environmental microbiology

Environmental Microbiology (2020) **00**(00), 00–00 doi:10.1111/1462-2920.15373

Time series metagenomic sampling of the Thermopyles, Greece, geothermal springs reveals stable microbial communities dominated by novel sulfur-oxidizing chemoautotrophs

A. Meziti,[1,2*] **E. Nikouli**,[1,2] **J. K. Hatt**,[2] **K. T. Konstantinidis**[2,3] **and K. A. Kormas**[1]

[1]*Department of Ichthyology and Aquatic Environment, University of Thessaly, Volos, 38446, Greece.*

[2]*School of Civil and Environmental Engineering, Georgia Institute of Technology, Ford Environmental Science and Technology Building, 311 Ferst Drive, Atlanta, GA, 30332.*

[3]*School of Biological Sciences, Georgia Institute of Technology, Ford Environmental Sciences and Technology Building, 311 Ferst Drive, Atlanta, GA, 30332.*

Summary

Geothermal springs are essentially unaffected by environmental conditions aboveground as they are continuously supplied with subsurface water with little variability in chemistry. Therefore, changes in their microbial community composition and function, especially over a long period, are expected to be limited but this assumption has not yet been rigorously tested. Toward closing this knowledge gap, we applied whole metagenome sequencing to 17 water samples collected between 2010 and 2016 from the Thermopyles sulfur-rich geothermal springs in central Greece. As revealed by 16S rRNA gene fragments recovered in the metagenomes, *Epsilonproteobacteria*-related operational taxonomic units (OTUs) dominated most samples and grouping of samples based on OTU abundances exhibited no apparent seasonal pattern. Similarities between samples regarding functional gene content were high, with all samples sharing >70% similarity in functional pathways. These community-wide patterns were further confirmed by analysis of metagenome-assembled genomes (MAGs), which showed that novel species and genera of the chemoautotrophic *Campylobacterales* order dominated the springs. These MAGs carried different pathways for thiosulfate or sulfide oxidation coupled to carbon fixation pathways. Overall, our study showed that even in the long term, functions of microbial communities in a moderately hot terrestrial spring remain stable, presumably driving the corresponding stability in community structure.

Introduction

Microbial communities of geothermal springs are of special interest for understanding early life as they are considered analogues of the first habitable environments on Earth (Konhauser *et al*., 2001; Damer and Deamer, 2019). These habitats are dominated by several thermophilic and hyperthermophilic species (López-López *et al*., 2013) that mediate nutrient cycling (e.g. carbon, nitrogen, and sulfur) (Falkowski *et al*., 2008). The best-studied geothermal habitat is Yellowstone National Park (YNP; USA) which contains >14 000 of sites with geothermal activity, covering a wide range of pH (2–10), temperature (40–92°C) and geochemical properties (Nordstrom *et al*., 2005; Rye and Truesdell, 2007; McCleskey *et al*., 2010). Most of these sites exhibit temporal stability in these geochemical properties (Inskeep *et al*., 2013).

Several molecular studies of the microbial diversity present in YNP hot springs have revealed the dominance of uncultivated thermophilic archaea, *Aquificales*, *Chloroflexi*, *Chlorobi* and *Cyanobacteria* at different sites (Barns *et al*., 1994; Barns *et al*., 1996; Boomer *et al*., 2002; Hiras *et al*., 2016; Hugenholtz *et al*., 1998; Meyer-Dombard *et al*., 2005; Reysenbach *et al*., 2006; Spieck *et al*., 2020; Toplin *et al*., 2008). Subsequent studies in YNP, including a large metagenomic survey (Inskeep *et al*., 2013), tried to elucidate the geochemical features that drive microbial diversity patterns and how specific phylotypes are functionally differentiated from each other based on electron donors such as hydrogen or sulfide and electron acceptors. Other habitats that

Received 25 August, 2020; accepted 19 December, 2020. **For correspondence. E-mail ameziti@uth.gr; Tel. (+30) 242-109-3240; Fax (+30)242-109-3157.

have been studied, including thermophilic springs in Japan (Yamamoto *et al*., 1998; Iino *et al*., 2010), Iceland (Flores *et al*., 2008; Takacs-Vesbach *et al*., 2008; Tobler and Benning, 2011), New Zealand (Childs *et al*., 2008) and Kamchatka (Russia; Burgess *et al*., 2012; Wilkins *et al*., 2019) were found to harbour similar microbial communities. Most of these habitats exhibit high temperatures (>65°C), and thus mainly consist of hyperthermophilic communities.

Mesophilic sulfur-rich terrestrial springs, in particular, are comparatively much less studied and understood. Chemoautotrophic microbial communities use reduced sulfur compounds contained in underlying geothermal water to gain energy and fuel these systems. These communities are important in sulfur cycling and carbon fixation, thus mediating the transfer of energy from the geothermal source to other trophic levels (Hügler *et al*., 2010). Most studies of mesophilic terrestrial springs have analysed microbial community diversity and functions in order to detect key community members and functions performed in association with environmental or geothermal parameters (Engel *et al*., 2004; Reigstad *et al*., 2011; Headd and Engel, 2013; Chan *et al*., 2015). However, although steady geochemical conditions imply stability in microbial diversity and function, no study of mesophilic terrestrial springs has been performed on a temporal or seasonal scale. Temporal analysis is particularly important for mesophilic springs since these systems typically receive soil or water inputs from other sources (i.e. seawater or rainfall), in addition to groundwater. It currently remains unknown whether these inputs could influence major metabolic pathways and override the major input from geothermal fluids.

Greece harbours many terrestrial springs due to the geology of the country. The formation of terrestrial springs is related to recent volcanic activity and active tectonics for which magmatic and volcanic processes along with the high mountain chains and active fault systems favour the rise of deep waters discharged at the surface as geothermal springs (Lambrakis and Kallergis, 2005). Most springs are characterized by the mixing of deep thermal reservoir water with meteoric water, while coastal springs are also characterized by loss of heat via the mixing of geothermal water with seawater and freshwater.

The Thermopyles springs located in the eastern part of mainland Greece (38°47′36.35″N/22°31′43.04″E) consist of such typical coastal springs and comprise one of the larger active geothermal systems in Greece (Fig. S1). They are part of the Spercheios tectonic graben, which is considered to be an extension of the Anatolia strike-slip fault (Georgalas and Papakis, 1966; Marinos *et al*., 1973). The activity of trending faults contributes to the uprise of thermal water, which specifically for the Thermopyles springs mixes with seawater or freshwater resulting in lower temperatures (~40°C) compared with hyperthermophilic geothermal springs and pH values close to 6 (Duriez *et al*., 2008). The system is rich in ammonia and hydrogen sulfide produced by the reduction of sulfate originating from the oxidation of sulfide minerals or directly from seawater (Lambrakis *et al*., 2014).

Very few studies have been conducted in the Thermopyles and these were focused on geochemical features and measurements of physicochemical properties (Lambrakis and Kallergis, 2005; Verros *et al*., 2007; Duriez *et al*., 2008; Lambrakis *et al*., 2014; Zarikas *et al*., 2014) or the investigation of travertine deposits in association with *Cyanobacteria* (Kanellopoulos *et al*., 2016). Investigation of geochemical features has shown the increased concentration of hydrogen sulfide mainly produced by pyrite oxidation and reduced species of sulfates (Lambrakis and Kallergis, 2005; Duriez *et al*., 2008). The microbial community of the Thermopyles has only been studied once (Kormas *et al*., 2009) but this previous study was focused on a cross-sectional comparison of populations to those in springs in other parts of Greece and was based on SSU rRNA gene clone libraries that provided only coarse resolution.

Here we analyse shotgun metagenomes from 17 Thermopyles samples collected during a 7-year period and in different seasons along with environmental data to evaluate (i) potential seasonal changes in microbial diversity and function, (ii) the novelty of the microbial species that are always prevalent in the community using 16S rRNA and metagenome assembled genomes (MAGs) and (iii) key metabolic functions that fuel the microbial communities in terms of energy and carbon sources for further biotechnological applications. Our findings suggest that seasonal and temporal changes were not significant in shaping microbial community composition and function while microbial communities were characterized by several novel species that have a key role in habitat function.

Results

Microbial community structure of thermopyles

In the metagenomic datasets, a total of 887 602 individual reads encoding fragments of the 16S rRNA gene were detected (Table S1). [Sample naming scheme: the two letters and the two numbers reflect the month and the year in which the sample was collected. For example, AU13 represents a sample collected in August 2013]. Bacterial sequences predominated since only 0.05% of the total 16S rRNA gene-carrying reads were assigned to *Archaea*. The highest diversity as indicated by the Shannon index of 16S rRNA gene-based operational

taxonomic units (or OTUs) was observed in the DE14 sample (5.7) while the lowest value was observed for DE16 (3.61; Table S1). In terms of Chao richness, the highest value was observed in MY16 (2875 OTUs) and the lowest in FE15 (1136 OTUs). Good's coverage values exceeded 0.98 for all samples indicating that approximately 98% of the total OTUs were recovered by sequencing. In total 5956 OTUs were observed in all time points. Only 66% of these OTUs had a >97% identity match against the SILVA database; the remaining OTUs represented 'novel' taxa at this level (Table S1). These novel OTUs comprised a substantial portion of the total community, accounting for 16%–40% of the total reads, depending on the sample considered (Table S1).

At the whole-genome level, the majority (60.9%–93.5%) of the metagenomic reads assembled in contigs longer than 500 bp (Table S1). The coverage of the microbial community achieved by the corresponding metagenomic dataset was also estimated based on the redundancy of the reads using the Nonpareil algorithm (Rodriguez-R and Konstantinidis, 2014), which confirmed the relatively low complexity of all samples as evidenced by coverage values ranging from 0.77 to 0.94 (Table 1).

The taxonomic composition was estimated based on the short reads encoding 16S rRNA gene fragments recovered in the metagenomes. Proteobacteria were the most abundant phylum across all samples (Fig. S2). Class level taxonomic distributions revealed the dominance of *Epsilonproteobacteria* (>60% of total sequences) in most samples, usually followed by *Gamma-* or *Alphaproteobacteria* (Fig. 1A). [Note that the classification of *Epsilonproteobacteria* has been the target of several recent studies and the proposal of a separate phylum, Epsilonbacteraeota, has been suggested (Waite *et al*., 2017). Since then, different databases (e.g. SILVA) have followed this suggestion while others have not (e.g. GenBank). In this study, different databases were used for classification, but we chose to use the GenBank classification of *Epsiloproteobacteria* as a Class within the Proteobacteria phylum for consistency with most earlier studies].

Out of the 5956 OTUs observed in all time points, 165 represented shared OTUs that were present in all samples, comprising 18.14%–91.64% (average 74.09%) of the total 16S rRNA gene carrying reads per sample (Table S1). Given the similar sequencing depth across the samples (Table 1, S1) and overall high sequence coverage (Table 1), the comparison of the number of detected shared OTUs can provide a reliable picture of the size and composition of the 'core' bacterial community in the springs. This core bacterial community mainly consisted of representatives of the *Piscirickettsiaceae*, *Sulfurovaceae* and *Campylobacteraceae* families, and the *Halothiobacillus*, *Sulfurovum*, *Arcobacter*, *Sulfuricurvum* and *Sulfurimonas* genera. The relative abundance of these core taxa did not exhibit any apparent seasonal pattern (Fig. 1C).

