

## Supplemental Text: Methods and Additional Results

### Brazelton et al. “Metabolic strategies shared by basement residents of the Lost City hydrothermal field”

## SUPPLEMENTAL METHODS

### Collection of hydrothermal fluid samples

Hydrothermal fluid samples were collected from actively venting chimneys at the Lost City hydrothermal field (**Figure 1**) using ROV *Jason* during the 2018 Lost City expedition aboard R/V *Atlantis* (AT42-01). The present study included hydrothermal fluid samples from seven chimney locations: Camel Humps, Sombrero, Marker 3, Marker C, Calypso, Marker 2, and Marker 8 (**Table 1; Supplemental Table S1**). Several fluid samples collected from the Beehive chimney, where the highest fluid temperatures have been measured, yielded only low-quality DNA sequences and are not included in this study. The Sombrero site was sampled on two different ROV *Jason* dives, and samples from the separate dives are labeled as Sombrero1 or Sombrero2 when appropriate. Fluid samples collected from Markers C, 2, 3, and 8 were included in an early microbial diversity study (Brazelton et al., 2006), but microbial diversity data from the other chimneys are reported here for the first time.

On the seafloor, venting fluids were slowly pumped through 0.2 µm Millipore Sterivex cartridge filters or into acid-washed Kynar bags with the HOG sampler (Lang & Benitez-Nelson, 2021). Samples intended for RNA extractions were collected into 2 L Kynar bags containing 67 mL of a stop solution (97.5% 200 proof ethanol, 2.5% Trizol LS; Thermo Fisher). Fluid temperatures were monitored in real-time during sampling with a probe embedded into the sampler intake. Concentrations of sulfate, hydrogen sulfide, and magnesium were measured according to standard methods (Butterfield & Massoth, 1994). The estimated precision of the analyses is 3% for magnesium, 2% for sulfate, and 4% for hydrogen sulfide (all as relative standard deviation).

Chemical analyses were conducted on aliquots of the same fluid samples used for DNA and RNA sequencing. Separate chemical measurements were conducted on separate fluid samples dedicated for the chemistry analyses reported by (Aquino et al., In Revision).

Immediately upon shipboard recovery of ROV *Jason*, all Sterivex filters were stored at -80 °C. Kynar bags were subsampled for shipboard aqueous chemistry measurements (pH, magnesium, sulfate, and sulfide) and filtered through Sterivex filters, which were subsequently frozen at -80 °C. All frozen filters were shipped overnight to the University of Utah on dry ice or in liquid nitrogen vapor shippers.

#### **Cell counts**

Aliquots of all fluid samples were preserved in 3.7% formaldehyde and stored at 4 °C. In the Anderson lab (Carleton College), preserved fluids were filtered onto 0.22 µm glass fiber filters and stained with DAPI (Porter & Feig, 1980). Cells were enumerated directly with an epifluorescence microscope, and a minimum of 20 fields of view were counted per sample. The median number of cells per field of view was recorded as the result for that sample, and the median value among all replicates is reported as the value for each location in **Table 1**.

#### **Bags supplemented with formate or methane**

On each dive, some of the 2 L Kynar bags were supplemented with <sup>13</sup>C-enriched bicarbonate, formate, or methane (Cambridge Isotope Laboratories) to test for conversion of these compounds into dissolved inorganic carbon (DIC) or methane inside the Kynar bags. Thus, these Kynar bags served as seafloor incubation experiments. In addition, each bag contained 0.2 g of dithiothreitol, intended as a redox buffer. Upon shipboard recovery, the bags were subsampled and processed for DNA sequencing as described above, and additional aliquots of the fluid sample were collected for analysis of <sup>13</sup>C enrichment of methane and DIC. In the Lang lab (U.

South Carolina), the ratio of  $^{13}\text{C}/^{12}\text{C}$  of headspace  $\text{CO}_2$  was analyzed using a Thermo Scientific GasBench-Delta V Plus Isotope Ratio Mass Spectrometer (IRMS). The  $^{13}\text{C}/^{12}\text{C}$  of  $\text{CH}_4$  was analyzed with a Thermo Scientific Gas Chromatograph-IsolinkII-IRMS equipped with an Agilent GS - CarbonPlot column (30 m x 0.320 mm i.d., 1.50  $\mu\text{m}$  film thickness).

None of the bags supplemented with  $^{13}\text{C}$ -enriched formate or DIC produced methane with a  $\delta^{13}\text{C}$  that was distinguishable from that of methane native to Lost City fluids (approximately -10 ‰) (Kelley et al., 2005; Proskurowski et al., 2008). Therefore, we conclude that no methane was produced from DIC or formate inside the Kynar bags.

Bags supplemented with  $^{13}\text{C}$ -enriched methane and containing fluids from Beehive, Sombrero, Calypso, Marker 8, and perhaps Marker 2 produced DIC with  $\delta^{13}\text{C}$  values significantly greater than that of DIC native to Lost City fluids (< 1.5 ‰) (Kelley et al., 2005; Proskurowski et al., 2008). Furthermore, all of the bags supplemented with  $^{13}\text{C}$ -enriched formate produced DIC with enriched  $\delta^{13}\text{C}$  values (**Supplemental Table S8**). Therefore, we conclude that methane and formate were oxidized to DIC in most of the Kynar sample bags.

Amplicon sequencing was conducted on each fluid sample separately, and no consistent differences in microbial community composition were detected among fluids collected in incubation bags compared to *in situ*-filtered Sterivex filters (**Supplemental Table S2**). Nevertheless, separate metagenome libraries were constructed with DNA prepared from *in situ*-filtered Sterivex filters and from incubation bags (**Supplemental Table S1**).

