 to 599 mg L^{-1} and 73.3 to 571 mg L^{-1}, respectively (Table 1). Compared with other YNP springs (shown in Fig. 2), the concentration of carbonates and bicarbonates in all the pH 4 sites was low. Waters in MKL and MRY had the highest SO_4^{2-}/Cl^- ratios, while NOR had the lowest SO_4^{2-}/Cl^- ratio site (Fig. 2). SMH, the site selected for metagenomic analysis, had a moderate SO_4^{2-}/Cl^- ratio (Fig. 2).

Taxonomic profiles of 16S rRNA gene amplicon sequencing

A total of 6432 bacterial 16S rRNA gene sequences were obtained from the four pH 4 sites after denoising and

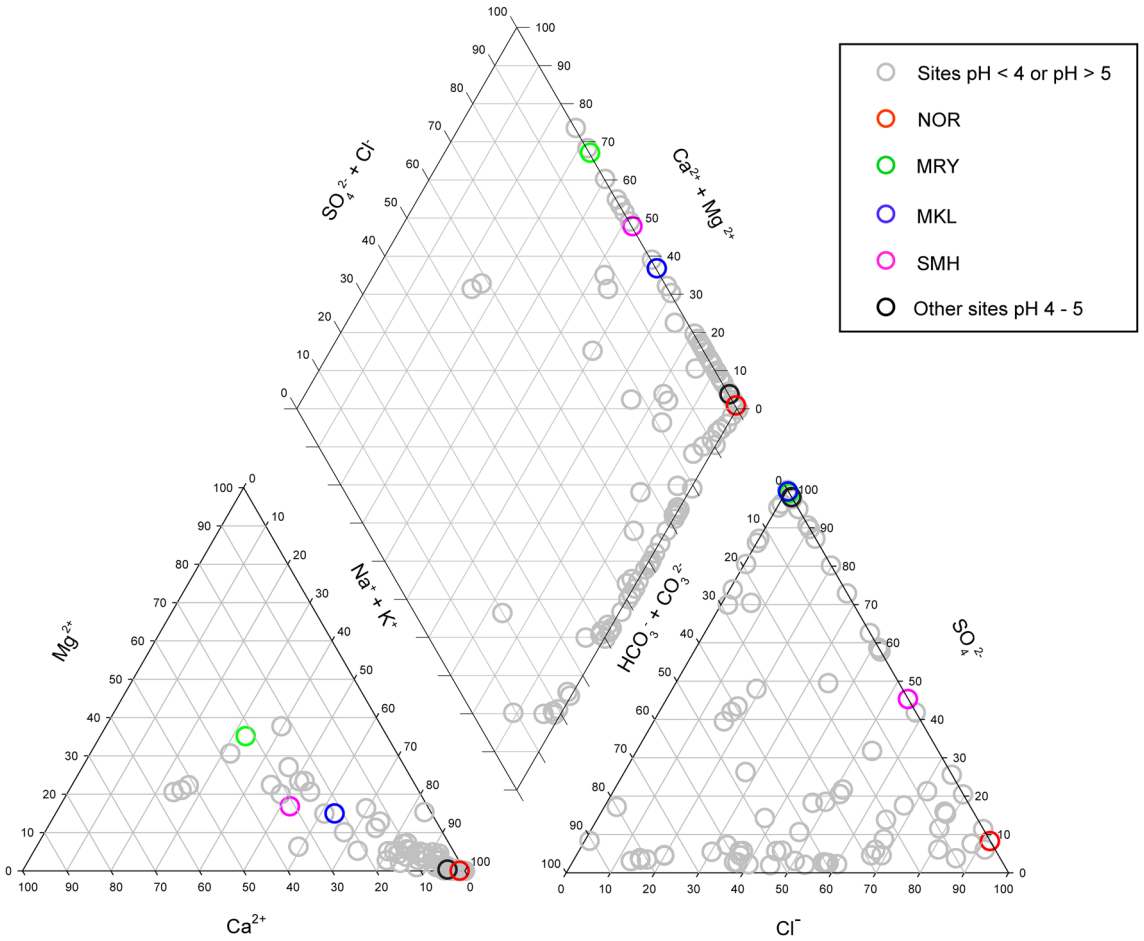

Fig. 2 Piper diagram indicating major ion compositions of four sites in this study together with the other YNP inventory sites

removing low-quality or chimeric sequences. No archaeal 16S rRNA gene sequences were amplified with the Arch349F and Arch806R primers despite varying PCR conditions. Negative PCR results were confirmed by adding Archaeal genomic DNA to PCR reactions containing sample DNA, indicating that either there was no Archaea template in the sample DNA or our PCR primers did not target any Archaea that were present in these samples. The bacterial 16S rRNA gene sequences clustered into 226 OTUs at 97% DNA similarity. Good's coverage (Good 1953), which provides an estimate of survey completeness, ranged from 91.1 to 98.5% (mean = 95.9%). Cyanobacteria, Proteobacteria, and Chloroflexi were the predominant bacterial phyla within NOR (38.5, 23.4 and 7.9% of 16S rRNA gene sequences, respectively, Fig. 3), whereas Cyanobacteria, Chlorobi, and Chloroflexi were the three most abundant bacterial groups within MRY (82.5, 7.9, and 5% of 16S rRNA gene sequences, respectively, Fig. 3). MKL was comprised mostly of Aquificae (93.5% of the 16S rRNA gene sequences), while SMH was dominated by Armatimonadetes, Chloroflexi, and Bacteroidetes (52.8, 18.6 and 18% of total 16S rRNA gene sequences, respectively, Fig. 3).

Metagenome sequencing, coverage, and overview of microbial community groups

Metagenome sequencing generated 848,583 reads (mean length = 438 bp) totaling 372 Mbp for SMH (Table 2). Assembly of the metagenomic sequence data set yielded 19,346 contigs, with an N50 contig length of 4303 bp. The annotation results of SMH metagenome from MG-RAST and IMG are detailed in Table 3.

PCoA analysis based on COG functional categories indicated that the 20 public YNP metagenomes (Table S1), along with SMH, clustered into three distinct groups that could be characterized by their dominant members as: (1) archaeal communities; (2) Aquificales communities; and (3) phototrophic communities (Fig. 4). The grouping detected through the PCoA analysis was also confirmed by

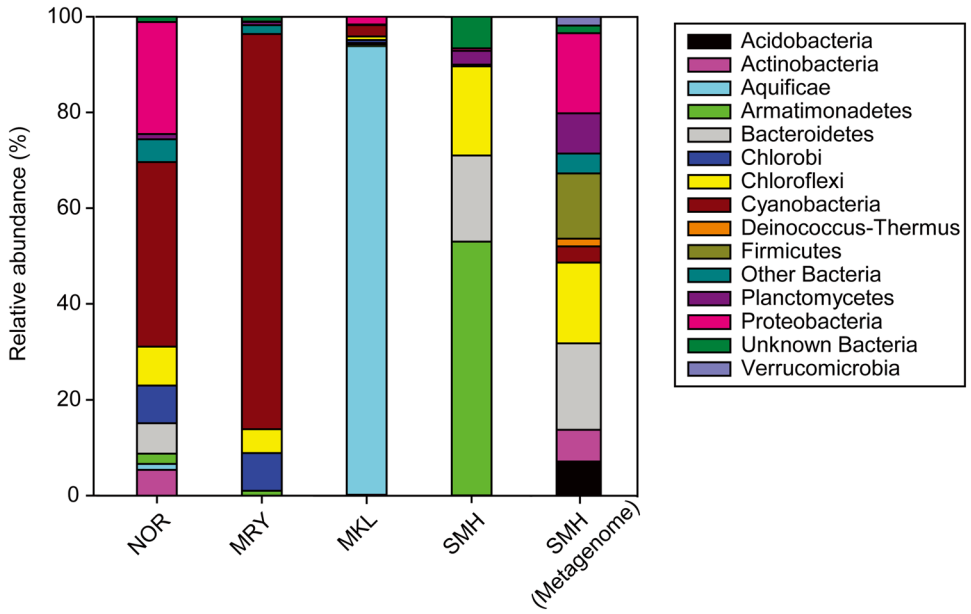

Fig. 3 Bacterial taxonomic classification and comparison of 16S rRNA gene and metagenomic reads. Taxonomic classification based on 16S rRNA gene amplicon pyrosequencing using the Greengenes database. Taxonomic classification based on metagenome using the MG-RAST M5NR database