More specifically, cluster analysis performed on the similarities between the abundance distributions of all identified OTUs in each sample (Morisita similarity; Wolda, 1981) exhibited lack of seasonal patterns (seasons were defined using either calendar days or climatological parameters). Four major clusters (A–D) were observed, exhibiting high intra-cluster similarities

Table 1. Metagenomic dataset statistics associated with Thermopyles dataset.

WGS dataset statistics	AP10	JN10	JN11	DE11	MY12	DE12	AU13	DE13	MY14
Total size (Gb)	3.47	3.31	3.54	2.24	2.37	3.17	3.02	3.84	2.82
# Predicted genes	44 818	92 624	40 118	49 796	39 349	45 342	37 093	54 317	45 364
# Contigs	12 590	23 409	9720	17 517	11 372	11 625	11 178	16 985	10 063
Average contig length	2867	3277	3512	2319	2842	3242	2611	2566	3802
Longer contig (bp)	186 023	250 868	138 677	104 928	186 610	239 545	204 647	240 475	240 302
Coverage (Nonpareil)	0.85	0.77	0.92	0.84	0.81	0.86	0.88	0.85	0.93
N50	3691	4791	5851	2521	3874	4658	3132	3067	7823
% Reads mapping on abundant MAGs genes	36.16	20.60	38.01	21.79	32.99	37.40	38.83	34.87	36.06

WGS dataset statistics	DE14	FE15	AP15	JN15	DE15	MY16	OC16	DE16
Total size (Gb)	3.35	1.70	3.34	3.24	3.28	4.11	2.71	3.48
# Predicted genes	112 357	16 152	31 367	102 440	61 816	80 003	35 720	29 561
# Contigs	21 625	4931	8295	24 137	14 613	22 072	8128	8011
Average contig length	4693	2591	3118	3603	3519	2988	3731	2994
Longer contig (bp)	260 721	240 237	229 951	269 767	240 362	240 890	691 025	164 877
Coverage (Nonpareil)	0.86	0.95	0.85	0.88	0.92	0.87	0.94	0.93
N50	10 965	2833	5336	5558	5519	4062	7697	4228
% Reads mapping on abundant MAGs genes	17.01	45.62	34.44	25.59	35.64	32.35	38.99	38.92

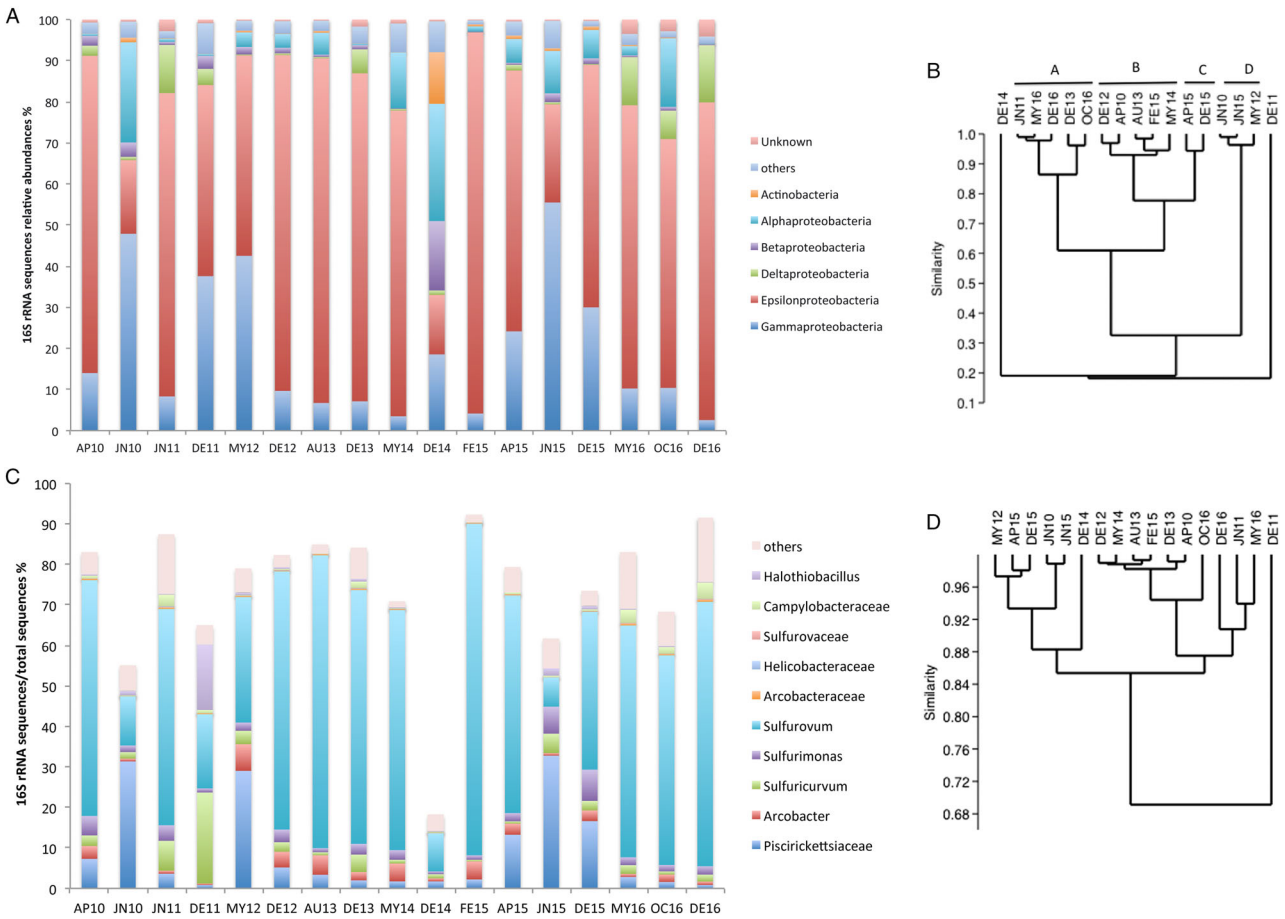

Fig 1. Taxonomic composition and cluster analysis of Thermopyles metagenomic datasets.
A. Taxonomic composition (class level) of the Thermopyles geothermal springs microbial communities.
B. Clustering of the Thermopyles samples based on 16S rRNA gene-based OTU (97% identity threshold) abundance profiles (Morisita similarities).
C. Taxonomic composition of 'core' OTUs Thermopyles geothermal springs microbial communities.
D. Clustering of the Thermopyles samples based on seed subsystems abundance profiles (Morisita similarities).

(e.g. >85%) and including samples from different years (e.g. JN11 vs. DE16; group A or JN10 vs. JN15; group D), while only two samples (DE11, DE14) were <20% similar to all the remaining samples (Fig. 1B). Furthermore, three clusters (A–C) exhibited inter-cluster similarities >60% and they were dominated by *Sulfurovum* sp.-related OTUs, implying the presence of a relatively stable community across all of our samples (i.e. AP10-OC16), occasionally interrupted by *Piscirickettsiaceae*-dominated communities or the two outliers (Fig. 1B and C). Apart from similarities in terms of taxonomic diversity, the size of the microbial community remained stable during the 7 years as evidenced by DAPI counts ranging from 85 434 to 98 888 cells ml^{-1} (Table S1).

Consistent with the results reported above based on individual OTUs, our non-metric multidimensional scaling (NMDS) (stress: 12.57%) analysis performed at the whole community level based on the normalized abundances for metagenome size of all identified OTUs revealed no measured environmental parameters (e.g. season, precipitation, pH) to be significant for the ordination of the samples (Fig. S3).

Microbial functional diversity and comparisons to other habitats

The total assembly size for each metagenome (no co-assembly was performed) ranged from 1.70 Gbp in FE15 to 4.11 Gbp in MY16, while N50 values ranged from 2521 bp in DE11 to 10 965 bp in DE14. Average contig length ranged from 2318 bp (DE11) to 4692 bp (DE14), while the largest contigs observed ranged from 104 928 bp (DE11) to 691 025 bp (OC16) (Table 1). The number of predicted genes from the metagenomic assemblies ranged from 16 152(FE15) to 112 357

(DE14) (Table 1), reflecting the underlying microbial community diversity of the samples.

Functional gene distributions were evaluated based on the classification of genes in the subsystems hierarchical annotation scheme of the SEED database, resulting in 13.94% (DE11) to 54.16% (AG13) of total reads mapping on assembled genes that have SEED subsystems annotations (Table S1). Cluster analysis performed on the abundance distributions of different functions revealed higher Morisita similarities than the taxonomic (OTUs) similarities mentioned above, suggesting a greater number of shared functions among all samples (Fig. 1D). Consistent with the taxonomic results, DE11 exhibited the lowest functional similarities with the rest of the samples (<70%), while DE14 was more similar with the rest of the samples compared with the taxonomic diversity comparisons described above (Fig. 1B and D). Pairwise comparisons, using DeSeq, of the functional distributions between all Thermopyles samples revealed relatively similar gene content with none of the 1114 identified subsystems exhibiting significantly different abundances ($p_{adj} > 0.05$). However, when using *p*-values without correction for multiple samples, DE11 exhibited higher abundances ($p < 0.05$) of CO_2 fixation, metabolism of aromatic compounds, ammonia assimilation, phages prophages and transposable elements, sulfate reduction and CRISPRs-related genes, and relatively lower abundance of denitrification, dissimilatory nitrite reduction and nitrate/nitrite ammonification compared with the other metagenomes (Fig. 2A). In general, these functions did not exhibit any specific pattern regarding season or month, similar to the OTU patterns mentioned above.

When comparisons were performed with other habitats such as Kalamas River in Greece (sample kal2feb; Meziti *et al*., 2016) and YNP in the USA (sample CIS_19; Inskeep *et al*., 2013), several significant differences ($p_{adj} < 0.05$) were observed in functional gene diversity (Table S2, S3, Fig. 2B). Overall, Thermopyles samples exhibited increased allelic diversity (i.e. a higher number of distinct variants of the same protein function at the 97% amino-acid similarity threshold) compared with both Kalamas and YNP for all genes involved in sulfur oxidation pathways including all genes from the Sox pathway (*soxABCDXYZ*) as well as genes for flagellar motility, potassium homeostasis (potassium channel and efflux system proteins genes) and the carboxysome (Rubisco operon transcription regulators and activation proteins involved in Calvin–Benson–Basham [CBB] cycle for CO_2 fixation).