## **Extraction and purification of DNA and RNA**

Extraction of DNA from all Sterivex filters (including *in situ*-filtered fluids and fluids collected in Kynar bags and filtered shipboard) was conducted as described previously (Brazelton et al.,

2017; Thornton et al., 2020). The full protocol is available in the Zenodo-archived GitHub repository (DOI: 10.5281/zenodo.5798015) and on protocols.io (DOI: dx.doi.org/10.17504/protocols.io.bykqpuvw). Cell lysis was performed at 65 °C in a pH 10 extraction buffer followed by bead-beating with 0.1 mm glass beads. Lysates were purified with pH 8 phenol/chloroform/isoamyl alcohol extractions, precipitated overnight in ethanol, resuspended in low-EDTA TE, and further purified with magnetic beads (Rohland & Reich, 2012).

Total RNA was extracted from the Sterivex filters with a modification of the DNA extraction protocol optimized for RNA. The full protocol is available in the Zenodo-archived GitHub repository (DOI: 10.5281/zenodo.5798015) and on protocols.io (DOI: dx.doi.org/10.17504/protocols.io.bykspuwe). Cell lysis was performed at room temperature in a pH 7 extraction buffer, followed by bead-beating with 0.1 mm glass beads. Lysates were purified with pH 5.2 phenol/chloroform/isoamyl alcohol extractions, precipitated overnight in ethanol, resuspended in ultrapure water, and further purified with magnetic beads (Rohland & Reich, 2012). First-strand synthesis of cDNA was performed with SuperScript IV Reverse Transcriptase and random hexamers (Thermo Fisher).

### **Sequencing of 16S rRNA amplicons**

Sequencing of amplicons generated from 16S rRNA genes and cDNA was performed at the Genomics Core Facility at Michigan State University on an Illumina MiSeq instrument using dual-indexed Illumina fusion primers targeting the V4 region of the 16S rRNA gene (Kozich et al., 2013). PCR products were batch-normalized and pooled using Invitrogen SequalPrep DNA Normalization plates. The pool was cleaned up using AmpureXP magnetic beads and quality-controlled and quantified using a combination of Qubit dsDNA HS, Agilent 4200 TapeStation HSDNA1000, and Kapa Illumina Library Quantification qPCR assays. This pool was loaded onto

an Illumina MiSeq standard v2 flow cell, and sequencing was performed in a 2 x 250 bp paired-end format using a 500 cycle v2 reagent cartridge. Base calling was done by Illumina Real Time Analysis (RTA) v1.18.54, and the output of RTA was demultiplexed and converted to fastq format with Illumina Bcl2fastq v2.19.1.

### **Analysis of 16S rRNA ASVs**

Amplicon sequence variants (ASVs) were inferred from 16S rRNA amplicon sequences with DADA2 v. 1.10.1 (Callahan et al., 2016) after removal of primer sequences with cutadapt v. 1.15 (Martin, 2011). Sequences from different sequencing runs were processed with DADA2 separately, and then the resulting ASVs from all sequencing runs were merged to form a final, non-redundant count table of ASVs. Potential contaminants were removed with the decontam package (Davis et al., 2018) using both the “frequency” and “prevalence” modes. In “frequency” mode, ASVs were removed if their counts were significantly correlated ( $p < 0.02$ ) with DNA concentration (as measured by Qubit fluorometric quantification, Thermo Fisher). In “prevalence” mode, ASVs were removed if they were significantly more likely ( $p < 0.02$ ) to be present in one of three likely contamination sources (ambient lab air, surface seawater, or DNA extraction blanks) than in any sample of hydrothermal fluid. Samples of ambient air in the R/V *Atlantis* shipboard laboratory and our laboratory at the University of Utah were obtained as previously described (Motamedi et al., 2020). Samples of surface seawater were obtained during a previous study at the same location above the Atlantis Massif (Motamedi et al., 2020) and also from the shipboard laboratory “tap” water produced from surface seawater by the R/V *Atlantis*. DNA extraction blanks were obtained by Motamedi et al. by subjecting sterile Sterivex filters to the DNA extraction protocol described above. All contamination control samples were sequenced on the same sequencing runs as described above for the hydrothermal fluid samples. A total of 1,823 ASVs (9% of all ASVs) representing 109,824 sequence counts (2% of

all sequence counts) were removed from downstream analyses by decontam. These removed ASVs are provided in **Supplemental Table S2**.

Taxonomic classification of all ASVs was performed with DADA2 using the SILVA reference alignment (SSURefv132) and taxonomy outline (Quast et al., 2013; Yilmaz et al., 2014). Bubble plots of ASVs were drawn with ggplot2 using proportional abundances. The ordination plot was generated with phyloseq v1.26.1 (McMurdie & Holmes, 2013) using an unconstrained NMDS ordination of Morisita-Horn dissimilarity values. The stress of the fit to two dimensions was 0.05 after 20 tries without a convergent solution. Very similar results were produced with MDS and CCA ordinations and with Bray-Curtis dissimilarity values.

Differential abundances of ASVs between the Marker 3 and Camel Humps locations were calculated with DESeq2 (Love, Huber & Anders, 2014) as implemented by phyloseq (McMurdie & Holmes, 2013). ASVs with variance  $<10^{-5}$  were filtered out prior to the test, and those ASVs with an adjusted p-value  $<0.05$  were considered to be differentially abundant.