Table 2 454 pyrosequencing and newbler assembly metrics of the metagenomic DNA sample from the site SMH (04YSMH020)

Parameter	SMH (04YSMH020)
Total number of reads	8,48,583
Mean read length	438 bp
Metagenome size (unassembled reads)	372 Mbp
Metagenome size (assembled reads)	38 Mbp (10.2%)
Number of reads in contigs	6,41,401
Number of contigs	19,346
Reads/contig	31.15
Largest contig (bp)	1,87,560
Mean contig length (bp)	2554
N50 contig length (bp)	4303
Number of singletons	1,29,660

the cluster analysis (Fig. S1). The site SMH clustered most closely with the phototrophic community group (Figs. 4, S1).

Taxonomic profiles of the metagenomics

Overall, community structure analysis performed with the M5NR database on MG-RAST indicated that SMH was dominated by bacteria (81.04%, Table 4). The remaining sequences from SMH matched Archaea (6.36%), Eukaryota (0.21%), and unassigned sequences (12.38%), or unclassified sequences (0.06%). The taxonomic distribution of numerically abundant phyla derived from the metagenome of SMH indicated that Chloroflexi (~17.8%), Bacteroidetes (~17.7%), Proteobacteria (~13.5%), and

Table 3 Features of the thermal spring metagenome SMH (04YSMH020) based on MG-RAST and IMG/M annotations

Annotation platform	MG-RAST[a]	IMG/M[b]
Total number of reads post MG-RAST quality control	7,30,387	–
Total DNA scaffolds post IMG/M data processing	–	19,293
Average GC content	52 ± 10%	–
Protein coding sequences	4,26,466	53,332
Protein coding sequences with function prediction	185,985 (43.6%)	32,441 (60.16%)
rRNA genes	402	63

– Not applicable or not determined

[a] Features from unassembled reads that passed MG-RAST quality control

[b] Features from Newbler assembled reads post IMG/M data processing

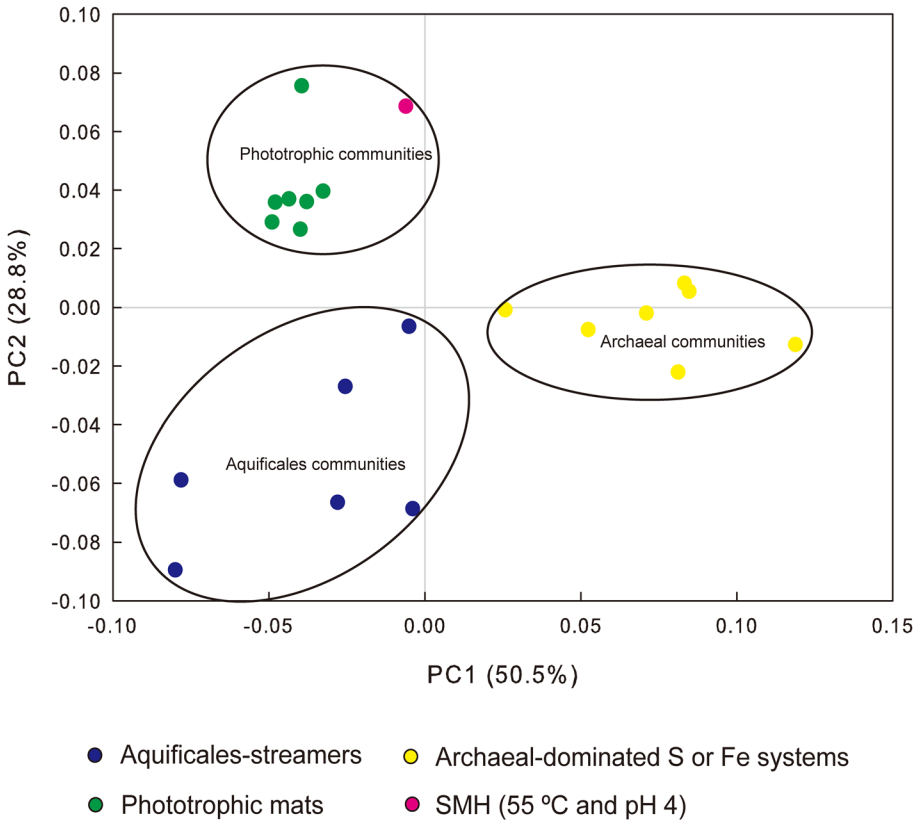

Fig. 4 Principal coordinates analysis (PCoA) of 21 YNP metagenomes based on COG categories

Table 4 Domain distribution on the metagenomic sample based on M5NR database

Domain	%
Archaea	6.36
Bacteria	81.04
Eukaryota	0.21
Viruses	–
Other sequences	0.01
Unassigned	12.32
Unclassified sequences	0.06

Unassembled reads annotated on MG-RAST were analyzed using the classification tool based on M5NR (60% minimum identity) with maximum e-value cutoff of 1×10^{-5} and minimum alignment length of 50 bp)

– Not applicable or not determined

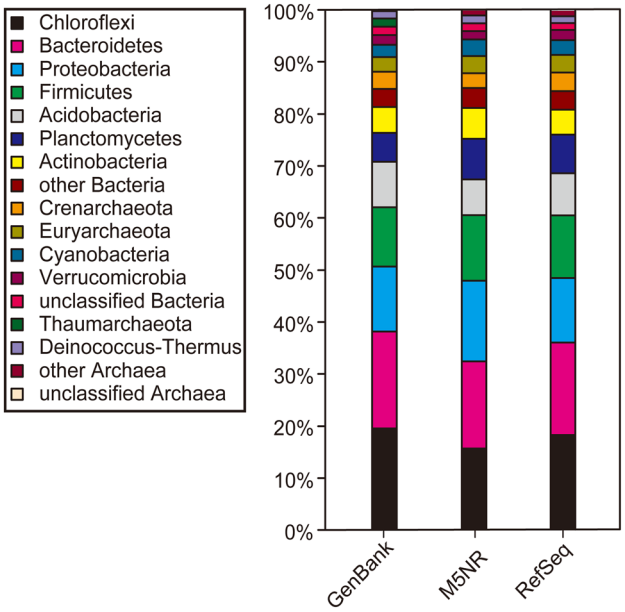

Fig. 5 Comparison of the taxonomic assignment of unassembled SMH metagenomic sequences based on GenBank (NCBI-nr), M5NR, and RefSeq databases

Firmicutes (~12.0%) were the four most abundant phyla, according to the GenBank, M5NR, and RefSeq databases (Fig. 5). Ktedonobacteria was the dominant class (~58.0%) in phylum Chloroflexi based on all three databases (Fig. S2). Sphingobacteria was the dominant class in the phylum Bacteroidetes (~52.4%, Fig. S2). Class Deltaproteobacteria was most abundant in phylum Proteobacteria (~36.5%, Fig. S2). At the class level, Clostridia accounted for ~66.8% of all Firmicutes reads, followed by Bacilli (~28.8%, Fig. S2), Negativicutes (~4.2%, Fig. S2), and Erysipelotrichi (~0.3%, Fig. S2). The majority of archaeal sequences were related to the Crenarchaeota

and Euryarchaeota (with ~2.87 and ~3.28%, respectively), however, because Archaea comprised only ~6% of the SMH sequences, the remainder of our analysis focused on the Bacteria.