Flagellar motility genes are generally increased in springs relative to riverine samples since microorganisms use motility to find the niche with the most favourable concentrations of oxygen and nutrients. The relatively lower frequency of flagellar motility genes found in the YNP samples could be attributed to the decreased water flow (i.e. samples were collected mostly from bottom or suspended sediment) in this ecosystem versus the samples from Thermopyles. Taxonomic analysis of the carboxysome-related genes showed that they were classified in the *Piscirickettsiaceae* family (*Gammaproteobacteria*), which also included many of the core OTUs (Figs 1A and C). Indeed, it has been shown that the CBB cycle is a common CO_2 fixation pathway for gammaproteobacterial autotrophs (Mangiapia and Scott, 2016); this pathway was not present in the YNP sample analysed here. Similarly, heat shock (*dna*K) genes were increased in Thermopyles populations compared with YNP. *Dna*K is mainly found in *Bacteria* while it is detected only in some groups of *Archaea* via lateral gene transfer events (Gribaldo *et al*., 1999), explaining the low abundance of *dna*K genes in the archaeal-dominated YNP site compared with Thermopyles.

Gene (allelic) diversity for CRISPR genes as well as ferredoxin oxidoreductase homologues, which are important in the reverse TCA cycle, often observed in sulfidic and anaerobic environments dominated by *Epsilonproteobacteria* and *Archaea* (Fig. 2B) was increased in both Thermopyles and YNP compared with Kalamas. Diversity in CRISPR-related genes was much higher in Thermopyles where 160 different CRISPR genes were detected in total including genes related to cas1-9, Cmr3/4/6 and Csd1/2/5/7 gene families, whereas in Kalamas only the Cas2-related proteins were detected.

Sulfur and nitrogen metabolism are of special interest for the microbial ecology of Thermopyles due to the chemical composition of the groundwater feed (Lambrakis *et al*., 2014), and thus the corresponding pathways were examined in more detail using KEGG annotations. For sulfur metabolism, three pathways were assessed, i.e. M00595 for thiosulfate oxidation by the Sox pathway (*soxABCDXYZ*), M00596 for dissimilatory sulfate reduction (DSR) and M00176 for assimilatory sulfate reduction (ASR) (Table S4). Most samples contained all genes of the Sox pathway and partial detection was observed in the rest, similar to the ASR pathway (Fig. 3B; Table S4). The complete DSR pathway was detected in eight samples, whereas partial detection was observed in three samples and complete absence found in the rest. For the samples in which partial detection was observed, the calculation of genome equivalents (i.e. what fraction of total cells encode the gene of interest assuming a single copy per genome) showed that absence was probably due to low coverage rather than true absence. Genome equivalent analysis also showed higher abundance (3- to 50-fold) of *soxCDY* genes relative to the rest of the genes of the pathway (Table S5), implying that members of the community could follow an alternative pathway for sulfur oxidation with the absence of *soxABZX* as suggested previously for other systems (Lahme *et al*., 2020).

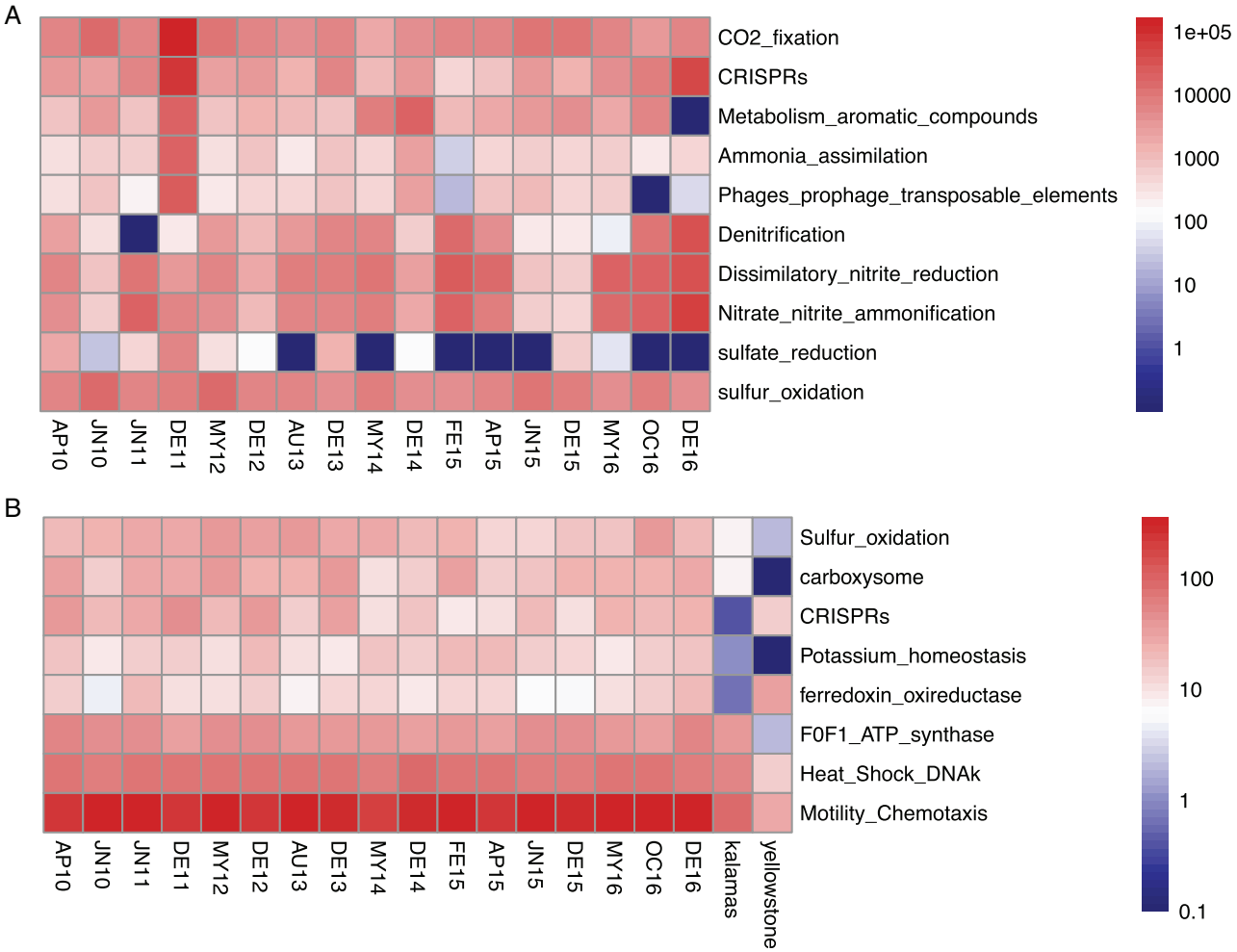

Fig 2. Functional profiles of the Thermopyles microbial communities.
A. Subsystems (SEED database) with differences in abundance through time based on negative binomial test using DeSeq and $p < 0.05$ (not-adjusted). Scale corresponds to log number of normalized reads.
B. Statistically significant differences in gene content (derived from the number of different genes that could be assigned to a subsystem) between Thermopyles, Yellowstone and Kalamas samples based on negative binomial test using DeSEQ and $p_{adj} < 0.05$ (p_{adj}-values were corrected for multiple testing; scale corresponds to normalized log number of different genes).

Regarding nitrogen metabolism, KEGG pathways predicted to be present in the Thermopyles samples included nitrogen fixation (*nif*; KEGG ID M00175), assimilatory nitrate reduction (ANR; KEGG ID M00531), as well as dissimilatory nitrate reduction (DNR; KEGG ID M00530) pathways. Nitrogen fixation genes, *nifHDK* and *vnfH* were present in all metagenomes except for FE15. The DNR pathway was complete in all metagenomes and ANR was complete in all metagenomes apart from MY14 and DE14 in which gene *nirA* for the transformation of nitrite to ammonia was absent (Fig. 3C).

Among the total SEED subsystems detected in at least one of the samples, 250 out of 1234 were common in all the habitats analysed and were present in eight of the Thermopyles samples. Genes coding for these proteins were considered homologous and were further analysed for their %G + C content. The %G + C content in Thermopyles (53.6%) was significantly higher and lower than the respective %G + C contents in Kalamas (47.2%) and YNP (66.3%) (*t*-test, $p < 0.0001$). This finding agreed with the hypothesis that %G + C content provides greater thermotolerance to microorganisms (Zheng and Wu, 2010) since the Kalamas site exhibits the lowest annual temperature (11°C) (Meziti *et al*., 2016) among the three sites, and YNP the highest (78–80°C) (Inskeep *et al*., 2013).

Novelty of the metagenome-assembled genomes. In total, 78 good quality MAGs, defined as completeness >70% and contamination (<10%), were recovered from the 17 samples. After ANI pairwise comparisons, the MAGs were grouped using a cutoff of 95% (threshold for