### **Sequencing of metagenome libraries**

Metagenome libraries were constructed with size-selected, sonicated DNA fragments of 500-700 bp with the NEBnext Ultra DNA II library kit for Illumina (E7645S), as previously described (Thornton et al., 2020). Paired-end sequencing (2 x 125 bp) of metagenomic libraries was conducted at the University of Utah High-Throughput Genomics Core Facility at the Huntsman Cancer Institute with an Illumina HiSeq2500 platform. Sequencing libraries (25 pM) were chemically denatured and applied to an Illumina HiSeq v4 paired end flow cell using an Illumina cBot. Hybridized molecules were clonally amplified and annealed to sequencing primers with reagents from an Illumina HiSeq PE Cluster Kit v4-cBot. Following transfer of the flowcell to an

Illumina HiSeq 2500 instrument (HCS v2.2.38 and RTA v1.18.61), a 125 cycle paired-end sequence run was performed using HiSeq SBS Kit v4 sequencing reagents (FC-401-4003).

### **Sequencing of metatranscriptome libraries**

Only two samples (one from Marker 2 and one from Sombrero) contained sufficient total RNA to attempt metatranscriptome sequencing. The two metatranscriptome libraries were constructed and sequenced by the University of Utah High-Throughput Genomics Core Facility at the Huntsman Cancer Institute. Total RNA was hybridized with NEBNext rRNA Depletion Solution Bacteria (E7850L) to substantially diminish rRNA from the samples. Stranded RNA sequencing libraries were prepared using the NEBNext Ultra II RNA Library Prep Kit for Illumina (E7770L). Purified libraries were qualified on an Agilent Technologies 2200 TapeStation using a D1000 ScreenTape assay. The molarity of adapter-modified molecules was defined by quantitative PCR using the Kapa Biosystems Kapa Library Quant Kit. Individual libraries were normalized to 10 nM, and equal volumes were pooled. Sequencing libraries were chemically denatured and applied to an Illumina NovaSeq flow cell using the NovaSeq XP workflow (20043131). Following the transfer of the flowcell to an Illumina NovaSeq 6000 instrument, a 150 cycle paired-end sequence run was performed using a NovaSeq 6000 S4 reagent Kit v1.5 (20028312).

### **Quality control and taxonomic classification of metagenome and metatranscriptome sequences**

Demultiplexing and conversion of the raw sequencing base-call data were performed through the CASAVA v1.8 pipeline. Adapter sequences and PhiX were removed from all reads with BBDuk (part of the BBTools suite, v35.85 (Bushnell, Rood & Singer, 2017)). Quality trimming was performed with our seq-qc package (<https://github.com/Brazelton-Lab/seq-qc>) as previously described (Thornton et al., 2020). Each library yielded 22-192 million reads (for a total of 1.4 billion reads among all metagenome libraries) after these quality control steps, representing 58-

82% of the number of raw reads. The two metatranscriptome libraries yielded 249 million reads (Marker 2) and 467 million reads (Sombrero) after quality filtering, representing 82% and 88%, respectively, of the original raw reads. Ribosomal RNA sequences were identified in the two metatranscriptomes with SortMeRNA (Kopylova, Noé & Touzet, 2012), resulting in the removal of 77% of reads from the Marker 2 metatranscriptome and 90% of reads from the Sombrero metatranscriptome. Quality-filtered, unassembled reads were assigned taxonomy with Kaiju and the NCBI nr+euk database (Menzel, Ng & Krogh, 2016). Kaiju was unable to classify 46-86% of reads, with the lowest percentage of unclassified reads in metatranscriptomes and the highest percentage of unclassified reads in the two Marker 3 metagenomes and one Sombrero metagenome (**Supplemental Table S3**). An interactive Krona plot is provided in the Zenodo-archived GitHub repository (DOI: 10.5281/zenodo.5798015).

### **Metagenomic assembly**

Sequences from metagenomic libraries were assembled with Megahit v1.1.1 (Li et al., 2016), using kmers of 27 to 127. A pooled, “all fluids” Megahit assembly was performed with metagenomic reads from all 13 libraries (representing 37 fluid samples collected in Sterivex filters and Kynar bags; see **Supplemental Table S1**). Genes were predicted with Prodigal v2.6.3 (Hyatt et al., 2010) in meta mode. Predicted protein sequences were queried against the KEGG release 83.2 prokaryotes database with Diamond v0.9.14 (Buchfink, Reuter & Drost, 2021).

In addition to the “all fluids” Megahit assembly, chimney-specific assemblies were conducted with metaSPAdes v3.13.0 (Nurk et al., 2017) as implemented by the KBase platform (kb\_SPAdes v.1.2.4) (Arkin et al., 2018). A chimney-specific assembly with reads obtained from Marker 3 was performed with Megahit v1.1.1 because metaSPAdes repeatedly failed when it



attempted to assemble Marker 3 reads. An overview of the metagenomic analysis workflow is illustrated in **Supplemental Figure S3**.