Gene functions enriched in the metagenome of SMH

The metagenome of SMH provided information on the functional capabilities of a microbial community in a relatively low-temperature, pH 4 site. PCoA and cluster analyses showed that the functional profile of the SMH community was most similar to seven previously published YNP metagenomes (Inskeep et al. 2013b) characterized as phototrophic communities (Fig. 4; Fig. S1). However, differences were detected in SMH compared with these seven phototrophic communities. Twenty-seven COG functions were significantly overrepresented in the SMH data set in at least six of the seven comparisons (Table S2). "Energy production and conversion", "Transcription" and "Carbohydrate metabolism and transport" were the three most abundant COG categories among these functions (25.9, 18.5, and 11.1%, respectively) and were dominant among those with highest enrichment *d* scores (Table S2; Fig. 6). The 30 most abundant COG functions are listed in Table 5.

Energy metabolism mapping

The functional assignment of the unassembled SMH metagenomic data set provided information about possible functions in this community. A total of 14,957 reads were assigned to energy metabolism using BLASTX against the NCBI-nr database, and the majority of reads were related to the domain Bacteria (~92%). Among the bacterial reads, most were mapped to Bacteroidetes, Chloroflexi, Firmicutes, Proteobacteria, and Planctomycetes involved in diverse pathways, such as oxidative phosphorylation, methane metabolism, nitrogen metabolism, carbon fixation pathways in prokaryotes, carbon fixation in photosynthetic organisms, sulfur metabolism, and photosynthesis (Fig. 7). Major KEGG function categories and unique hits assigned to each category are listed in Tables S3 and S4. Genes involved in nitrate reduction were among the abundant categories associated with nitrogen metabolism (Table S3; Fig. S3). Genes encoding sulfate adenylyltransferase, cysteine synthase, and sulfite reductase were also highly enriched in the metagenome of SMH (Table S4; Fig. S4) compared with other similar published metagenomes.

Discussion

Despite the diverse types of springs observed among thermal areas in YNP, their fluids ultimately originate from meteoric water. The parent water is modified through various processes, including water–rock interactions, subsurface mixing, and boiling and cooling during transport to the surface (Fournier 1989, 2005; Truesdell and Fournier 1976; Truesdell et al. 1977). Thermal waters in circumneutral to alkaline springs represent deeply sourced meteoric water enriched in carbonate, chloride, and silica, which emerge from faults located at relatively low elevations. By contrast, sulfate waters in acidic, vapor-dominated springs are generally discovered at the higher elevation unfractured lava caps (Fournier 2005). H_2S rich steam separates from the underlying chloride-rich neutral water and enters perched pools of ground water, where H_2S is oxidized to H_2SO_4 either abiotically or biotically (Fournier 1989, 2005; Guo et al. 2014). The presence of impermeable rock caps derived from young lava flows covering old geothermal areas in YNP about 150,000–100,000 years ago effectively segregates circumneutral to alkaline and acidic waters, which may explain the paucity of springs with intermediate pH in YNP (Guo et al. 2014; Hurwitz et al. 2007; Morgan et al. 2005; White 1957).

Although the pH range of all our sites is narrow, profound differences exist in SO_4^{2-}/Cl^- ratios across sampling sites, reflecting the different extent of mixing between acid-sulfate waters and circumneutral to alkaline, chloride-rich waters (Fig. 2). Acidic geothermal waters (pH < 6) in YNP can be further classified into springs with high SO_4^{2-}/Cl^- ratios and low SO_4^{2-}/Cl^- ratios (Fournier 1989; Guo et al. 2014). The MKL and MRY samples represent sites with high SO_4^{2-}/Cl^- ratios, whereas SMH and NOR have low SO_4^{2-}/Cl^- ratios due to elevated Cl^- content. The SMH waters possibly result from the mixing of high SO_4^{2-}/Cl^- ratio waters with circumneutral to alkaline, chloride meteoric waters (Guo et al. 2014). The low SO_4^{2-}/Cl^- ratio found in the NOR water sample (Fig. 2) is consistent with the previous observations, indicating that mixing between the two types of waters is common in Norris Geyser Basin (Fournier et al. 2002; Nordstrom et al. 2009). Owing to their unique geochemistry, springs with intermediate pH may harbor distinct microbial communities.

In this study, we employed several approaches to examine the microbial communities in a relatively understudied YNP thermal spring niche, focusing on four pH 4 springs, with temperatures ranging from 55 to 84 °C. Site SMH had the lowest temperature (55 °C), while the three other sites were much warmer (≥72 °C, see Inskeep et al. 2013b). Although we did not detect the Archaea by 16S rRNA gene sequencing in any of the pH 4 sites, archaeal DNA sequences were detected in the SMH metagenome (~6.36% of the total sequences). This discrepancy might be explained by the potential bias of archaeal primers (Colman et al. 2015). Given the dominance of the Bacteria in both data sets, we focused on bacterial communities in this