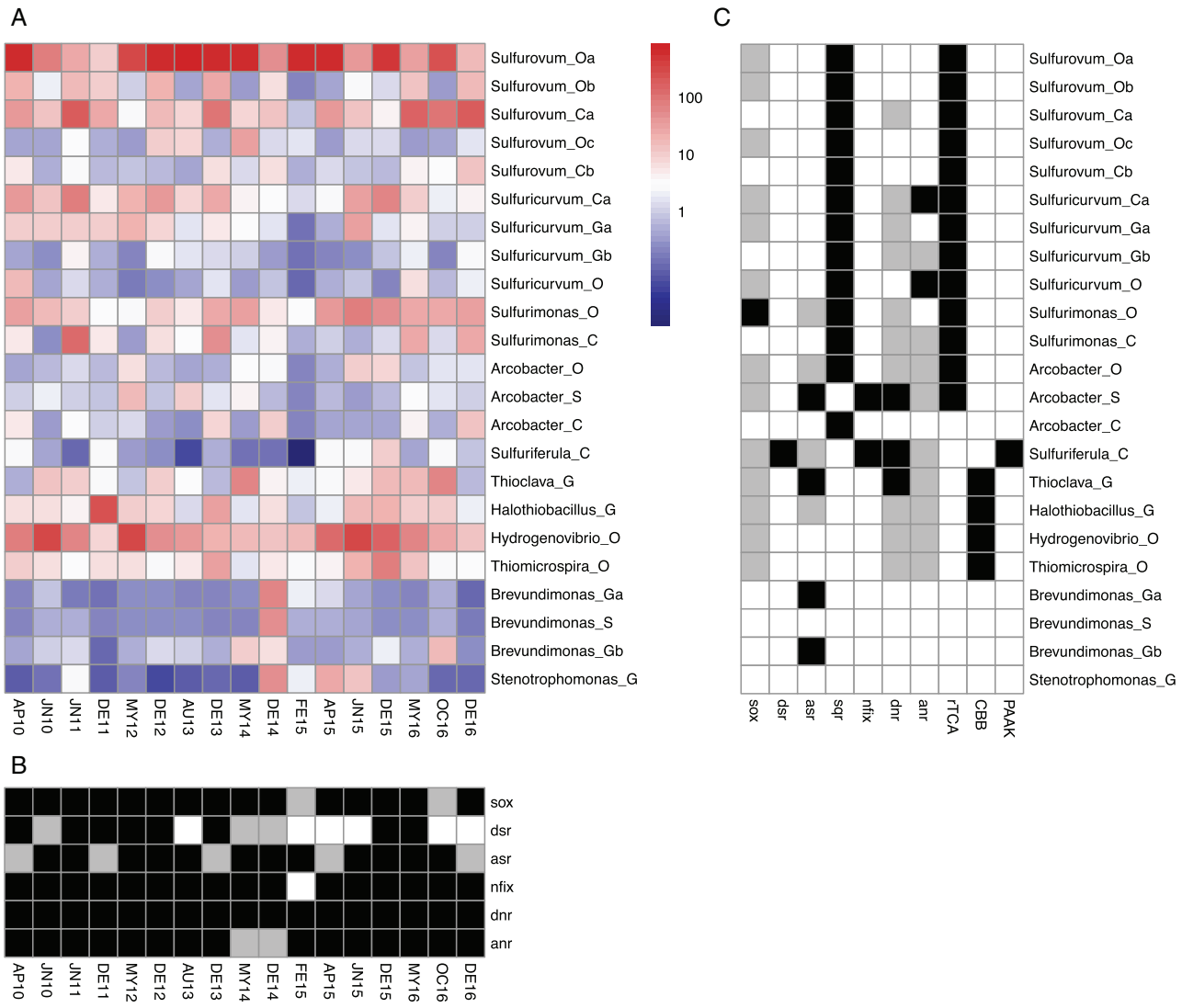

Fig 3. The abundance of MAGs and presence of specific pathways in metagenomes and MAGs.
A. Abundance of the identified bacterial populations (MAGs); scale: sequence depth coverage.
B. Presence/absence of specific pathways in different metagenomes.
C. Presence/absence of specific pathways in different MAGs.
DSR: dissimilatory sulfate reduction, ASR: assimilatory sulfate reduction, sqr: sulfide oxidoreductase, nifix: nitrogen fixation, DNR: dissimilatory nitrate reduction, ANR: assimilatory nitrate reduction, denitr: denitrification, rTCA: reductive TCA cycle, CBB: Calvin–Benson–Basham, PAAK: Phosphate acetyltransferase-acetate kinase pathway; black boxes stand for full pathway, grey boxes for partial pathway and white for absence.

species) in 42 genomospecies (GSP) for further analysis (Table S6). The best quality (i.e. highest completeness, lowest contamination) MAG of each genomospecies was used as a representative.

Our assessment using MiGA (Rodriguez-R *et al*., 2018) revealed that many of the MAGs (15/42) belonged to *Epsilonproteobacteria* followed by *Alphaproteobacteria* (10), *Gammaproteobacteria* (9), *Actinobacteria* (4), *Bacteroidetes* (2) and *Betaproteobacteria* (1) (Table S6) consistent with the 16S rRNA gene-based OTU results mentioned above (Fig. 1A and C). The majority of the *Epsilonproteobacteria* affiliated MAGs were only classified to the order or class level compared with all previously named species of isolates suggesting that they represent novel families, not previously classified and with no genome representatives. Results from GTDB-tK classifications for the *Epsilonproteobacteria* MAGs were consistent, showing that our MAGs probably represent novel genera or families (Table S6). This finding partially agreed with the 16S rRNA gene-based results that showed ~15% of the epsilonproteobacterial classified OTUs belong to novel genera and 8% of them to novel families (Table S1).

To better study the taxonomy and function of abundant MAGs, 23 MAGs exhibiting high relative abundance

(Table 1, S7) and potential involvement in the sulfur cycle based on their predicted gene content were chosen for further investigation (Table 2; Fig. 3A). [MAGs naming scheme: the first part reflects the closest relative of the MAG and the second part the lowest taxonomic rank the two share according to the MiGA TypeMAT/NCBI database ($p < 0.1$), i.e. C: class, O: order, F: family. G: Genus, S: Species. For instance, we use Sulfurimonas_O for a MAG that had a *Sulfurimonas* sp. as the closest relative and was classified - at the lowest level with statistical confidence - to the order *Campylobacterales*].

Five of these MAGs (Sulfurovum_X) belonged to closely related species exhibiting ANI similarities below 95% but above 72% among them (Table S8). The closest relative of these MAGs was *Sulfurovum lithotrophicum* NZ CP011308 (Jeon *et al*., 2017) exhibiting AAI similarities between 52.68% and 63.05% based on MiGA (Table 2); thus, they share only class or order with their closest relative and represent a novel genus, if not family. The *Sulfurovum* MAGs were dominant in most samples with coverage and relative abundances exceeding 100× and 20% of total community respectively (Fig. 3A; Table S7), in agreement with the 16S rRNA gene-based analysis of which OTUs represented the 'core' community (Fig. 1C), showing *Sulfurovum* and *Sulfurovaceae*-related OTUs to be highly abundant and core.

The rest of the MAGs were taxonomically classified ($p < 0.1$) to *Epsilonproteobacteria* or *Campylobacterales* having *Sulfuricurvum*, *Arcobacter* and *Sulfurimonas* genera as best matches with their closest relatives to be *Sulfuricurvum kujiense* NC 014762T (Kodama and Watanabe, 2004), *Arcobacter* sp. L NC 017192 (Toh *et al*., 2011) and *Sulfurimonas autotrophica* NC 014506T (Inagaki *et al*., 2003) respectively (Table 2). ANI values between MAGs with the same closest relative varied from 71.14% to 90.32%, implying that they most likely represented distinct species of the same family or genus. These findings highlighted the substantial intra-genus diversity at Thermopyles for the above-mentioned genera, similar to the *Sulfurovum* MAGs mentioned earlier (Table S8), and in agreement with the 16S rRNA gene-based core OTUs. The remaining MAGs were taxonomically classified to *Brevundimonas*, *Halothiobacillus*, *Thioclava*, *Stenotrophomonas*, *Thiotrhichales* and *Betaproteobacteria* (Table 2). Notably, taxonomic assignments and persistence over time of most MAGs (e.g. *Epsilonproteobacteria*, *Thiotrichales* and *Halothiobacillus*) corroborated with 16S rRNA gene data, further supporting the presence of a core community.

Key metabolic functions of the MAGs. The 23 abundant MAGs (Table 2) suspected to be involved in the sulfur cycle were examined for key energy pathways including pathways for sulfur and nitrogen cycling as well as for carbon fixation. In addition to the KEGG pathways for sulfur and nitrogen cycling mentioned above, pathways for CO_2 fixation, such as the reductive TCA cycle (M00173), phosphate acetyltransferase-acetate kinase pathway (M00579), the reductive pentose phosphate cycle (M00165)/CBB, the dark and light Crassulacean acid metabolism (M00168/9) and the reductive acetyl-CoA pathway (M00377) were searched along with sulfide quinone reductase (*sqr*; E.C.1.8.5.4) for the oxidation of sulfide to elemental sulfur.

Using a similar approach and thresholds as mentioned above for assembled contigs and genes, *sqr* was detected in all *Epsilonproteobacteria* classified MAGs indicating the potential for elemental sulfur formation. Genes for the oxidation of thiosulfate by the Sox pathway were partially observed in some *Sulfurovum*, *Sulfuricurvum*, *Arcobacter*, *Halothiobacillus*, *Thiomicrospira*, *Hydrogenovibrio* and *Thioclava* MAGs, while the complete complex was observed only in *Sulfurimonas*_O (Fig. 3C; Table S4), in agreement with previous studies (Han and Perner, 2015). The absence of some genes from the *sox* complex could be attributable to low completeness in some cases such as with the *Sulfurovum*_O, *Hydrogenovibrio*_O, *Thiomocrospira*_O and *Thioclava*_G MAGs. However, the absence of specific genes (*soxCD*) in the *Sulfuricurvum* MAGs could indicate alternative pathways as previously detected in *Sulfuricurvum* sp., in which thiosulfate oxidation ends up with the accumulation of sulfur globules or polysulfide (Friedrich *et al*., 2005; Frigaard and Dahl, 2009).

Regarding DSR, only *Sulfuriferula*_C possessed the complete pathway while no related genes were observed in the other MAGs. For ASR, the complete pathway was only detected in *Thioclava*, *Arcobacter*_S and *Brevundimonas*_Ga, whereas *Sulfurimonas*, *Arcobacter*, *Sulfuriferula*_C, *Halothiobacillus* and *Brevundimonas* MAGs possessed some genes of the pathway mostly limited to only the gene for the first step, that is the transformation of sulfate to adenosine 5′ phosphosulfate (APS) (Fig. 3B; Table S4). Almost all of the MAGs possessed the gene encoding the enzyme for sulfate adenylytransferase (*sat*), one of the enzymes responsible for the formation of APS in DSR and ASR, without possessing any other enzymes of the pathway. In these cases, both pathways were considered absent.

For nitrogen metabolism, the genes encoding the enzymes for nitrogen fixation were detected in *Arcobacter* and *Sulfuriferula* MAGs, along with the complete DNR pathway (Table S4; Fig. 3b). The DNR pathway was also present in *Thioclava*_G and a partial pathway, missing the enzymes for the production of nitrite, was present in *Sulfurovum*, *Sulfuricurvum* and *Sulfurimonas*. The remaining MAGs possessed only a few or

Table 2. Taxonomic identification of MAGs and assembly statistics.