### **Binning of Metagenome-Assembled Genomes (MAGs)**

Binning of metagenome-assembled genomes (MAGs) from the “all fluids” and Marker 3 Megahit assemblies was conducted with BinSanity using the Binsanity-lc workflow v0.2.6.2 and a minimum contig size of 3 kb (Graham, Heidelberg & Tully, 2017). Binning of MAGs from the chimney-specific metaSPAdes assemblies was conducted with MaxBin2 v2.2.4 (kb\_maxbin v.1.1.1) (Wu, Simmons & Singer, 2016), MetaBAT2 v2.2 (kb\_metabat v.2.3.0) (Kang et al., 2019), and DAS Tool v1.1.2 (kb\_das\_tool v.1.0.6) (Sieber et al., 2018). The ANME-1 MAG was reconstructed from a Megahit assembly of Calypso reads with MetaBAT2 and DAS Tool. MAGs were assigned taxonomic classifications with GTDB-Tk v1.5.1 (reference data version r202;(Chaumeil et al., 2020)). Individual contigs were assigned taxonomic classifications with MMseqs2 (Mirdita et al., 2021). Gene products were predicted by annotation with Prokka v1.14.5 and its default databases (Seemann, 2014) with further function prediction provided by GhostKOALA v2.2 (Kanehisa, Sato & Morishima, 2016). Protein identifications and predicted functions were supplemented by results from InterProscan 5 (v5.52-86.0) (Jones et al., 2014), FeGenie (Garber et al., 2020), and dbCAN2 (Zhang et al., 2018). Completeness and contamination of initial automated MAGs were estimated with CheckM v1.0.5 (Parks et al., 2015), and the completeness and redundancy of the final refined MAGs were estimated with anvi'o v7 (Eren et al., 2021). Selected MAGs were refined by re-assembly with metaSPAdes v3.11.1 using reads mapped to MAGs and contigs with matching taxonomic classifications, followed by manual curation in anvi'o v7 (Eren et al., 2015) according to the differential coverage of contigs among samples. The final, refined MAGs represent 0.02 – 0.84% of the total read coverage in each metagenome or metatranscriptome library.

## Coverages of genes, contigs, and MAGs

Quality-filtered metagenome and metatranscriptome sequences were mapped to the “all fluids” Megahit assembly with Bowtie2 v2.3.2 (Langmead & Salzberg, 2012). Bowtie2 mapping rates to the “all fluids” assembly were 67-71% for Camel Humps, 82-89% for Sombrero1, 60-71% for Sombrero2, 95% for Marker 3, 67-74% for Calypso, and 56-60% for Marker 2. Bowtie2 mapping rates to the “all fluids” assembly were 21% before rRNA removal and 14% after rRNA removal for the Marker 2 metatranscriptome and 83% before rRNA removal and 74% after rRNA removal for the Sombrero1 metatranscriptome. Artificial duplicates were removed from the bam files using the MarkDuplicates function of Picard Tools v2.17.8 (*Picard Toolkit*, 2019).

The coverage for each predicted protein was calculated as transcripts (or fragments) per million (TPM) with count\_features v1.3.0, part of our seq-annot package (Thornton et al., 2020). TPM is a proportional unit (multiplied by one million for convenience) that is normalized to the length of each predicted protein sequence as well as to the total library size. Coverages of both metagenomes and metatranscriptomes were calculated in the same way and reported in the same units (TPM), although the metagenomic coverages reflect fragments instead of transcripts (Thornton et al., 2020).

The coverage of each MAG was calculated as the weighted sum of the normalized, proportional coverages (in TPM) of its member contigs. The contig coverages were obtained by mapping all unassembled reads against the re-assembled, refined MAGs with Bowtie2 and then calculating the average coverage per position of each contig with the genomecov command (with the option -pc) in bedtools v2.30.0 (Quinlan & Hall, 2010). Normalized coverages in units of TPM were calculated by dividing the average coverage per position by the total number of read pairs for that library.

## Phylogenetic Analyses

Phylogenetic trees of 16S rRNA genes were constructed with alignments obtained from the SILVA alignment server (Pruesse, Peplies & Glöckner, 2012) and RAxML v8.2.11 (Stamatakis, 2014) using the gamma model of rate heterogeneity and the “-f a” option to construct 100 rapid bootstrap searches and 20 maximum likelihood searches. The *Thermodesulfovibrionales* ASV from the Samail Ophiolite, Oman (Rempfert et al., 2017) was generated with DADA2 as described above with reads accessed via SRA accession SRR5000240. Phylogenies of McrA, GrdB, and carbonic anhydrase were constructed with Clustal Omega alignments (Madeira et al., 2019) and RAxML as described above but with automated selection of the rate model. All sequences and accessions are provided in the Zenodo-archived GitHub repository accessible via DOI: 10.5281/zenodo.5798015.

---

## SUPPLEMENTAL RESULTS:

### DETAILED COMPARISONS OF HYDROTHERMAL FLUID SAMPLES

#### Markers 3 and C

The strong similarity of community compositions from Markers 3 and C is remarkable considering that fluids from Marker 3 were sampled at much lower temperatures (<20 °C) than those from Marker C (up to 80 °C). Temperatures exceeding 55 °C were measured at Marker 3 during the same dive, but the temperatures of the fluid samples (i.e. measured in-line, during sampling) used for DNA-based analyses were lower, most likely due to cooling of the fluids as they exit the chimney. In addition, the pH, sulfate, sulfide, and magnesium levels of Marker 3 fluids are more similar to those of seawater, compared to Marker C fluids. These results suggest that the cooler fluids venting from Marker 3 experience more dilution with ambient seawater compared with the warmer fluids venting from Marker C, but that they share a common subsurface source, which may explain the similar microbial compositions.

284

285 Marker 3 fluids were rich in metagenomic sequences classified as family *Methanosarcinaceae*,  
286 which includes the dominant archaeal phylotype previously detected in Lost City chimneys  
287 (Schrenk et al., 2004; Brazelton et al., 2010, 2011). Archaeal sequences were much more  
288 abundant in metagenomes than in the 16S rRNA amplicon libraries. For example,  
289 *Methanosarcinaceae* is the most abundant family in Marker 3 metagenomic reads (1-2% of all  
290 reads and 20-21% of all reads that could be classified to the family level by Kaiju;  
291 **Supplemental Table S3**), even though *Methanosarcinaceae* ASVs have lower relative  
292 abundance than several bacterial ASVs in the same samples, suggesting a bias against  
293 archaeal sequences in the ASV dataset.