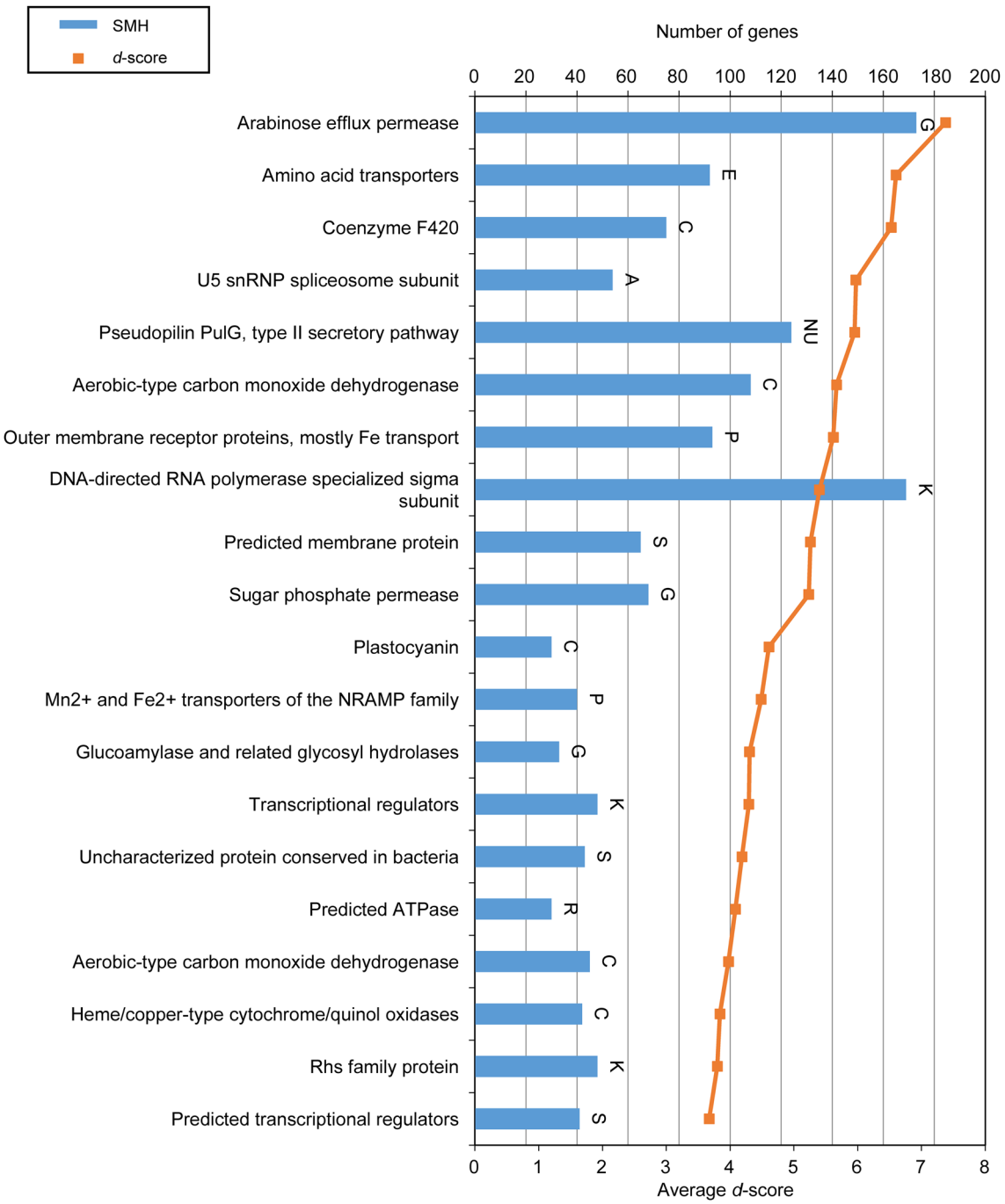

Fig. 6 Overrepresented COG functions in the metagenome of SMH relative to seven other phototrophic metagenomes. These represent the 20 most enriched functions from a total of 27 COG functions that received significant enrichment scores in six of seven comparisons (Table S2). Average *d* score represents the mean enrichment score over all seven comparisons. *Letters above graphs* represent COG category

study. Distinct community assemblages were recovered at sites with different temperatures (Fig. 3), despite similar pH values (Table 1). For example, the Aquificae dominated 16S rRNA gene data from MKL (Fig. 5a), one of the high-temperature sites (72 °C). The Aquificae normally predominates in springs with temperatures above 70 °C (Inskeep et al. 2013b; Reysenbach et al. 2005), or high temperature reaches within a spring where photosynthesis is temperature limited (Cole et al. 2013; Everroad et al. 2012; Hall et al. 2008; Huber and Stetter 2001). In contrast, the lower temperature in situ, together with the mixing of acid-sulfate waters with circumneutral to alkaline, chloride-rich

Table 5 Top 30 (by sequence count) COG functions represented in SMH (04YSMH020) metagenomic assembled sequences based on IMG/M annotations

COG category	COG ID	COG name	Sequence count	Ranking
M	COG0438	Glycosyltransferase	247	1
G	COG2814	Arabinose efflux permease	173	2
K	COG1595	DNA-directed RNA polymerase specialized sigma subunit, sigma24 homolog	169	3
TK	COG0745	Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain	161	4
R	COG0673	Predicted dehydrogenases and related proteins	146	5
IQR	COG1028	Dehydrogenases with different specificities (related to short-chain alcohol dehydrogenases)	145	6
RTKL	COG0515	Serine/threonine protein kinase	142	7
H	COG2226	Methylase involved in ubiquinone/menaquinone biosynthesis	141	8
V	COG1131	ABC-type multidrug transport system, ATPase component	138	9
NU	COG2165	Type II secretory pathway, pseudopilin PulG	124	10
R	COG4783	Putative Zn-dependent protease, contains TPR repeats	122	11
NU	COG3063	Tfp pilus assembly protein PilF	119	12
M	COG0463	Glycosyltransferases involved in cell wall biogenesis	117	13
T	COG2204	Response regulator containing CheY-like receiver, AAA-type ATPase, and DNA-binding domains	109	14
C	COG1529	Aerobic-type carbon monoxide dehydrogenase, large subunit CoxL/CutL homologs	108	15
MG	COG0451	Nucleoside-diphosphate-sugar epimerases	107	16
E	COG0747	ABC-type dipeptide transport system, periplasmic component	107	17
I	COG1960	Acyl-CoA dehydrogenases	105	18
P	COG1629	Outer membrane receptor proteins, mostly Fe transport	93	19
E	COG0531	Amino acid transporters	92	20
T	COG0642	Signal transduction histidine kinase	91	21
I	COG1024	Enoyl-CoA hydratase/carnithine racemase	90	22
O	COG1225	Peroxiredoxin	89	23
R	COG0596	Predicted hydrolases or acyltransferases (alpha/beta hydrolase superfamily)	88	24
TK	COG2197	Response regulator containing a CheY-like receiver domain and an HTH DNA-binding domain	88	25
C	COG1012	NAD-dependent aldehyde dehydrogenases	86	26
IQ	COG0318	Acyl-CoA synthetases (AMP-forming)/AMP-acid ligases II	84	27
V	COG0841	Cation/multidrug efflux pump	82	28
T	COG2205	Osmosensitive K^+ channel histidine kinase	78	29
C	COG2141	Coenzyme F420-dependent N^5,N^{10}-methylenetetrahydromethanopterin reductase and related flavin-dependent oxidoreductases	75	30

waters in SMH (Fig. 2), may contribute to the broader taxonomic composition observed in this spring (e.g., Armatimonadetes, Bacteroidetes, and Chloroflexi, Fig. 3). Although Armatimonadetes sequences dominated the 16S rRNA gene sequencing reads, they were not detected in the metagenomic annotation (Figs. 3, 5). We did not detect a similar discrepancy between the amplicon and metagenomic sequencing among any of the other bacterial phyla. The Armatimonadetes is a newly described phylum and is estimated to comprise 12 groups, occurring in a variety of environments (Dunfield et al. 2012), but only a few strains of Armatimonadetes have been isolated to date (Dunfield et al. 2012). The lack of isolates and classification may be attributed to inaccurate annotations, which results in incongruent results. We detected 16S rRNA gene amplicons from Cyanobacteria in NOR and MRY (Fig. 3), where temperatures were 84 and 80 °C, respectively (Table 1), which is above the temperature limit for photosynthesis (~70–75 °C) (Hamilton et al. 2012; Klatt et al. 2011; Rothschild and Mancinelli 2001). Samples from sites NOR and MRY were collected from the upper 3 cm of sediment, where dead cells from allochthonous phototrophic microorganisms inhabiting cooler margins of the stream channel may deposit and accumulate.

Although the four springs had similar pH values, it is likely that the distinct geochemistry of each spring