Sample	Bin	MAG name	Taxonomy NCBI ($p < 0.10$)	Closest_relative NCBI (AAI)	Completeness	Contamination	# Genes	Length (Mbp)	GC%	N50
DE13	14	Arcobacter_C	*Epsilonproteobacteria* (Class)	Arcobacter sp. L NC 017192 (47.14% AAI)	82.9	9.9	1637	1.37	34.74	3397
JN15	18	Arcobacter_O	*Campylobacterales* (order)	Arcobacter sp. L NC 017192 (55.41% AAI)	94.6	9	4141	3.32	33.02	5788
AU13	5	Arcobacter_S	*Arcobacter* sp. L (Species)	Arcobacter sp. L NC 017192 (86.86% AAI)	84.7	9.9	3030	2.36	28.07	3008
DE14	2	Brevundimonas_Ga	*Brevundimonas* (genus)	Brevundimonas diminuta NZ CP021995 (81.06% AAI)	75.7	1.8	3446	3.49	66.88	46,130
OC16	6	Brevundimonas_Gb	*Brevundimonas* (genus)	Brevundimonas sp. DS20 NZ CP012897 (83.75% AAI)	93.7	6.3	3844	3.53	68.25	34,665
DE14	3	Brevundimonas_S	*Brevundimonas* diminuta	Brevundimonas diminuta NZ CP021995 (96.83)	73	1.8	3439	3.41	67.76	51,630
JN15	14	Halothiobacillus_G	*Halothiobacillus (Genus)*	Halothiobacillus neapolitanus c2 NC 013422T (89.46%)	83.8	2.7	2684	2.25	56.06	4350
MY16	11	Hydrogenovibrio_O	*Thiotrichales* (Order)	Hydrogenovibrio crunogenus XCL 2 NC 007520 (63.68% AAI)	89.2	0.9	2345	2.14	45.1	8061
AP15	6	Stenotrophomonas_G	*Stenotrophomonas* (genus)	Stenotrophomonas rhizophila NZ CP007597T (86.11% AAI)	86.5	0.9	3624	4.04	67.72	56,024
DE12	3	Sulfuricurvum_C	Epsilonproteobacteria (Class)	Sulfuricurvum kujiense DSM 16994 NC 014762T (55.21% AAI)	88.3	4.5	2470	2.19	45.61	9117
JN15	6	Sulfuricurvum_Ga	*Sulfuricurvum* (Genus)	Sulfuricurvum kujiense DSM 16994 NC 014762T (79.16% AAI)	91	2.7	1787	1.53	45.58	9298
DE16	5	Sulfuricurvum_Gb	*Sulfuricurvum* (Genus)	Sulfuricurvum kujiense DSM 16994 NC 014762T (79.05% AAI)	92.8	2.7	1693	1.5	44.39	10 849
AP10	10	Sulfuricurvum_O	*Campylobacterales* (order)	Sulfuricurvum kujiense DSM 16994 NC 014762T (55.56%)	72.1	0.9	1817	1.6	57.92	16 172
DE15	11	Sulfuriferula_C	*Betaproteobacteria* (Class)	Sulfuriferula sp. AH1 NZ CP021138 (53.45%	95.5	8.1	4449	3.97	65.8	34 175
MY16	8	Sulfurimonas_C	*Epsilonproteobacteria* (Class)	Sulfurimonas autotrophica DSM 16294 NC 014506T (54.43% AAI)	84.7	1.8	1645	1.46	44.24	8986
MY14	6	Sulfurimonas_O	*Campylobacterales* (order)	Sulfurimonas autotrophica DSM 16294 NC 014506T (64.07% AAI	91	7.2	2287	1.99	37.39	10,100
DE12	5	Sulfurovum_Ca	Epsilonproteobacteria (Class)	Sulfurovum sp. NBC37 1 NC 009663 (56.17% AAI)	82	6.3	1913	1.43	34.37	3088
DE16	8	Sulfurovum_Cb	Epsilonproteobacteria (Class)	Sulfurovum lithotrophicum NZ CP011308T (52.68% AAI)	93.7	3.6	1914	1.68	33.98	11 871
AP10	1	Sulfurovum_Oa	*Campylobacterales* (order)	Sulfurovum lithotrophicum NZ CP011308T (63.3% AAI)	92.8	4.5	1839	1.76	41.56	72 213
AP10	8	Sulfurovum_Ob	*Campylobacterales* (order	Sulfurovum lithotrophicum NZ CP011308T (56.37% AAI)	89.2	3.6	1861	1.74	42.47	7244
MY14	5	Sulfurovum_Oc	*Campylobacterales* (order	Sulfurovum lithotrophicum NZ CP011308T (63.07% AAI)	93.7	1.8	2082	1.96	46.8	25 751
JN11	13	Thioclava_G	*Thioclava* (Genus)	Thioclava nitratireducens NZ CP019437T (82.65% AAI)	92.8	2.7	4007	3.85	64.05	19 975
DE12	2	Thiomicrospira_O	*Thiotrichales* (Order)	Thiomicrospira aerophila AL3 NZ CP007030T (56.15% AAI)	84.7	0.9	2296	2.06	51.78	5551

Best taxonomic classification ($p < 0.1$) and closest relative according to MiGA NCBI Prok Database.

no genes of the pathway. The ANR pathway was complete only in *Sulfuricurvum* and genes for enzymes catalysing the first step of ANR (i.e. the formation of nitrite) were detected in *Arcobacter*_S, *Thioclava*_G, *Sulfurimonas*_C and *Thiotrichales* MAGs (Fig. 3B). These findings agree with previous studies showing that nitrate can be used as an alternative electron acceptor for members of the *Sulfurovum*, *Sulfuricurvum*, *Sulfurimonas* and *Arcobacter* genera (Hamilton *et al*., 2015; Han and Perner, 2015).

Finally, the TCA cycle was observed in all MAGs as expected but the key genes for the reductive TCA cycle, i.e. ATP citrate lyase; fumarate reductase and 2-oxoglutarate:ferredoxin oxidoreductase were observed only in epsilonproteobacterial MAGs apart from Arcobacter_C (Fig. 3C) indicating that these species are capable of CO_2 fixation via the reductive TCA cycle. On the other hand, the relatively abundant gammaproteobacterial MAGs, *Halothiobacilus*_G, *Hygrogenovibrio*_O, *Thiomicrospira*_O, *Thioclava*_G were capable of CO_2 fixation via the CBB cycle. The latter three MAGs belong to *Piscirickettsiales*, which was consistent with the increased carboxysome-related subsystems detected in the metagenomes analysis (Fig. 2B). Finally, the genes encoding the phosphate acetyltransferase-acetate kinase pathway were detected in *Sulfuriferula*_C (Fig. 3B).

Correlation of MAG abundances with environmental parameters. In general, only a few rather weak correlations were observed between MAG relative abundances and environmental parameters (Table S9). Most notably, pH value was positively correlated with the abundance of *Sulfuricurvum*_G, the temperature was negatively correlated with *Sulfurovum*_O and positively with *Sulfurimonas*_O abundances, while the conductivity was positively correlated with *Halothiobacillus* abundance. Several correlations in abundances were observed between MAGs (Table S9), indicating possible synergistic or antagonistic interactions. Canonical correspondence analysis (CCA) confirmed the above-mentioned results on differences in community composition characterizing samples DE11 and DE14 as outliers also at the MAG level, while pH and conductivity emerged as important factors ($p < 0.05$) for the ordination of the samples based on MAG abundances (Fig. S4).

Discussion. As mentioned above, Thermopyles are characterized by high sulfide concentrations; hence it is expected that sulfides are an important energy source for autotrophic microorganisms (Porter *et al*., 2009; Han *et al*., 2012; Rossmassler *et al*., 2016). Moreover, *Epsilonproteobacteria* have been detected to be key players in sulfur metabolism in sulfide-rich environments (Han *et al*., 2012; Hamilton *et al*., 2015; Rossmassler *et al*., 2016) and were also prevalent in the Thermopyles samples collected in 2005 (Kormas *et al*., 2009). In agreement with these expectations and previous results, *Epsilonproteobacteria* were prevalent in our time series and their relative abundances were largely unaffected by environmental factors as revealed by 16S rRNA gene and MAG data (Figs 1A and 3A). Furthermore, the presence of a relatively large number of core species, which together made up, on average, about 75% of the total communities sampled, revealed that Thermopyles microbial communities are dominated by specific microbial populations, which probably interact with each other (synergistically or competitively) and represent mostly novel genera, if not novel families. A previous study of the YNP geothermal hot springs also revealed the presence of core communities despite a limited number of samples used for analysis (three samples collected over 3 years; De León *et al*., 2013). A riverine ecosystem (in Kalamas; Greece) in the same geographic area and studied with similar methods and sequencing effort as the Thermopyles showed temporal Morisita similarities <30%, and no MAGs in high abundance (e.g. >10× coverage) in more than one sample over a 3-year sampling period (Meziti *et al*., 2016; Meziti *et al*., 2018). Although surface riverine ecosystems are known to undergo more temporal disturbances and seasonal fluctuations than geothermal springs and direct comparisons to geothermal springs need to be interpreted with caution, these results do contrast with the high stability (>60% Morisita similarities between most samples, Fig. 1B) and enrichment of sulfur-oxidizing species in the Thermopyles (coverage >10× over time, Fig. 3A).

Although it was expected that sulfur cycling–related pathways would be prevalent in the Thermopyles, our study represents the first actual documentation of prevalent microbially mediated processes related to sulfur and thiosulfate oxidation as well as assimilatory and dissimilatory sulfate reduction in a terrestrial mesophilic sulfur-rich geothermal spring. Our analysis also revealed a large functional gene diversity, especially in terms of the number of distinct co-occurring variants (alleles) of specific genes, for the relatively low taxonomic diversity detected (Fig. 1B and D). Core functions that dominated Thermopyles regardless of the prevailing environmental conditions (Fig. 2A) compared with another (non-sulfur rich, non-high temperature) freshwater habitat included CRISPRs, sulfur oxidation and reductive TCA cycle-related genes (Fig. 2B). These results also agreed with previous results from similar habitats (Sikorski *et al*., 2010; López-López *et al*., 2013). Increased diversity of CRISPR-related genes, representing different Cas, Cmr and Csc proteins, has also been detected in Yumthang geothermal spring system (India) that exhibits

similar physicochemical properties with Thermopyles (Najar *et al*., 2020), CRISPR systems are considered stress regulators since they represent defensive tools by attacking viruses and plasmids (Louwen *et al*., 2014) and are more abundant in thermophilic than mesophilic bacteria, as has been shown in some environmental (Makarova *et al*., 2006; Westra *et al*., 2019) and predictive modelling studies (Weissman *et al*., 2019).

The prevalence of genes encoding proteins related to flagellar motility agrees with results from other geothermal habitats (Badhai *et al*., 2015) but the lower abundances in sulfur-rich sediment sample from YNP pointed out the importance of water chemistry characterizing different geothermal springs (Inskeep *et al*., 2013). Similarly, the prevalence of carboxysome genes, related to Rubisco operon transcription regulators and activation proteins, indicated the importance of *Piscirickettsiaceae*-related genera in CO_2 fixation via the CBB cycle, in addition to the reductive TCA cycle, at the Thermopyles. This feature (CBB) was absent from the YNP site, which could be attributable to archaea dominating the YNP site (archaea do not use this pathway for CO_2 fixation; Berg *et al*.,2010).

Furthermore, at the genome level, concurrent with previous data for mesophilic sulfide-rich habitats (Hamilton *et al*., 2015), all the *Epsilonproteobacteria* MAGs analysed possessed pathways for sulfur oxidation and CO_2 fixation via the reductive TCA cycle. The presence of complete or incomplete sulfur oxidation pathways in the genome of the close relatives of the MAGs reported here, coupled to the frequent co-presence of sulfide quinone reductase (*sqr*) genes, further corroborated the metabolic versatility of *Epsilonproteobacteria* (Friedrich *et al*., 2005; Frigaard and Dahl, 2009; Wright *et al*., 2013; Hamilton *et al*., 2015; Han and Perner, 2015; Rossmassler *et al*., 2016).

The presence of different pathways oxidizing different sulfur compounds and producing either sulfate or elemental sulfur, along with correlation analysis results (Table S9; Fig. S4), implied that synergistic or antagonistic relationships between different populations could influence taxonomic and functional diversity at the Thermopyles. Another explanation for the high intra-genus species and pathway diversity could be the ability to utilize alternative sulfur oxidation pathways when conditions are unfavourable or if different affinities of different alleles for the same substrate exist. The importance of other environmental factors that were not measured here (i.e. metal concentrations) should be noted as a probable factor that could be included in future studies to better explain the prevalence of specific species at different time points.