294

295 Venting fluids collected near Markers C and 3 were particularly enriched in taxa representing  
296 potential sulfate-reducing bacteria (SRB). *Thermodesulfovibrionia* and *Desulfotomaculum* ASVs  
297 dominated these fluids, but they were much less abundant in the RNA fraction of the Marker C  
298 sample (**Figure 3**). ASVs classified as *Desulfocapsa* were abundant in Marker 3 fluids, present  
299 in the RNA fraction from Marker C, and absent in the DNA fraction from Marker C.  
300 *Desulfobulbus* ASVs were absent in both Markers C and 3 except for the RNA fraction from  
301 Marker C, and they were prevalent in Sombrero fluid samples. *Desulfobulbus* and *Desulfocapsa*  
302 both belong to the family *Desulfobulbaceae*, which had 30-fold less coverage in Marker 3  
303 metagenomes than *Methanosarcinaceae* and 8-fold less coverage than the *Nitrospiraceae*  
304 family that includes *Thermodesulfovibrio* (**Supplemental Table S3**).

305

306 Marker 3 metagenomes were also distinctive in their high proportions of Candidatus  
307 Patescibacteria (**Supplemental Figure S2**). Patescibacteria were rare in all samples except for  
308 Marker 3 fluids, where they represented 33-36% of all classified sequences (**Supplemental**  
309 **Table S3**). Hundreds of ASVs were classified as Patescibacteria (predominantly Paceibacteria

and Gracilibacteria), but they did not represent a large fraction of the counts in any of the fluid samples (**Supplemental Table S2**).

### **Camel Humps and Sombrero**

Fluids venting from Camel Humps contained a remarkably even distribution of ASVs that included *Sulfurovum*, *Sulfurospirillum*, and *Thiomicrothabodus* at similar abundances as taxa typically associated with ambient seawater (e.g. *Alteromonas*, *Roseobacter*, *Halomonas*). This community composition strongly contrasts with that of Marker 3, even though both locations are at the summit of the central Poseidon structure (**Figure 1**). Of the 100 ASVs most common in Camel Humps fluids, 70 of these were significantly less abundant in Marker 3 fluids (**Supplemental Table S2**). The differences between Camel Humps and Marker 3 cannot be explained by mixing of the same hydrothermal fluid with ambient seawater because the bacteria at Camel Humps belong to lineages associated with hydrothermal environments (e.g., *Sulfurovum*, *Sulfurospirillum*, *Thiomicrothabodus*), not ambient seawater. Moreover, the temperatures measured during the sampling of Camel Humps fluids were higher than those of Marker 3 fluids, which is the opposite of what would be expected if dilution by seawater had been responsible for the observed trends. Instead, these results suggest a difference in the subsurface source of the hydrothermal fluids venting from these locations. Although Marker 3 and Camel Humps are located next to each other, they are visually distinct chimney structures that could potentially host fluids venting from distinct subsurface sources (**Supplemental Figure S2**). Alternatively, the distinctive assemblage of taxa at Camel Humps may represent biofilm communities that were flushed into nearby venting fluids, swamping the less dense organisms derived from the subsurface. The two explanations are not mutually exclusive, and both indicate that the fluids collected from Camel Humps during this study are not representative of the subseafloor.

The Sombrero chimney is located on a ridge stretching from the main vent field toward the eastern wall of the Atlantis Massif (**Figure 1**), and it was sampled on two separate ROV *Jason* dives, with the second dive (Sombrero2) collecting fluids with higher temperatures and lower sulfate concentrations than the first dive (Sombrero1) (**Table 1**). Despite these different measurements at the time of sampling, the overall microbial composition was remarkably consistent across all Sombrero samples ranging in temperature from 10 – 74 °C (**Figure 1**). Minor differences are nevertheless visible (**Figure 2**); for example, *Thiomicrothrix* and *Sulfurospirillum* dominated Sombrero fluids with lower temperature, higher sulfate, and lower sulfide. Warmer and more sulfidic Sombrero fluids included greater proportions of taxa that were also abundant in fluids from Markers 3 and C (**Figure 2**).

#### **Markers 2 and 8**

The most sulfidic fluids were collected from chimneys near Marker 2 and Marker 8, which are located on the western edge of the Lost City hydrothermal field (**Table 1**; **Figure 1**; **Supplemental Figure S1**), and their microbial compositions were distinct from all other fluids. One of the fluid samples from Marker 2 was dominated by a single ASV identical to the 16S rRNA gene of *Alteromonas macleodi* (**Supplemental Table S2**), a ubiquitous marine bacterium (Koch et al., 2020), suggesting substantial dilution of the sample with ambient seawater. Other samples from Marker 2 also contained bacteria that have been previously associated with sulfur oxidation in Lost City chimney biofilms, including the genera *Sulfurovum*, *Sulfurospirillum*, and *Thiomicrothrix* (Brazelton et al., 2010; Brazelton & Baross, 2010; McGonigle, Lang & Brazelton, 2020). Therefore, they are likely to be adapted to chimney habitats where sulfidic hydrothermal fluids are mixing with oxic seawater, and they are probably not representative of microbial communities inhabiting anoxic, subsurface environments. The ribosomal RNA fraction of Marker 2 fluids contained elevated relative abundances of a *Sulfurospirillum* ASV and a *Thiomicrothrix* ASV (**Figure 2**). Microbial taxa detected in fluids collected near Marker 8

were broadly similar to those of Marker 2 fluids, except that they were dominated by one *Sulfurovum* ASV that was rare in Marker 2 fluids (**Figure 2**).