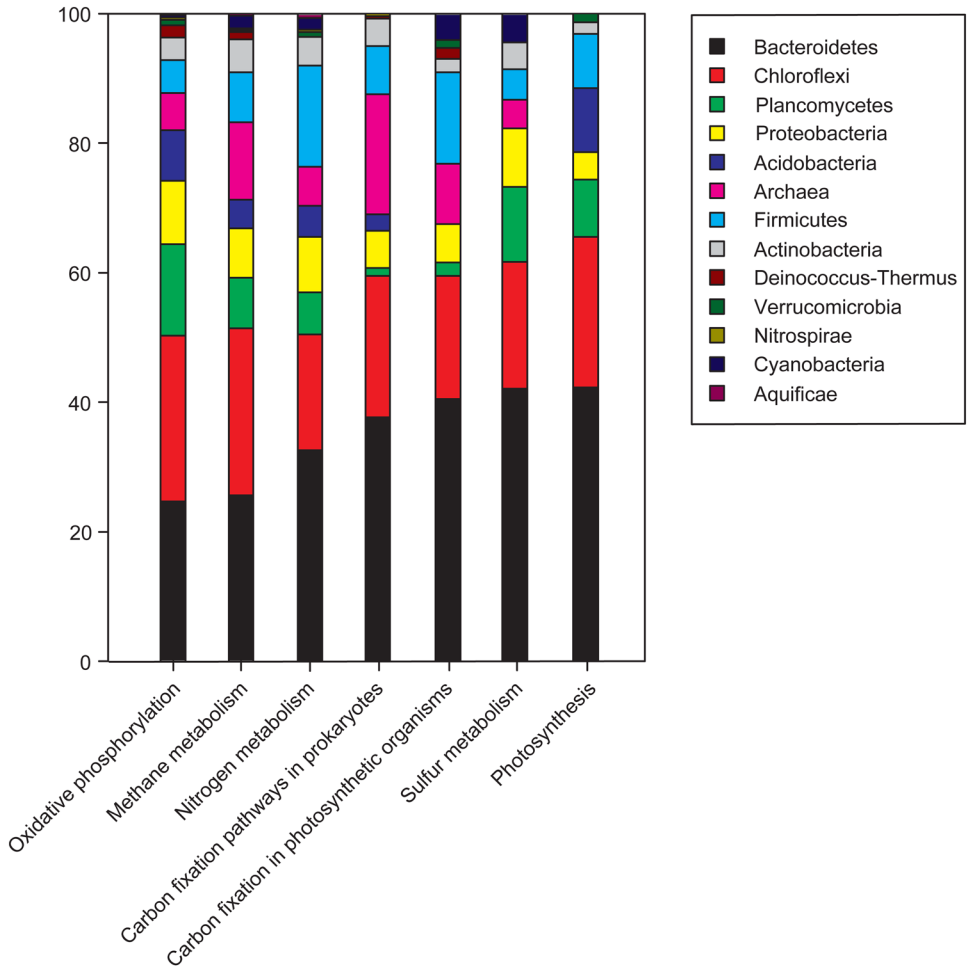

Fig. 7 Taxonomic assignment of metagenomic reads from the site SMH related to energy metabolism (KEGG identifiers)

additionally contributed to the community differences observed. In particular, biologically important species, such as SO_4^{2-}, NH_4^+, and NO_3^-, varied among the springs and concentrations were high compared with other springs sampled parkwide. The concentration of SO_4^{2-} in MKL, for example, was about seven orders of magnitude higher than that in NOR (Table 1; Fig. 2). Of note, in contrast to the paucity of NO_3^- in NOR and MKL, the NO_3^- concentration (37.0 mg L^{-1}) in SMH was the highest among 104 inventoried sites representative of the diversity of sites parkwide (http://vlab.lternet.edu/ynp_inv_data_products.html).

We compared the metagenome of the low-temperature (55 °C) site SMH, to 20 YNP metagenomes from a previous report (Inskeep et al. 2013b), to better understand how the microbial community of this uninvestigated habitat compared with other relatively well-characterized communities from YNP. We expected site SMH to group with the archaeal sites, because they were all from low-pH sites (pH 2–4). Instead, site SMH, which was collected at a temperature 17 °C cooler than the coldest archaeal site (One Hundred Spring Plain, 72 °C, Table S1), clustered with phototrophic sites (Fig. 4; Fig. S1), likely due to the predominance of phototrophic organisms and the low abundance of Archaea (~6%) in SMH.

The metagenome of SMH provided a first glimpse of the metabolic functional profile of a low-temperature, pH 4 site of YNP. Functions overrepresented in the SMH metagenome compared with other YNP samples were responsible for energy production and represented important redox reactions and key steps in electron transport. Some enriched COG functions (Table S2; Fig. 6), such as coenzyme F420 (COG 2141), heme Cu oxidase (COG1622), and carbon monoxide dehydrogenase (COG1529, COG2080, and COG1319) and proteins (COG3794 and COG0723), involved in electron transport can be related directly to the microbial community and their metabolic potentials. For example, COG 2141 associated with coenzyme F420-dependent 5,10-methylenetetrahydromethanopterin reductase belongs to the family of oxidoreductases responsible for redox reactions in many Actinobacteria and methanogenic Archaea (Deppenmeier 2002). Heme-copper-type oxidases (COG 1622) representing the terminal energy-transfer enzymes of respiratory chains play a

significant role in aerobic metabolism (Garcia-Horsman et al. 1994). COGs related to carbon monoxide dehydrogenase can be involved in diverse biochemical pathways, including aerobic carboxydotrophic, sulfate-reducing, and hydrogenogenic bacteria.

Among the overrepresented functions, we observed many relevant to carbohydrate metabolism and transport. For instance, COG2814 belongs to the family of "arabinose efflux permease". Proteins of this COG function belong to the major facilitator superfamily (MFS) that can transport a diverse array of substrates, including amino acids, drugs, ions, and sugars across the membrane (Law et al. 2008). Another overrepresented function COG2271, a sugar phosphate permease, which is responsible for carbohydrate transport derived from the environment, is also affiliated with the MFS. Functions (e.g., COG 1131 and COG 0841, Table 5) associated with multidrug resistance are known to export antibiotics and toxic molecules (Piddock 2006). Bacteria bearing these functions can defend against toxic compounds produced by competitors (Piddock 2006). Functions (COG1629 and COG1914) related to inorganic ion transport and metabolism are significantly abundant in SMH. Microorganisms in thermal springs are expected to encounter heavy metals (Inskeep et al. 2010) and possess genes involved in heavy metal transport.

Pathways involved in nitrogen and sulfur metabolism frequently contribute significantly to energy generation by thermal spring microbial communities, because alternative electronic acceptors, such as arsenate, CO_2, elemental sulfur, ferric iron, nitrate, sulfate, and thiosulfate, are often abundant at spring sources (Hall et al. 2008; Inskeep et al. 2010; Jimenez et al. 2012). Approximately 81.1% of the genes detected in SMH associated with nitrogen metabolism were related to the Bacteroidetes, Chloroflexi, Firmicutes, Proteobacteria, and Planctomycetes (Fig. 7). Nitrate as an electron acceptor is energetically favorable over broad pH ranges, which may explain the widespread distribution of nitrate reduction genes in different types of YNP springs (Shock et al. 2010; Swingley et al. 2012). In addition, the nitrate concentration in SMH was the highest among all our YNP inventory sites (37.0 mg L^{-1}, http://vlab.lternet.edu/ynp_inv_data_products.html), thus representing an abundant nutrient and energy resource for nitrate reducers.

Genes necessary for dissimilatory nitrate reduction to N_2 via denitrification, including dissimilatory nitrate reductase gene clusters (e.g., *nar*G, *nar*H, *nar*I, and *nar*J) and nitric oxide reductase gene clusters (e.g., *nor*B, *nor*C, *nor*D, and *nor*Q) were prevalent in SMH (Table S3; Fig. S3). In addition, the gene coding for nitrite reductase (*nir*K) was detected (Table S3; Fig. S3). Nitrite reductase (*nir*K or *nir*S) is a pivotal enzyme of dissimilatory nitrate reduction pathway. According to models of dissimilatory nitrate reduction in bacteria (Gonzalez et al. 2006; Moreno-Vivián et al. 1999; Richardson et al. 2001), a nitrite reductase (*nir*K or *nir*S) is requisite for producing NO, which is a substrate for nitric oxide reductase (e.g., *nor*B, *nor*C) to produce N_2O. The sequences related to gene *nos*Z coding for nitrous oxide reductase (associated with class Aquificae, Bacteroidetes, Flavobacteriia, Ignavibacteriae, Sphingobacteriia, and Thermomicrobia) are important for the last step of denitrification, which converts N_2O to N_2 (Table S3; Fig. S3).