Altogether, it appears that chemoautotrophic microbial communities that mainly oxidize different reduced sulfur compounds using oxygen or nitrate as electron acceptors dominate the Thermopyles thermal springs. Although all epsilonproteobacterial sulfur-oxidizing MAGs detected in this study probably consist of new genera that are members of a new family, they are moderately or distantly related to the genera *Sulfurovum*, *Sulfuricurvum*, *Sulfurimonas* and *Arcobacter* (Table 2). Similar communities have been observed in hyperthermophilic and mesophilic sulfur-rich environments in the past (Hamilton *et al*., 2015; Hotaling *et al*., 2019; Huegler *et al*., 2010; Rossmassler *et al*., 2016; Wright *et al*., 2013), but only once in a terrestrial geothermal spring (Reigstad *et al*., 2011) and cross-sectional (as opposed to a time series of 7 years here) data. Notably, comparisons with genomes from some of these studies (when available) showed that even the (taxonomically) uncharacterized *Epsilonproteobacteria* detected in these previous studies (e.g. MAGs) were remotely affiliated with those recovered here. This indicated, once more, that the bacteria prevalent in the Thermopyles are members of 'novel' genera or families. Nonetheless, these novel taxa possess similar sulfur-oxidizing properties to known species, further supporting a universal functional profile of *Epsilonproteobacteria* in sulfur-rich environments.

The community composition of the two outlier samples with respect to stable core community (DE11 and DE14; Figs 1A and 3A, Fig. S1B) was the only deviation from the stable, core taxa that characterized the Thermopyles and could be attributed to environmental perturbations as evidenced by the distinct conductivity and pH observed in these samples, although the exact perturbation event remains to be confirmed. For instance, the *Halothiobacillus* MAG prevalence in DE11, along with its positive correlation with conductivity, was further supported by the known halotolerant character of *Halothiobacillus* sp. not requiring salt in order to grow but growing optimally when salt concentrations increase (Sievert *et al*., 2000). Altogether, these findings point to a saltwater intrusion event driving the composition of the DE11 sample. Coverage and relative abundance for the *Halothiobacillus* MAG along with 16S rRNA gene data (Figs 1C and 3A; Table S6) support the hypothesis that *Halothiobacillus* sp. is one of the core members of the Thermopyles community that thrives when conditions are favourable (i.e. increased conductivity) and remains at low abundance during the rest of the time.

On the other hand, the increased rainfall prior to sampling the DE14 sample, coupled with a decrease in the pH of the sample and the 16S rRNA and MAG relative abundance and coverage data showing decreased *Epsilonproteobacteria* and increased *Alpha*-, *Betaproteobacteria* and *Actinobacteria* abundances, might indicate influences from the soil. The *Sulfuricurvum*_Ga MAG dropped in abundance in MY14 and DE14 samples, consistent with the positive correlation of its abundance with pH as revealed by our data

(Table S9). This is further supported by previous studies showing that the type species of the genus, *Sulfuricurvum kujiense* (Kodama and Watanabe, 2004), has a pH range between 6 and 8 while it grows optimally at pH 7 (Han *et al*., 2012). However, the prevalence of *Brevundimonas*-related species in DE14 could not be linked to the pH drop since *Brevundimonas* species are alkalophiles, and this is further supported by their presence in alkaline thermal springs (Tekere *et al*., 2011; Gupta *et al*., 2017). Hence, our findings overall are not conclusive about the exact underlying cause of the unique diversity observed in sample DE14, although it was most certainly related to inputs from non-geothermal sources.

Collectively, our results showed that the Thermopyles microbial communities are mainly fuelled by chemolithotrophic processes performed by core sulfide-oxidizing bacterial taxa that persist over time. Short intervals might be driven by changes in environmental parameters but changes in taxonomy never override the persistence of core functional gene potential. The outcome of this study, that is, gene and genome sequences from 'novel' sulfur-oxidizing bacteria, will be useful for the metagenomic investigation of other, similar environments. Future studies could assist in further understanding of evolutionary relationships and functionality of these and related microbial communities.

Experimental procedures

Sample collection and processing

Samples were collected from the Thermopyles geothermal springs from June 2010 until December 2016 (Table S1; Fig. S1). Water samples of 5 L were collected in pre-sterilized dark polyethylene bottles from a seepage point with visual water flow and bubbling. Upon return to the lab and within 3 h, each sample was pre-filtered through a sterile 180 μm mesh nylon filter (Millipore, Burlington, MA, USA) to exclude sampling of microorganisms attached to large particles, and then a volume of 1.8–4.5 L of filtrate from the pre-filtration was filtered through a 0.2 μm isopore polycarbonate filter (Sartorius, Goettingen, Germany) under mild vacuum (<100 mmHg) to collect microbial biomass. Subsequently, filters were folded aseptically, placed in sterile cryovials and stored at −80°C until further processed. *In situ* measurements, including temperature, pH and conductivity (Table S1), were taken by a portable multisensor instrument (WTW/Xylem, Rye Brook, NY, USA). Subsamples of 10–15 ml were fixed with 2% formaldehyde final concentration and kept at 4°C in the dark for cell counts. These subsamples were filtered to black Nuclepore filters (pore size of 0.2 μm) and stained with DAPI (0.1 μg ml^{-1}). DNA extraction was performed with the MoBio Power Soil kit (MoBio, Carlsbad, CA, USA) following its standard protocol. For all samples, libraries were prepared using the Illumina Nextera XT DNA library prep kit according to manufacturer's instructions, except that the protocol was terminated after isolation of cleaned double-stranded libraries. An equimolar mixture of the libraries was sequenced on an Illumina HiSEQ 2500 instrument (High Throughput Sequencing Core, Georgia Institute of Technology) for 300 cycles (2 × 150 bp paired-end rapid run). Adapter trimming and demultiplexing of sequenced samples was carried out by the instrument.

Metagenomic read sequence trimming and assembly

Illumina reads were trimmed using a Q = 20 Phred quality score cut-off using SolexaQA++ (Cox *et al*., 2010) and only trimmed reads longer than 50 bp were considered for further analysis. Metagenomic reads were assembled using IDBA with default settings for metagenomes (Peng *et al*., 2012). Protein-coding genes were predicted from contigs longer than 500 bp using MetaGeneMark.hmm with default parameters (Zhu *et al*., 2010). Sequencing and assembly statistics for the metagenomic datasets are provided in Table 1. Metagenomic datasets have been deposited in NCBI SRA under the bioproject PRJNA611516.

Metagenomic reads carrying fragments of the 16S rRNA gene were identified with Parallel-META (v.2.1) (Su *et al*., 2012), and were subsequently processed for OTU picking (>97% sequence identity threshold) and taxonomic identification with SILVA128 database (Pruesse *et al*., 2007) using MacQIIME 1.8.0 and the pick_closed_reference_otus.py script with default parameters (Caporaso *et al*., 2010). However, this script cannot detect novel diversity for sequences that do not match the reference sequences at the threshold used. For this reason, unassigned sequences were further analysed using mothur 1.35.0 (Schloss *et al*., 2009). Sequences were aligned and classified using the SILVA 128 database following mothur's instructions (cutoff = 80). Thus aligned sequences were classified accordingly to higher taxonomic levels (i.e. Class, Family, etc.) and OTUs were built using a 97% identity threshold as described above. The relative abundance of OTUs or taxa (phylum, class and genus level) in each dataset was estimated based on the sum of the abundances of the corresponding OTUs or taxa, normalized for dataset size (i.e. total number of reads) using the DESeq package version 1.26.0 (Anders and Huber, 2010).

Population genome binning

Contigs longer than 500 bp were binned into MAGs using MaxBin v2.1.1 with default settings (Wu *et al*., 2014). In each binning run, only contigs from the assembly of an

individual sample were used (no co-assembly was performed). CheckM and the MiGA web server (www.microbial-genomes.org) were used to estimate completeness and contamination of each MAG based on the recovery of single-copy universal bacterial proteins (Parks *et al*., 2015; Rodriguez-R *et al*., 2018). Recruitment plots and coverage for MAG contigs and genes were calculated using the 'enveomics.R' package v1.4.1 from the Enveomics Collection (Rodriguez-R and Konstantinidis, 2016). Final MAGs are available through http://enve-omics.ce.gatech.edu/data/ and in NCBI SRA under the bioproject PRJNA611516. The relative abundance of MAGs was calculated as $100 \times MAG_{coverage} \times (MAG_{size}(bp)/Metagenome_{size}(bp))$. Taxonomy of the MAGs was estimated by MiGA with the NCBI Genome (Prokaryotes) and TypeMAT databases containing all quality controlled genomes from prokaryotic type material contained in NCBI until June 2020 (MiGA online; June 2020). Genomes were also classified using GTDB-tK in order to also include previously published MAGs for comparisons (Chaumeil *et al*., 2020).

Functional annotation of predicted genes and determination of differentially abundant features

The predicted protein sequences encoded in the MAGs and the assembled contigs were searched against uniprot-TrEMBL(2018) to assign functional annotation based on best matches, and ≥40% similarity, ≥60 bitscore and ≥80% alignment length cutoffs for a match. Predicted genes were further grouped in functional categories based on their best match against the SEED database using the subsystems categories (Overbeek *et al*., 2005) and the KEGG-orthology using GhostKOALA (Kanehisa *et al*., 2016).

Metagenomic reads were mapped on the predicted genes from the assembly using BLAT (Kent, 2002) with at least 95% identity and 50% of query length aligned. The abundance of each gene on each dataset was estimated by the number of reads that mapped on the genes with the above cutoffs (gene coverage) after normalizing for gene length. The relative abundance of annotation terms (subsystems) in each dataset was estimated based on the sum of the abundances of the corresponding genes, normalized for dataset size (i.e. total number of reads). Differentially abundant categories between samples were identified with the DESeq package version 1.26.0 (Anders and Huber, 2010) using the binomial test with a 0.05 adjusted *p*-value as a threshold for differential abundance (Benjamini–Hochberg correction; (Benjamini and Hochberg, 1995).

Diversity and similarity indices, multivariate and correlation analysis. Shannon and Chao indices, and richness of observed OTUs were calculated using PAST (Hammer *et al*., 2020). NMDS analysis and Cluster analysis using Morisita similarities (Wolda, 1981) were performed as described in Meziti *et al*. (2015). The significance of environmental parameters for the ordination of the samples was calculated using the function envfit of the R package vegan. Spearman correlations between MAG abundances and environmental factors were calculated using the PAST software (Hammer *et al*., 2020). CCA between MAG abundance and environmental parameters was performed using the Vegan package in R (Oksanen *et al*., 2019).

Comparisons to other metagenomes

Available protein-coding genes datasets from the Kalamas River metagenomes in Greece (kal2feb; Meziti *et al*., 2016) and YNP (CIS_19; Inskeep *et al*., 2013) metagenomes were analysed and annotated similarly to the Thermopyles datasets. These datasets were used to compare microbial community function between the Thermopyles and another freshwater environment in Greece (Kalamas) and a geothermal spring in the United States (Yellowstone). Comparisons between metagenomic datasets were performed based on the gene counts of the different functional annotation terms (subsystems). To make these counts comparable between datasets, protein sequences were first clustered using the CD-HIT algorithm (Fu *et al*., 2012) with the following parameters: $S = 97$ (similarity threshold) and $L = 0.5$ (minimum length coverage). Representative proteins from each cluster were annotated based on their best match against the SEED database. Comparisons between gene counts of different subsystems were performed using DESeq as described above.