## **Calypso**

The Calypso chimney sits on the eastern wall of the Atlantis Massif, approximately 75 m from the large Poseidon edifice that dominates the Lost City field (**Figure 1**). The fluids venting from Calypso had higher sulfide concentrations than the fluids from Markers 3 and C, but less sulfide than in Markers 2 and 8 (**Table 1**). The overall microbial community structure resembled that of Sombrero fluids (**Figure 1**), although the ASVs with the highest counts in Calypso were also common in Markers 3 and C (**Figure 2**). The top *Thermodesulfovibrionia* ASV in Calypso fluids differed by one base from the ASV that dominated the fluids from Markers C and 3. This ASV was further enriched in the RNA fraction from Calypso, suggesting that it could have been metabolically active prior to sampling.

ASVs classified as the ANME-1b group of archaeal methanotrophs were most abundant in fluids from Calypso (**Figure 2**). ANME-1 sequences were rare or absent in almost all other fluid samples except for a few samples from Markers C and 3 and the RNA (but not DNA) fraction from one sample of fluids at Marker 2. These results are consistent with our previous studies that documented very high relative abundances of ANME-1 in cooler chimneys at the periphery of the field and trace levels of ANME-1 DNA in hot chimneys of the central Poseidon complex (Brazelton et al., 2006, 2010).

Calypso fluids were rich in Chloroflexi ASVs and metagenomic sequences primarily classified as class *Dehalococcoidia* (**Supplemental Tables S2-S3**), but ASVs classified as *Anaerolineae*, TK10, and KD4-96 were also present at low levels. The Chloroflexi MAG from Lost City biofilms previously reported in (McGonigle, Lang & Brazelton, 2020) is a member of the *Anaerolineae*.

*Dehalococcoidia* ASVs were generally most abundant in Calypso fluids, but most ASVs were broadly distributed (**Figure 2**).

---

## **SUPPLEMENTAL RESULTS:**

### ***DETAILED DESCRIPTIONS OF METAGENOME-ASSEMBLED GENOMES (MAGs)***

#### ***Methanosarcinaceae* MAG**

The predominant *Methanosarcinaceae* ASV was classified as genus *Methanosalsum* in the SILVA database, but our phylogenies of 16S rRNA and mcrA indicate that the Lost City *Methanosarcinaceae* are not monophyletic with any previously characterized genera (**Supplemental Figure S5**). Over 254 bases, the *Methanosarcinaceae* ASV is 94% similar to the 16S rRNA gene of *Methanococcoides methylutens* and 95.7% similar to that of *Methanohalophilus mahii*.

The *Methanosarcinaceae* MAG contained 1276 coding sequences (CDS) and was estimated to be 84% complete with 1% redundancy. It shares 99% average nucleotide identity (ANI) with the MAG we previously recovered from a Lost City biofilm (McGonigle, Lang & Brazelton, 2020). It is most abundant in Marker 3 fluids (**Figure 3**), and it is more abundant in those Sombrero fluids that show more mixing with seawater (lower temperature, more sulfate, less sulfide), consistent with the distribution of *Methanosarcinaceae* ASVs (**Figure 2**).

*Methanosarcinaceae* MAG-1276 has a remarkably low GC content (29%), which may help explain why the first metagenomic studies of Lost City biofilms recovered surprisingly few archaeal sequences (Brazelton & Baross, 2009, 2010). It encodes the core pathway for methanogenesis from carbon dioxide (**Supplemental Table S5**). It contains predicted



sequences for F<sub>420</sub>-reducing hydrogenase (FrhAB), which is required by all methanogens that reduce carbon dioxide with H<sub>2</sub> (Mand & Metcalf, 2019). In addition, two genes predicted to encode subunits of Ech hydrogenase (EchCE) are present. FrhB and EchCE are located on the same contig as a multicomponent Na<sup>+</sup>:H<sup>+</sup> antiporter (MrpACBD) and the MAG's only gene annotated as a subunit of NADH-quinone oxidoreductase (NuoH).

The MAG encodes AMP-forming acetyl-CoA synthetase, as previously reported for *Methanosarcinaceae* in Lost City biofilms (McGonigle, Lang & Brazelton, 2020), which may enable acetoclastic methanogenesis. It also encodes MtaA and MtaC (**Supplemental Table S5**), two of the three proteins that enable the use of methanol as a substrate for methanogenesis in some methanogens.

*Methanosarcinaceae* MAG-1276 contains a complete 14-gene cluster (mbhA-N) encoding membrane-bound hydrogenase (**Figure 5; Supplemental Figure S8**). The same gene cluster, with conserved synteny, is also found in methanogens belonging to the order *Methanomicrobiales* (e.g. *Methanospirillum hungatei*) and in heterotrophs of the order *Thermococcales* (e.g. *Thermococcus kodakarensis*) (Thauer et al., 2010). The mbhL subunits from these methanogens have only 42-45% identities with the Lost City mbhL sequences reported here, which have greater similarity (~49% identities) to mbhL sequences from *Thermococcus*.

As we reported previously for chimney biofilms (Lang et al., 2018), the *Methanosarcinaceae* MAG encodes a formate dehydrogenase (FDH) that is similar to that of *Methanobrevibacter* species, which are unable to use formate as a carbon source. The contig encoding this FDH also encodes one subunit of F<sub>420</sub>-non-reducing hydrogenase (MvhD) and heterodisulfide reductase (HdrABC).

440

441 We identified additional, divergent FDH-like sequences associated with *Methanosarcinaceae*,  
442 ANME-1, *Thermodesulfovibrionales*, and *Bipolaricaulota* (**Figure 6; Supplemental Figure S9**).