Given the abundance of ammonium (16.9 mg L^{-1}) in SMH and the high-energy costs of biological nitrogen fixation, the absence of *nif*K (Table S3; Fig. S3), a gene involved in the synthesis of molybdenum-dependent nitrogenase (Dos Santos et al. 2012), is not surprising. The source of ammonium in SMH is most likely abiotic (Holloway et al. 2011), because genes associated with the dissimilatory nitrate reduction to ammonium (DNRA) pathway (e.g., *nap*A, *nrf*A) were completely absent from SMH and have not been reported for YNP. The absence of DNRA in SMH may be due to factors, such as the lack of dissolved organic carbon (below detection) and high nitrate availability (37.0 mg L^{-1}; Table 1) in situ. The previous studies suggest that DNRA activities can outcompete denitrification activities under high C/NO_3^- ratios (Tiedje et al. 1983) and low nitrate availability (van den Berg et al. 2015). Ammonium assimilation is possible in SMH by genes coding for assimilatory nitrate reduction to ammonia (e.g., *nas*A, *nir*A, and *nir*B). In this bacteria dominated spring, we did not detect genes (e.g., *amo*A) coding for the ammonium monooxygenase, which is able to use ammonia as a substrate. Our result is consistent with the previous reports of the absence of nitrification genes from other YNP thermal springs (Inskeep et al. 2010; Swingley et al. 2012). Nitrification may not predominate among YNP thermal springs, because the oxidation of ammonium is rarely thermodynamically favorable under in situ conditions (Shock et al. 2010), despite the occurrence of the archaeal *amo*A-like genes in the YNP axenic culture and environmental samples (de la Torre et al. 2008; Zhang et al. 2008).

Most of the genes involved in sulfur metabolism are related to the conversion of sulfate into adenylylsulfate and to the subsequent production of sulfite and H_2S (Table S4; Fig. S4), similar to what has been previously reported in other thermal springs (Jimenez et al. 2012; Swingley et al. 2012). Given that sulfide was below detection in SMH, whereas sulfate was 306 mg L^{-1} (Table 1), sulfide must be oxidized or incorporated rapidly. The abundance of available sulfate provides a large energy source for sulfate-reducing microbes, which is further supported by the pathway reconstruction of sulfate reduction based on the metagenomic gene content (Fig. S4).

Genes responsible for cysteine synthase A (*cys*K) and B (*cys*M) are implicated in the formation of adenylylsulfate

(Table S4; Fig. S4). The environmental *apr*A and *apr*B sequences coding for adenosine-5′-phosphosulfate (APS) reductase (Apr) exhibit closest matches to members of Betaproteobacteria and Thermoprotei (Table S4; Fig. S4). Based on the current models of dissimilatory sulfate reduction and sulfur oxidation in prokaryotes, adenosine-5′-phosphosulfate (APS) reductase (Apr) is a pivotal enzyme. During the process of sulfate reduction, the function of Apr is to convert APS to sulfite. Once sulfate is activated to APS by ATP-sulfurylase at the expense of ATP, sulfite is subsequently reduced to sulfide by dissimilatory sulfite reductases (DSRs, Meyer and Kuever 2008). The alpha subunit of Apr enzymes is considered to be ubiquitous in all known sulfate-reducing and most of the sulfur-oxidizing prokaryotes (Meyer and Kuever 2008). For example, environmental *apr*A reads found in site SMH have high identity to those annotated in the *Thiobacillus plumbophilus* and *Caldivirga maquilingensis* genomes (*e* value < 10^{-35} Table S4; Fig. S4). *T. plumbophilus* requires H_2S as an electron donor (Drobner et al. 1992), whereas *C. maquilingensis* respires sulfur, thiosulfate, or sulfate (Itoh et al. 1999). Genes coding for the assimilatory and dissimilatory reduction of adenylylsulfate to sulfite (e.g., *apr*A, *apr*B, and *cys*H) and the subsequent assimilatory reduction of sulfite to H_2S (e.g., *cys*I and *sir*) observed in site SMH suggest that sulfate and sulfite reduction pathways are dominant processes in the environment studied here. Unlike previously reported for a YNP alkaline spring (Swingley et al. 2012) where there is a genomic potential for sulfur oxidation, enzymes, such as sulfite oxidase (*sox*), were not detected in our data set, either due to insufficient sequencing depth, low abundance, or absence.

Conclusions

In this study, the microbial community of pH 4 springs was studied using 16S rRNA gene pyrosequencing. We found that the microbial assemblages varied among the four different sites studied, despite the narrow range of pH values sampled. Temperature and geochemistry of the waters likely contributed to the differences we observed. In addition, we assessed the functional profiles of the microbial community in a low-temperature, pH 4 spring that was previously unexplored, using shotgun metagenome sequencing. Functional cluster analyses revealed that this unexplored geobiological community was most similar to other phototrophic communities sampled from YNP, although the pH was more consistent with sites dominated by Archaea. Taxonomically, this spring community included a microbial assemblage and functional profile that were distinct from other phototrophic YNP communities, which are typically circumneutral. Apart from Chloroflexi that are commonly found in phototrophic communities, Bacteroidetes, Proteobacteria, and Firmicutes were abundant in this spring. The taxonomic diversity resulted in metabolic diversity (e.g., chemotrophs and heterotrophs), as described in the metagenomic data presented here. Compared with other YNP phototrophic communities, the annotation based on COG database indicated a relative enrichment of functions involved in energy production and conversion, transcription, and carbohydrate transport. The identification of genes coding for nitrogen and sulfur cycling revealed a microbial population involved in the dissimilatory and assimilatory reduction of nitrate, and conversion of sulfate into adenylylsulfate, sulfite, and H_2S. Low-temperature, intermediate-pH terrestrial hydrothermal springs in YNP harbor unique communities with diverse metabolisms that deserve further attention. This research not only provides an initial survey that serves as a foundation for understanding microbial communities in these less common springs, but also offers a framework for future microbial studies in pH 4 YNP thermal springs. It is notable that the waters from the four pH 4 sites in this study underwent different water mixing processes. Efforts to better integrate the role of water source and history may prove useful in understanding the microbial ecology of thermal springs.

Acknowledgements We are very grateful for the valuable suggestions from three anonymous reviewers. We appreciate helpful discussions with David J. Van Horn and Daniel R. Colman related to data analyses performed in the study, as well as sample collection from Kendra R. Mitchell, and sample processing by George Rosenberg of the Center for Evolutionary and Theoretical Immunology (CETI), University of New Mexico. This work was supported by NSF Biodiversity Surveys and Inventories Grant 02-06773 to CTV and NIH-sponsored CETI 454 Voucher Program funding to XJ. Research reported in this publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number P30GM110907. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

Aditiawati P, Yohandini H, Madayanti F, Akhmaloka (2009) Microbial diversity of acidic hot spring (Kawah Hujan B) in geothermal field of Kamojang area, West Java-Indonesia. Open Microbiol J 3:58–66

Aiken GR (1992) Chloride interference in the analysis of dissolved organic carbon by the wet oxidation method. Environ Sci Technol 26:2435–2439

Amend JP, Shock EL (2001) Energetics of overall metabolic reactions of thermophilic and hyperthermophilic Archaea and Bacteria. FEMS Microbiol Rev 25:175–243

APHA (1985) Method 428C. Methylene blue method for sulfide. In: Standard methods for the examination of water and wastewater, 14th edn. American Public Health Association, pp 403–405