Acknowledgements

This work was partly funded by the US National Science Foundation, awards #1831582 and #1759831 (to KTK), and by internal funding from the University of Thessaly (KAK). We thank the Partnership for an Advanced Computing Environment (PACE) at the Georgia Institute of Technology, which enabled the computational tasks associated with this study.

References

Anders, S., and Huber, W. (2010) Differential expression analysis for sequence count data. *Genome Biol* **11**: R106.

Badhai, J., Ghosh, T.S., and Das, S.K. (2015) Taxonomic and functional characteristics of microbial communities and their correlation with physicochemical properties of four geothermal springs in Odisha, India. *Front Microbiol* **6**: 1166.

Barns, S.M., Delwiche, C.F., Palmer, J.D., and Pace, N.R. (1996) Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. *Proc Natl Acad Sci U S A* **93**: 9188–9193.

Barns, S.M., Fundyga, R.E., Jeffries, M.W., and Pace, N.R. (1994) Remarkable archaeal diversity detected in a Yellowstone National Park hot spring environment. *Proc Natl Acad Sci U S A* **91**: 1609–1613.

Benjamini, Y., and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B Methodol* **57**: 289–300.

Boomer, S.M., Lodge, D.P., Dutton, B.E., and Pierson, B. (2002) Molecular characterization of novel red green non-sulfur bacteria from five distinct hot spring communities in Yellowstone National Park. *Appl Environ Microbiol* **68**: 346–355.

Burgess, E.A., Unrine, J.M., Mills, G.L., Romanek, C.S., and Wiegel, J. (2012) Comparative geochemical and microbiological characterization of two thermal pools in the Uzon Caldera, Kamchatka, Russia. *Microb Ecol* **63**: 471–489.

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

Chan, C.S., Chan, K.-G., Tay, Y.-L., Chua, Y.-H., and Goh, K.M. (2015) Diversity of thermophiles in a Malaysian hot spring determined using 16S rRNA and shotgun metagenome sequencing. *Front Microbiol* **6**: 177.

Chaumeil, P.-A., Mussig, A.J., Hugenholtz, P., and Parks, D. H. (2020) GTDB-Tk: a toolkit to classify genomes with the genome taxonomy database. *Bioinformatics* **36**: 1925–1927.

Childs, A.M., Mountain, B.W., O'Toole, R., and Stott, M.B. (2008) Relating microbial community and physicochemical parameters of a hot Spring: Champagne Pool, Wai-o-tapu, New Zealand. *Geomicrobiol J* **25**: 441–453.

Cox, M.P., Peterson, D.A., and Biggs, P.J. (2010) SolexaQA: at-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinformatics* **11**: 485.

Damer, B., and Deamer, D. (2019) The hot Spring hypothesis for an origin of life. *Astrobiology* **20**: 429–452.

De León, K., Gerlach, R., Peyton, B., and Fields, M. (2013) Archaeal and bacterial communities in three alkaline hot springs in heart Lake Geyser Basin, Yellowstone National Park. *Front Microbiol* **4**: 330.

Duriez, A., Marlin, C., Dotsika, E., Massault, M., Noret, A., and Morel, J.L. (2008) Geochemical evidence of seawater intrusion into a coastal geothermal field of Central Greece: example of the Thermopylae system. *Environ Geol* **54**: 551–564.

Engel, A.S., Porter, M.L., Stern, L.A., Quinlan, S., and Bennett, P.C. (2004) Bacterial diversity and ecosystem function of filamentous microbial mats from aphotic (cave) sulfidic springs dominated by chemolithoautotrophic "*Epsilonproteobacteria*.". *FEMS Microbiol Ecol* **51**: 31–53.

Falkowski, P.G., Fenchel, T., and Delong, E.F. (2008) The microbial engines that drive Earth's biogeochemical cycles. *Science* **320**: 1034–1039.

Flores, G.E., Liu, Y., Ferrera, I., Beveridge, T.J., and Reysenbach, A.-L. (2008) *Sulfurihydrogenibium kristjanssonii* sp. nov., a hydrogen- and sulfur-oxidizing thermophile isolated from a terrestrial Icelandic hot spring. *Int J Syst Evol Microbiol* **58**: 1153–1158.

Friedrich, C.G., Bardischewsky, F., Rother, D., Quentmeier, A., and Fischer, J. (2005) Prokaryotic sulfur oxidation. *Curr Opin Microbiol* **8**: 253–259.

Frigaard, N.-U., and Dahl, C. (2009) Sulfur metabolism in phototrophic sulfur bacteria. *Adv Microb Physiol* **54**: 103–200.

Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012) CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**: 3150–3152.

Georgalas, G. and Papakis, N. (1966) Observations sur les sources ordinaries et thermominerales radioactives de la region karstique de Kammena Vourla (Grece Centrale). In *Proceedings of the International Association of Hydrogeologists*. Knjiga, N., pp. 221–227.

Gribaldo, S., Lumia, V., Creti, R., Conway de Macario, E., Sanangelantoni, A., and Cammarano, P. (1999) Discontinuous occurrence of the hsp70 (dnaK) gene among Archaea and sequence features of HSP70 suggest a novel outlook on phylogenies inferred from this protein. *J Bacteriol* **181**: 434–443.

Gupta, V., Gupta, N., Capalash, N., and Sharma, P. (2017) Bio-prospecting bacterial diversity of Hot Springs in northern Himalayan region of India for Laccases. *Indian J Microbiol* **57**: 285–291.

Hamilton, T.L., Jones, D.S., Schaperdoth, I., and Macalady, J.L. (2015) Metagenomic insights into S(0) precipitation in a terrestrial subsurface lithoautotrophic ecosystem. *Front Microbiol* **5**: 756.

Hammer, O., Harper, D.A.T., and Ryan, P.D. (2001) PAST: Paleontological Statistics Software Package for Education and Data Analysis. 9.

Han, C., Kotsyurbenko, O., Chertkov, O., Held, B., Lapidus, A., Nolan, M., *et al*. (2012) Complete genome sequence of the sulfur compounds oxidizing chemolithoautotroph *Sulfuricurvum kujiense* type strain (YK-1T). *Stand Genomic Sci* **6**: 94–103.

Han, Y., and Perner, M. (2015) The globally widespread genus *Sulfurimonas*: versatile energy metabolisms and adaptations to redox clines. *Front Microbiol* **6**: 989.

Headd, B., and Engel, A.S. (2013) Evidence for niche partitioning revealed by the distribution of sulfur oxidation genes collected from areas of a terrestrial sulfidic Spring with differing geochemical conditions. *Appl Environ Microbiol* **79**: 1171–1182.

Hiras, J., Wu, Y.-W., Eichorst, S.A., Simmons, B.A., and Singer, S.W. (2016) Refining the phylum *Chlorobi* by resolving the phylogeny and metabolic potential of the representative of a deeply branching, uncultivated lineage. *ISME J* **10**: 833–845.

Hotaling, S., Quackenbush, C.R., Bennett-Ponsford, J., New, D.D., Arias-Rodriguez, L., Tobler, M., and Kelley, J. L. (2019) Bacterial diversity in replicated hydrogen sulfide-rich streams. *Microb Ecol* **77**: 559–573.

Hugenholtz, P., Pitulle, C., Hershberger, K.L., and Pace, N. R. (1998) Novel division level bacterial diversity in a Yellowstone Hot Spring. *J Bacteriol* **180**: 366–376.

Hügler, M., Gärtner, A., and Imhoff, J.F. (2010) Functional genes as markers for sulfur cycling and CO_2 fixation in

microbial communities of hydrothermal vents of the Logatchev field. *FEMS Microbiology Ecology* **73**: 526–537.

Iino, T., Mori, K., Uchino, Y., Nakagawa, T., Harayama, S., and Suzuki, K.-I. (2010) *Ignavibacterium album* gen. nov., sp. nov., a moderately thermophilic anaerobic bacterium isolated from microbial mats at a terrestrial hot spring and proposal of Ignavibacteria classis nov., for a novel lineage at the periphery of green sulfur bacteria. *Int J Syst Evol Microbiol* **60**: 1376–1382.

Inagaki, F., Takai, K., Kobayashi, H., Nealson, K.H., and Horikoshi, K. (2003) *Sulfurimonas autotrophica* gen. nov., sp. nov., a novel sulfur-oxidizing epsilon-proteobacterium isolated from hydrothermal sediments in the mid-Okinawa Trough. *Int J Syst Evol Microbiol* **53**: 1801–1805.

Inskeep, W.P., Jay, Z.J., Tringe, S.G., Herrgård, M.J., Rusch, D.B., and YNP Metagenome Project Steering Committee and Working Group Members. (2013) The YNP metagenome project: environmental parameters responsible for microbial distribution in the Yellowstone geothermal ecosystem. *Front Microbiol* **4**: 67.

Jeon, W., Priscilla, L., Park, G., Lee, H., Lee, N., Lee, D., *et al*. (2017) Complete genome sequence of the sulfur-oxidizing chemolithoautotrophic *Sulfurovum lithotrophicum* 42BKTT. *Stand Genomic Sci* **12**: 54.

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2016) KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* **44**: D457–D462.

Kanellopoulos, C., Lamprinou, V., Mitropoulos, P., and Voudouris, P. (2016) Thermogenic travertine deposits in Thermopylae hot springs (Greece) in association with cyanobacterial microflora. *Carbonate Evaporite* **31**: 239–248.

Kent, W.J. (2002) BLAT - the BLAST-like alignment tool. *Genome Res* **12**: 656–664.

Kodama, Y., and Watanabe, K. (2004) *Sulfuricurvum kujiense* gen. nov., sp. nov., a facultatively anaerobic, chemolithoautotrophic, sulfur-oxidizing bacterium isolated from an underground crude-oil storage cavity. *Int J Syst Evol Microbiol* **54**: 2297–2300.

Konhauser, K.O., Phoenix, V.R., Bottrell, S.H., Adams, D.G., and Head, I.M. (2001) Microbial–silica interactions in Icelandic hot spring sinter: possible analogues for some Precambrian siliceous stromatolites. *Sedimentology* **48**: 415–433.

Kormas, K.A., Tamaki, H., Hanada, S., and Kamagata, Y. (2009) Apparent richness and community composition of *Bacteria* and *Archaea* in geothermal springs. *Aquat Microb Ecol* **57**: 113–122.

Lahme, S., Callbeck, C.M., Eland, L.E., Wipat, A., Enning, D., Head, I.M., and Hubert, C.R.J. (2020) Comparison of sulfide-oxidizing *Sulfurimonas* strains reveals a new mode of thiosulfate formation in subsurface environments. *Environ Microbiol* **22**: 1784–1800.

Lambrakis, N., and Kallergis, G. (2005) Contribution to the study of Greek thermal springs: hydrogeological and hydrochemical characteristics and origin of thermal waters. *Hydrgeol J* **13**: 506–521.