443 These divergent FDH-like sequences have not been previously described, though they are  
444 present in a few published genomes. The *Methanosarcinaceae* FDH-like sequence (**Figure 5**)  
445 shares 63% amino acid identities with a predicted protein in the *Desulfitibacter alkalitolerans*  
446 genome. In a previous study, we reported these divergent FDH-like sequences as bacterial  
447 based on their similarity to *Desulfitibacter* sequences (Lang et al., 2018). Although  
448 *Desulfitibacter alkalitolerans* can use formate as an electron donor, it encodes two additional  
449 homologs of *fdhA* that are not present in any Lost City metagenomes. A similar FDH-like  
450 sequence is encoded by *Dethiobacter alkaliphilus*, which is unable to grow on formate as its  
451 sole carbon source (Sorokin et al., 2008). The phylogeny in **Figure 5** is rooted with FdhF  
452 (anaerobic formate hydrogen lyase) from *Methanothermobacter thermautotrophicus* and  
453 *Methanothermobacter marburgensis*. These two species are unable to grow on formate (Kaster  
454 et al., 2011), and they lack a separate *fdhABC* operon that is found in *M. thermautotrophicus* st.  
455 CaT2, which can grow on formate.

456

457 No known transporters for formate were identified in this MAG, but one gene was predicted to  
458 encode a member of the oxalate:formate antiporter family. In *Oxalobacter formigenes*, this  
459 transporter enables uptake of oxalate with export of formate (Abe, Ruan & Maloney, 1996). Its  
460 presence was also observed in a formate-utilizing methanogen, which lacks any other formate  
461 transporters, in hyperalkaline groundwaters of the Samail Ophiolite (Fones et al., 2021).

462 Potentially homologous oxalate:formate antiporters were also identified in two NPL-UPA2 MAGs  
463 and two *Natronincolaceae* MAGs from Lost City fluids. The four bacterial sequences shared 59-  
464 63% amino acid identities with the *Methanosarcinaceae* sequence.

465

## ***Methanocellales* MAG**

A MAG classified as order *Methanocellales* (838 CDS; 84% complete with 1% redundancy) was only present in Calypso fluids (**Figure 3**). Curiously, no ASVs were classified as *Methanocellales*, and no previous studies have identified any methanogen taxa in Lost City samples other than *Methanosarcinaceae*.

*Methanocellales* MAG-838 encodes the key enzyme CODH/ACS, but it has an incomplete pathway for methanogenesis, including only four of the first five steps (FwdABCDGF, Ftr, Mtd, and Mer) and lacking all other steps, including the proteins required for methane production (methyl-coenzyme M reductase and heterodisulfide reductase) (**Supplemental Table S5**). Evidence for acetate utilization includes a predicted sequence for AMP-forming acetyl-CoA synthetase, which shares 73% amino acid identities with the homolog in the *Methanosarcinaceae* MAG, and a cation/acetate symporter (ActP). The MAG includes one [NiFe] hydrogenase and the Rnf complex (RnfCDGEAB), an energy-conserving ferredoxin:NAD<sup>+</sup>-oxidoreductase (Biegel et al., 2011).

## **ANME-1 MAG**

A MAG classified as ANME-1 (1099 CDS; 88% complete with 4% redundancy) was most abundant in Calypso fluids as well as the Marker 2 metatranscriptome, even though ANME-1 was very rare in the Marker 2 metagenomes. As expected for ANME-1 archaea, the Lost City ANME-1 MAG contains the core methanogenic pathway (**Supplemental Table S5**). The absence of cytochromes and presence of hydrogenases in this MAG was noted by (Chadwick et al., 2021) as consistent with the genomic features of the so-called “freshwater” clade of ANME-1, for which the genus “Candidatus Methanoalium” was proposed. One of the shared features within this clade, including the Lost City ANME-1 MAG, is a novel HdrABC-MvhADG complex, which is involved in the transfer of electrons derived from H<sub>2</sub> in methanogens.

492

493 The Lost City ANME-1 MAG contains the core methanogenic pathway (**Supplemental Table**  
494 **S5**), including F<sub>420</sub>-dependent methylenetetrahydromethanopterin reductase (Mer). This gene is  
495 required for methanogenesis from carbon dioxide, but it is typically absent in ANME genomes,  
496 with at least one exception previously reported (Beulig et al., 2019). The MAG lacks all but one  
497 of the subunits of N<sup>5</sup>-methyl-H<sub>4</sub>MPT:coenzyme M methyltransferase (Mtr), which catalyzes the  
498 penultimate step of methanogenesis (and putatively the second step of reverse  
499 methanogenesis). It is present in most but not all ANME-1 genomes (Chadwick et al., 2021).

500

501 The ANME-1 MAG-1 also encodes the complete mbhA-N gene cluster for membrane-bound  
502 hydrogenase (**Figure 5; Supplemental Figure S8**), and each predicted gene in the cluster has  
503 the greatest similarity to the homolog in the *Methanosarcinaceae* MAG than to any other  
504 sequences in public databases. It lacks any established genes for FDH, but it includes a  
505 divergent FDH-like sequence that appears to be homologous to those found in other Lost City  
506 MAGs (**Figure 6; Supplemental Figure S9**). Distant homologs of FDH have been previously  
507 identified in ANME genomes, but they do not share significant sequence similarity with the FDH-  
508 like sequences reported here.