Boyd ES, Hamilton TL, Spear JR, Lavin M, Peters JW (2010) [FeFe]-hydrogenase in Yellowstone National Park: evidence for

dispersal limitation and phylogenetic niche conservatism. ISME J 4:1485–1495
Boyd ES, Hamilton TL, Wang J, He L, Zhang CL (2013) The role of tetraether lipid composition in the adaptation of thermophilic archaea to acidity. Front Microbiol 4:62
Brock TD (1971) Bimodal distribution of pH values of thermal springs of the world. Geol Soc Am Bull 82:1393–1394
Brock T, Brock K, Belly R, Weiss R (1972) *Sulfolobus*: a new genus of sulfur-oxidizing bacteria living at low pH and high temperature. Arch Microbiol 84:54–68
Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R (2010a) PyNAST: a flexible tool for aligning sequences to a template alignment. Bioinformatics 26:266–267
Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R (2010b) QIIME allows analysis of high-throughput community sequencing data. Nat Methods 7:335–336
Cole JK, Peacock JP, Dodsworth JA, Williams AJ, Thompson DB, Dong H, Wu G, Hedlund BP (2013) Sediment microbial communities in Great Boiling Spring are controlled by temperature and distinct from water communities. ISME J 7:718–729
Colman DR, Thomas R, Maas KR, Takacs-Vesbach CD (2015) Detection and analysis of elusive members of a novel and diverse archaeal community within a thermal spring streamer consortium. Extremophiles 19:307–313
de la Torre JR, Walker CB, Ingalls AE, Konneke M, Stahl DA (2008) Cultivation of a thermophilic ammonia oxidizing archaeon synthesizing crenarchaeol. Environ Microbiol 10:810–818
Deppenmeier U (2002) Redox-driven proton translocation in methanogenic Archaea. Cell Mol Life Sci 59:1513–1533
Dequiedt S, Thioulouse J, Jolivet C, Saby N, Lelievre M, Maron PA, Martin MP, Prévost-Bouré NC, Toutain B, Arrouays D (2009) Biogeographical patterns of soil bacterial communities. Environ Microbiol Rep 1:251–255
DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol 72:5069–5072
Dos Santos PC, Fang Z, Mason SW, Setubal JC, Dixon R (2012) Distribution of nitrogen fixation and nitrogenase-like sequences amongst microbial genomes. BMC Genom 13:162
Drobner E, Huber H, Rachel R, Stetter KO (1992) *Thiobacillus plumbophilus* spec. nov., a novel galena and hydrogen oxidizer. Arch Microbiol 157:213–217
Dunfield PF, Tamas I, Lee KC, Morgan XC, McDonald IR, Stott MB (2012) Electing a candidate: a speculative history of the bacterial phylum OP10. Environ Microbiol 14:3069–3080
Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26:2460–2461
Engel AS, Johnson LR, Porter ML (2013) Arsenite oxidase gene diversity among Chloroflexi and Proteobacteria from El Tatio Geyser Field, Chile. FEMS Microbiol Ecol 83:745–756
Everroad RC, Otaki H, Matsuura K, Haruta S (2012) Diversification of bacterial community composition along a temperature gradient at a thermal spring. Microbes Environ 27:374–381
Fournier R (1989) Geochemistry and dynamics of the Yellowstone National Park hydrothermal system. Annu Rev Earth Planet Sci 17:13–53
Fournier RO (2005) Geochemistry and dynamics of the Yellowstone National Park hydrothermal system. In: Inskeep WP, TR M (eds) Geothermal biology and geochemistry in Yellowstone National Park. Thermal Biology Institute, Montana State University, Bozeman, pp 4–30
Fournier RO, Weltman U, Counce D, White L, Janik C (2002) Results of weekly chemical and isotopic monitoring of selected springs in Norris Geyser Basin, Yellowstone National Park. US Geological Survey
Garcia-Horsman JA, Barquera B, Rumbley J, Ma J, Gennis RB (1994) The superfamily of heme-copper respiratory oxidases. J Bacteriol 176:5587–5600
Glass EM, Wilkening J, Wilke A, Antonopoulos D, Meyer F (2010) Using the metagenomics RAST server (MG-RAST) for analyzing shotgun metagenomes. Cold Spring Harb Protoc 2010:pdb prot5368
Gonzalez P, Correia C, Moura I, Brondino C, Moura J (2006) Bacterial nitrate reductases: molecular and biological aspects of nitrate reduction. J Inorg Biochem 100:1015–1023
Good IJ (1953) The population frequencies of species and the estimation of population parameters. Biometrika 40:237–264
Guo Q, Kirk Nordstrom D, Blaine McCleskey R (2014) Towards understanding the puzzling lack of acid geothermal springs in Tibet (China): insight from a comparison with Yellowstone (USA) and some active volcanic hydrothermal systems. J Volcanol Geoth Res 288:94–104
Hall JR, Mitchell KR, Jackson-Weaver O, Kooser AS, Cron BR, Crossey LJ, Takacs-Vesbach CD (2008) Molecular characterization of the diversity and distribution of a thermal spring microbial community by using rRNA and metabolic genes. Appl Environ Microbiol 74:4910–4922
Hamilton TL, Vogl K, Bryant DA, Boyd ES, Peters JW (2012) Environmental constraints defining the distribution, composition, and evolution of chlorophototrophs in thermal features of Yellowstone National Park. Geobiology 10:236–249
Holloway JM, Nordstrom DK, Bohlke JK, McCleskey RB, Ball JW (2011) Ammonium in thermal waters of Yellowstone National Park: processes affecting speciation and isotope fractionation. Geochim Cosmochim Acta 75:4611–4636
Huber R, Stetter KO (2001) Aquificales. In: eLS. Wiley, New York
Hurwitz S, Evans W, Lowenstern J, Bergfeld D, Werner C, Heasler H, Jaworowski C (2007) Extensive hydrothermal rock alteration in a low pH, steam-heated environment: Hot Springs Basin, Yellowstone National Park. In: Proceedings of the water–rock interaction symposium-12, Kunming, pp 81–85
Huson DH, Auch AF, Qi J, Schuster SC (2007) MEGAN analysis of metagenomic data. Genome Res 17:377–386
Inskeep WP, Rusch DB, Jay ZJ, Herrgard MJ, Kozubal MA, Richardson TH, Macur RE, Hamamura N, Jennings R, Fouke BW, Reysenbach AL, Roberto F, Young M, Schwartz A, Boyd ES, Badger JH, Mathur EJ, Ortmann AC, Bateson M, Geesey G, Frazier M (2010) Metagenomes from high-temperature chemotrophic systems reveal geochemical controls on microbial community structure and function. PLoS One 5:e9773
Inskeep WP, Jay ZJ, Herrgard MJ, Kozubal MA, Rusch DB, Tringe SG, Macur RE, Jennings R, Boyd ES, Spear JR, Roberto FF (2013a) Phylogenetic and functional analysis of metagenome sequence from high-temperature archaeal habitats demonstrate Linkages between metabolic potential and geochemistry. Front Microbiol 4:95
Inskeep WP, Jay ZJ, Tringe SG, Herrgard MJ, Rusch DB (2013b) The YNP metagenome project: environmental parameters responsible for microbial distribution in the Yellowstone geothermal ecosystem. Front Microbiol 4:67
Itoh T, K-i Suzuki, Sanchez PC, Nakase T (1999) *Caldivirga maquilingensis* gen. nov., sp. nov., a new genus of rod-shaped crenarchaeote isolated from a hot spring in the Philippines. Int J Syst Bacteriol 49:1157–1163

Jackson CR, Langner HW, Donahoe-Christiansen J, Inskeep WP, McDermott TR (2001) Molecular analysis of microbial community structure in an arsenite-oxidizing acidic thermal spring. Environ Microbiol 3:532–542

Jimenez DJ, Andreote FD, Chaves D, Montana JS, Osorio-Forero C, Junca H, Zambrano MM, Baena S (2012) Structural and functional insights from the metagenome of an acidic hot spring microbial planktonic community in the Colombian Andes. PLoS One 7:e52069

Klatt CG, Wood JM, Rusch DB, Bateson MM, Hamamura N, Heidelberg JF, Grossman AR, Bhaya D, Cohan FM, Kuhl M, Bryant DA, Ward DM (2011) Community ecology of hot spring cyanobacterial mats: predominant populations and their functional potential. ISME J 5:1262–1278

Law CJ, Maloney PC, Wang DN (2008) Ins and outs of major facilitator superfamily antiporters. Annu Rev Microbiol 62:289–305

Madigan MT (2003) Anoxygenic phototrophic bacteria from extreme environments. Photosynth Res 76:157–171

Maechler M, Rousseeuw P, Struyf A, Hubert M, Hornik K (2012) Cluster: cluster analysis basics and extensions. R Package version 1:56