Lambrakis, N., Katsanou, K., and Siavalas, G. (2014) Geothermal fields and thermal waters of Greece: an overview. In *Geothermal Systems and Energy Resources*, London: CRC Press, pp. 63–84.

López-López, O., Cerdán, M.E., and González-Siso, M.I. (2013) Hot Spring metagenomics. *Life* **3**: 308–320.

Louwen, R., Staals, R.H.J., Endtz, H.P., van Baarlen, P., and van der Oost, J. (2014) The role of CRISPR-Cas systems in virulence of pathogenic *bacteria*. *Microbiol Mol Biol Rev* **78**: 74–88.

Makarova, K.S., Grishin, N.V., Shabalina, S.A., Wolf, Y.I., and Koonin, E.V. (2006) A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol Direct* **1**: 7.

Mangiapia, M., and Scott, K. (2016) From CO_2 to cell: energetic expense of creating biomass using the Calvin-Benson-Bassham and reductive citric acid cycles based on genome data. *FEMS Microbiol Lett* **363**: fnw054.

Marinos, P., Frangopoulos, J., and Stournaras, G. (1973) The thermomineral spring of Hypati (Central Greece): hydrogeological, hydrodynamical, geochemical and geotechnical study of the spring and the surrounding area. *Ann Geol Pays Hellen* **1**: 105–214.

McCleskey, R.B., Nordstrom, D.K., Susong, D.D., Ball, J.W., and Taylor, H.E. (2010) Source and fate of inorganic solutes in the Gibbon River, Yellowstone National Park, Wyoming, USA. II Trace element chemistry. *J Volcanol Geotherm Res* **196**: 139155.

Meyer-Dombard, D.R., Shock, E.L., and Amend, J.P. (2005) Archaeal and bacterial communities in geochemically diverse hot springs of Yellowstone National Park, USA. *Geobiology* **3**: 211–227.

Meziti, A., Tsementzi, D., Kormas, K.A., Karayanni, H., and Konstantinidis, K.T. (2016) Anthropogenic effects on bacterial diversity and function along a river-to-estuary gradient in Northwest Greece revealed by metagenomics. *Environ Microbiol* **18**: 4640–4652.

Meziti, A, Tsementzi, D, Rodriguez-R, LM, Hatt, JK, Karayanni, H, & Kormas, KA (2018). Quantifying the changes in genetic diversity within sequence-discrete bacterial populations across a spatial and temporal riverine gradient. *ISME*, **13**: 767–779.

Meziti, Kormas, & Moustaka-Gouni (2015). Spatially uniform but temporally variable bacterioplankton in a semi-enclosed coastal area. *Syst Appl Microbiol*, **38**: 358–367.

Najar, I.N., Sherpa, M.T., Das, S., and Thakur, N. (2020) Bacterial diversity and functional metagenomics expounding the diversity of xenobiotics, stress, defense and CRISPR gene ontology providing eco-efficiency to Himalayan Hot Springs. *Funct Integr Genomics* **20**: 479–496.

Nordstrom, D.K., Ball, J.W., and McCleskey, R.B. (2005) Ground water to surface water: chemistry of thermal outflows in Yellowstone National Park. 73–94.

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., *et al*. (2019) vegan: Community Ecology Package.

Overbeek, R., Begley, T., Butler, R.M., Choudhuri, J.V., Chuang, H.-Y., Cohoon, M., *et al*. (2005) The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res* **33**: 5691–5702.

Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., and Tyson, G.W. (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* **25**: 1043–1055.

Peng, Y., Leung, H.C.M., Yiu, S.M., and Chin, F.Y.L. (2012) IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **28**: 1420–1428.

Porter, M., Engel, A., Kane, T., and Kinkle, B. (2009) Productivity-diversity relationships from chemolithoautotrophically based Sulfidic karst systems. *Int J Speleol* **38**: 27–40.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J., and Glöckner, F.O. (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* **35**: 7188–7196.

Reigstad, L.J., Jorgensen, S.L., Lauritzen, S.-E., Schleper, C., and Urich, T. (2011) Sulfur-oxidizing chemolithotrophic proteobacteria dominate the microbiota in high Arctic thermal springs on Svalbard. *Astrobiology* **11**: 665–678.

Reysenbach, A.-L., Banta, A.B., Civello, S., Daly, J., Mitchel, K., Lalonde, S.V., *et al*. (2006) *Aquificales* in Yellowstone National Park.

Rodriguez-R, L.M., Gunturu, S., Harvey, W.T., Rosselló-Mora, R., Tiedje, J.M., Cole, J.R., and Konstantinidis, K.T. (2018) The microbial genomes atlas (MiGA) webserver: taxonomic and gene diversity analysis of *Archaea* and *Bacteria* at the whole genome level. *Nucleic Acids Res* **46**: W282–W288.

Rodriguez-R, L.M., and Konstantinidis, K.T. (2014) Nonpareil: a redundancy-based approach to assess the level of coverage in metagenomic datasets. *Bioinformatics* **30**: 629–635.

Rodriguez-R, L.M., and Konstantinidis, K.T. (2016) The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. *PeerJ Preprints* **4**: e1900v1.

Rossmassler, K., Hanson, T.E., and Campbell, B.J. (2016) Diverse sulfur metabolisms from two subterranean sulfidic spring systems. *FEMS Microbiol Lett* **363**: fnw162.

Rye, R., and Truesdell, A. (2007) *The Question of Recharge to the Deep Thermal Reservoir Underlying the Geysers and Hot Springs of Yellowstone National Park*, Menlo Park, CA: Publications of the US Geological Survey.

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

Sievert, S.M., Heidorn, T., and Kuever, J. (2000) *Halothiobacillus kellyi* sp. nov., a mesophilic, obligately chemolithoautotrophic, sulfur-oxidizing bacterium isolated from a shallow-water hydrothermal vent in the Aegean Sea, and emended description of the genus *Halothiobacillus*. *Int J Syst Evol Microbiol* **50**: 1229–1237.

Sikorski, J., Munk, C., Lapidus, A., Djao, O.D.N., Lucas, S., Glavina Del Rio, T., *et al*. (2010) Complete genome sequence of *Sulfurimonas autotrophica* type strain (OK10T). *Stand Genomic Sci* **3**: 194–202.

Spieck, E., Spohn, M., Wendt, K., Bock, E., Shively, J., Frank, J., *et al*. (2020) Extremophilic nitrite-oxidizing *Chloroflexi* from Yellowstone hot springs. *ISME J* **14**: 364–379.

Su, X., Xu, J., and Ning, K. (2012) Parallel-META: efficient metagenomic data analysis based on high-performance computation. *BMC Syst Biol* **6**: 1–11.

Takacs-Vesbach, C., Mitchell, K., Jackson-Weaver, O., and Reysenbach, A.-L. (2008) Volcanic calderas delineate biogeographic provinces among Yellowstone thermophiles. *Environ Microbiol* **10**: 1681–1689.

Tekere, M., Lötter, A., Olivier, J., Jonker, N., and Venter, S. (2011) Metagenomic analysis of bacterial diversity of Siloam hot water spring, Limpopo, South Africa. *Afr J Biotechnol* **10**: 18005–18012.

Tobler, D.J., and Benning, L.G. (2011) Bacterial diversity in five Icelandic geothermal waters: temperature and sinter growth rate effects. *Extremophiles* **15**: 473–485.

Toh, H., Sharma, V.K., Oshima, K., Kondo, S., Hattori, M., Ward, F.B., *et al*. (2011) Complete genome sequences of *Arcobacter butzleri* ED-1 and *Arcobacter* sp. strain L, both isolated from a microbial fuel cell. *J Bacteriol* **193**: 6411–6412.

Toplin, J.A., Norris, T.B., Lehr, C.R., McDermott, T.R., and Castenholz, R.W. (2008) Biogeographic and phylogenetic diversity of thermoacidophilic cyanidiales in Yellowstone National Park, Japan, and New Zealand. *Appl Environ Microbiol* **74**: 2822–2833.

Verros, G.D., Latsos, T., Anagnostou, K.E., Avlakiotis, P., Chaikalis, C., Liolios, C., *et al*. (2007) Physico-chemical characteristics of thermopylae natural hot Water Springs in Central Greece. *Chem Geothermometry* **963**: 412–415.

Waite, D.W., Vanwonterghem, I., Rinke, C., Parks, D.H., Zhang, Y., Takai, K., *et al*. (2017) Comparative genomic analysis of the class *Epsilonproteobacteria* and proposed reclassification to *Epsilonbacteraeota* (phyl. nov.). *Front Microbiol* **8**: 682.

Weissman, J.L., Laljani, R.M.R., Fagan, W.F., and Johnson, P.L.F. (2019) Visualization and prediction of CRISPR incidence in microbial trait-space to identify drivers of antiviral immune strategy. *ISME J* **13**: 2589–2602.

Westra, E.R., van Houte, S., Gandon, S., and Whitaker, R. (2019) The ecology and evolution of microbial CRISPR-Cas adaptive immune systems. *Philos Trans R Soc B: Biol Sci* **374**: 20190101.

Wilkins, L.G.E., Ettinger, C.L., Jospin, G., and Eisen, J.A. (2019) Metagenome-assembled genomes provide new insight into the microbial diversity of two thermal pools in Kamchatka, Russia. *Sci Rep* **9**: 3059.

Wolda, H. (1981) Similarity indices, sample size and diversity. *Oecologia* **50**: 296–302.

Wright, K., Williamson, C., Grasby, S., Spear, J., and Templeton, A. (2013) Metagenomic evidence for sulfur lithotrophy by *Epsilonproteobacteria* as the major energy source for primary productivity in a sub-aerial arctic glacial deposit, Borup Fiord Pass. *Front Microbiol* **4**: 63.

Wu, Y.-W., Tang, Y.-H., Tringe, S.G., Simmons, B.A., and Singer, S.W. (2014) MaxBin: an automated binning method to recover individual genomes from metagenomes

using an expectation-maximization algorithm. *Microbiome* **2**: 26.

Yamamoto, H., Hiraishi, A., Kato, K., Chiura, H.X., Maki, Y., and Shimizu, A. (1998) Phylogenetic evidence for the existence of novel thermophilic bacteria in Hot Spring sulfur-turf microbial mats in Japan. *Appl Environ Microbiol* **64**: 1680–1687.

Zarikas, V., Anagnostou, K.E., Avlakiotis, P., Kotsopoulo, S., Liolios, C., Latsos, T., *et al*. (2014) Measurement and analysis of physicochemical parameters concerning thermopylae natural Hot Spring waters. *J Appl Sci* **14**: 2331–2340.

Zheng, H., and Wu, H. (2010) Gene-centric association analysis for the correlation between the guanine-cytosine content levels and temperature range conditions of prokaryotic species. *BMC Bioinformatics* **11**: S7.

Zhu, W., Lomsadze, A., and Borodovsky, M. (2010) Ab initio gene identification in metagenomic sequences. *Nucl Acids Res* **38**: e132.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Appendix S1. Figures.

Appendix S2. Tables.