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437:376–380

Markowitz VM, Ivanova NN, Szeto E, Palaniappan K, Chu K, Dalevi D, Chen IM, Grechkin Y, Dubchak I, Anderson I, Lykidis A, Mavromatis K, Hugenholtz P, Kyrpides NC (2008) IMG/M: a data management and analysis system for metagenomes. Nucleic Acids Res 36:D534–D538

McCleskey RB, Ball JW, Nordstrom DK, Holloway JM, Taylor HE (2005) Water-chemistry data for selected hot springs, geysers, and streams in Yellowstone National Park, Wyoming 2001–2002. US Geological Survey Open-File Report 2004-1316

Meyer B, Kuever J (2008) Homology modeling of dissimilatory APS reductases (AprBA) of sulfur-oxidizing and sulfate-reducing prokaryotes. PLoS One 3:e1514

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

Meyer-Dombard DR, Swingley W, Raymond J, Havig J, Shock EL, Summons RE (2011) Hydrothermal ecotones and streamer biofilm communities in the Lower Geyser Basin, Yellowstone National Park. Environ Microbiol 13:2216–2231

Miller SR, Strong AL, Jones KL, Ungerer MC (2009) Bar-coded pyrosequencing reveals shared bacterial community properties along the temperature gradients of two alkaline hot springs in Yellowstone National Park. Appl Environ Microbiol 75:4565–4572

Mitchell K, Takacs-Vesbach C (2008) A comparison of methods for total community DNA preservation and extraction from various thermal environments. J Ind Microbiol Biotechnol 35:1139–1147

Moreno-Vivián C, Cabello P, Martínez-Luque M, Blasco R, Castillo F (1999) Prokaryotic nitrate reduction: molecular properties and functional distinction among bacterial nitrate reductases. J Bacteriol 181:6573–6584

Morgan L, Pierce K, McIntosh W (2005) Patterns of rhyolitic volcanism along the track of the Yellowstone hotspot. Geochim Cosmochim Acta Suppl 69:142

Nordstrom DK, McCleskey RB, Ball JW (2009) Sulfur geochemistry of hydrothermal waters in Yellowstone National Park: IV acid–sulfate waters. Appl Geochem 24:191–207

Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M (1999) KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res 27:29–34

Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara R, Simpson GL, Solymos P, Stevens MHH, Wagner H (2012) Vegan: community ecology package. R package version:2.0-5

Piddock LJ (2006) Multidrug-resistance efflux pumps—not just for resistance. Nat Rev Microbiol 4:629–636

Quince C, Lanzen A, Davenport RJ, Turnbaugh PJ (2011) Removing noise from pyrosequenced amplicons. BMC Bioinform 28(12):38

Reysenbach AL, Wickham GS, Pace NR (1994) Phylogenetic analysis of the hyperthermophilic pink filament community in Octopus Spring, Yellowstone National Park. Appl Environ Microbiol 60:2113–2119

Reysenbach AL, Ehringer H, Hershberger K (2000) Microbial diversity at 83 °C in Calcite Springs, Yellowstone National Park: another environment where the *Aquificales* and "Korarchaeota" coexist. Extremophiles 4:61–67

Reysenbach A-L, Banta A, Daly SC, Mitchel K, Lalonde S, Konhauser K, Rodman A, Rusterholtz K, Takacs-Vesbach C (2005) Aquificales in Yellowstone National Park. In: Inskeep WP, McDermott TR (eds) Geothermal biology and geochemistry in YNP. Montana State University, Thermal Biology Institute, Bozeman, pp 129–142

Rhoads DD, Wolcott RD, Sun Y, Dowd SE (2012) Comparison of culture and molecular identification of bacteria in chronic wounds. Int J Mol Sci 13:2535–2550

Richardson DJ, Berks BC, Russell DA, Spiro S, Taylor CJ (2001) Functional, biochemical and genetic diversity of prokaryotic nitrate reductases. Cell Mol Life Sci 58:165–178

Rothschild LJ, Mancinelli RL (2001) Life in extreme environments. Nature 409:1092–1101

Rye RO, Truesdell AH (2007) The question of recharge to the deep thermal reservoir underlying the geysers and hot springs of Yellowstone National Park. US Geol Surv Prof Pap 1717:1–35

Sen R, Maiti N (2014) Genomic and functional diversity of bacteria isolated from hot springs in Odisha, India. Geomicrobiol J 31:541–550

Shock EL, Holland M, Meyer-Dombard DA, Amend JP, Osburn GR, Fischer TP (2010) Quantifying inorganic sources of geochemical energy in hydrothermal ecosystems, Yellowstone National Park, USA. Geochim Cosmochim Acta 74:4005–4043

Song Z-Q, Wang F-P, Zhi X-Y, Chen J-Q, Zhou E-M, Liang F, Xiao X, Tang S-K, Jiang H-C, Zhang CL, Dong H, Li W-J (2013) Bacterial and archaeal diversities in Yunnan and Tibetan hot springs, China. Environ Microbiol 15:1160–1175

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

Swingley WD, Meyer-Dombard DR, Shock EL, Alsop EB, Falenski HD, Havig JR, Raymond J (2012) Coordinating environmental genomics and geochemistry reveals metabolic transitions in a hot spring ecosystem. PLoS One 7:e38108

Tatusov RL, Galperin MY, Natale DA, Koonin EV (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res 28:33–36

Team RDC (2011) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna

Tiedje JM, Sexstone AJ, Myrold DD, Robinson JA (1983) Denitrification: ecological niches, competition and survival. Antonie Van Leeuwenhoek 48:569–583

Truesdell AH, Fournier RO (1976) Conditions in the deeper parts of the hot spring systems of Yellowstone National Park. Wyoming, US Geological Survey

Truesdell AH, Nathenson M, Rye RO (1977) The effects of subsurface boiling and dilution on the isotopic compositions of Yellowstone thermal waters. J Geophys Res 82:3694–3704

van den Berg EM, van Dongen U, Abbas B, van Loosdrecht MC (2015) Enrichment of DNRA bacteria in a continuous culture. ISME J 9:2153–2161

Van Horn DJ, Van Horn ML, Barrett JE, Gooseff MN, Altrichter AE, Geyer KM, Zeglin LH, Takacs-Vesbach CD (2013) Factors controlling soil microbial biomass and bacterial diversity and community composition in a cold desert ecosystem: role of geographic scale. PLoS One 8:e66103

Wang Q, Garrity GM, 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

Wang S, Hou W, Dong H, Jiang H, Huang L, Wu G, Zhang C, Song Z, Zhang Y, Ren H, Zhang J, Zhang L (2013) Control of temperature on microbial community structure in hot springs of the Tibetan Plateau. PLoS One 8:e62901

Ward DM, Ferris MJ, Nold SC, Bateson MM (1998a) A natural view of microbial biodiversity within hot spring cyanobacterial mat communities. Microbiol Mol Biol Rev 62:1353–1370

Ward DM, Ferris MJ, Nold SC, Bateson MM (1998b) A natural view of microbial biodiversity within hot spring cyanobacterial mat communities. Microbiol Mol Biol Rev 62:1353–1370

White DE (1957) Thermal waters of volcanic origin. Geol Soc Am Bull 68:1637–1658

Xie W, Zhang CL, Wang J, Chen Y, Zhu Y, de la Torre JR, Dong H, Hartnett HE, Hedlund BP, Klotz MG (2015) Distribution of ether lipids and composition of the archaeal community in terrestrial geothermal springs: impact of environmental variables. Environ Microbiol 17:1600–1614

Zhang CL, Ye Q, Huang Z, Li W, Chen J, Song Z, Zhao W, Bagwell C, Inskeep WP, Ross C, Gao L, Wiegel J, Romanek CS, Shock EL, Hedlund BP (2008) Global occurrence of archaeal amoA genes in terrestrial hot springs. Appl Environ Microbiol 74:6417–6426