509

#### 510 **Bipolaricaulota MAGs**

511 ASVs classified as Bipolaricaulota (named Acetothermia in the SILVA database) clustered into  
512 three distinct clades (**Supplemental Figure S6**) that correspond to the taxonomic classifications  
513 of three distinct Bipolaricaulota MAGs. The most abundant of the three Bipolaricaulota MAGs  
514 (1207 CDS; 86% complete with 0% redundancy) was classified by GTDB as species UBA7950  
515 within class Bipolaricaulia, and it shares 99% ANI with the GTDB reference genome, which was  
516 assembled by (Parks et al., 2021) with sequences from a previously published Lost City biofilm  
517 metagenome (Lang et al., 2018). Bipolaricaulota MAG-1207 is most abundant in Sombrero and

Calypso, especially the Sombrero metatranscriptome (**Figure 3**). It encodes a nearly complete glycolysis pathway, an incomplete TCA cycle, an incomplete Wood-Ljungdahl pathway, the Rnf complex, and pyruvate formate-lyase (PflD) (**Supplemental Table S5**).

A second *Bipolaricaulota* MAG (1260 CDS; 89% complete with 1% redundancy) was classified as family UBA9294 within order UBA7950, and it shares only 77% ANI with the closest reference in GTDB (**Supplemental Table S4**). *Bipolaricaulota* MAG-1260 has a similar complement of genes as in MAG-1207, including an incomplete Wood-Ljungdahl pathway and the Rnf complex, though it lacks pyruvate formate-lyase and glycine reductase.

Neither of these *Bipolaricaulota* MAGs contains any known hydrogenases or formate dehydrogenases. A divergent FDH-like sequence that appears to be homologous with those in the *Methanosarcinaceae*, ANME-1, and *Thermodesulfobibrionales* MAGs was observed in multiple initial BinSanity bins classified as *Bipolaricaulota* (**Figure 6**, but it was not included in the three re-assembled, refined *Bipolaricaulota* MAGs. The most similar sequence in the NCBI NR database was from a MAG assembled by (Parks et al., 2021) from a Voltri Massif serpentinite spring (Brazelton et al., 2017).

A third *Bipolaricaulota* MAG (1503 CDS; 92% complete with 1% redundancy) was classified as genus UBA3574 within family Bipolaricaulaceae. It encodes a nearly complete Wood-Ljungdahl pathway, including the NADP-dependent formate dehydrogenase (FdhA) that is typical of acetogens. It does not encode the Rnf complex, but it contains genes encoding all subunits of the heterodisulfide reductase complex MvhAGD-HdrABC. It encodes a [NiFe] hydrogenase (HoxYH) and at least a partial gene cluster for membrane-bound hydrogenase, including the large catalytic subunit MbhL (**Figure 5; Supplemental Figure S8**). The predicted MbhL sequence is most closely related to two *Bipolaricaulota* MAGs from hydrothermal systems: the

Mid-Cayman Rise (Zhou et al., 2020) and Guaymas Basin (Dombrowski, Teske & Baker, 2018). Bipolaricaulota MAG-1503 also encodes an oxalate:formate antiporter distinct from those found in the *Methanosarcinaceae*, NPL-UPA2, and *Natrinincolaceae* MAGs (31-33% amino acid identities).

#### ***Thermodesulfovibrionales* MAG**

Two ASVs classified as *Thermodesulfovibrionia* (a class within phylum Nitrospirae) differed from each other by a single base. One of these was the top ASV in each of the fluid samples from Markers 3 and C (8-24% of all sequences), while the second ASV dominated the Calypso fluids (7-17% of all DNA fractions and 27% of the RNA fraction). Lost City *Thermodesulfovibrionia* metagenomic sequences and ASVs were notably rare in the low-sulfide fluids from Camel Humps and the high-sulfide fluids from Marker 2 (**Figures 2-3**).

Accordingly, a MAG (1293 CDS; 92% complete with 0% redundancy) classified as order *Thermodesulfovibrionales* within class *Thermodesulfovibrionia* was most abundant in Marker 3 and Calypso fluids. For heterotrophic metabolism, it encodes a nearly complete glycolysis pathway and TCA cycle, plus genes for lactate dehydrogenase, pyruvate ferredoxin oxidoreductase (PorABCD), and an oxalate:formate antiporter distinct from those found in the *Methanosarcinaceae*, NPL-UPA2, and *Natrinincolaceae* MAGs (21-24% amino acid identities),

*Thermodesulfovibrionales* MAG-1293 has a partial Wood-Ljungdahl pathway and a monomeric CO dehydrogenase (CooS) with two maturation factors (CooF and CooC). It encodes the NAD(P)-dependent formate dehydrogenase typical of acetogens (FdhA), as well as aerobic formate dehydrogenase (FdoG) and the divergent FDH-like sequence also observed in other Lost City MAGs (**Figure 6**). It encodes a [NiFe]-hydrogenase (HyaAB) classified as [NiFe] Group 1c, a group of respiratory H<sub>2</sub>-uptake hydrogenases that can use fumarate, sulfate, or

metals as terminal electron acceptors. The MAG contains only one subunit of membrane-bound hydrogenase (MbhJ), which shares 43-49% amino acid identities with the MbhJ sequences of the other Lost City MAGs shown in **Figure 5**. The key genes required for nitrogen fixation (nifHDK) and dissimilatory sulfate reduction (dsrAB) are present.

### ***Desulfotomaculum* MAGs**

Two MAGs classified as family *Desulfotomaculaceae* were resolved based on their distinct coverage patterns. Both MAGs were estimated to be 94% complete with 0% redundancy, but one had more predicted genes (1580) than the other (1144). *Desulfotomaculaceae* MAG-1580 was only abundant in one sample from Sombrero, while MAG-1144 was abundant in Marker 3, Sombrero, and Calypso fluids (**Figure 3**). Several ASVs were classified as genus *Desulfotomaculum* within the family *Desulfotomaculaceae*. One of these was most abundant in Marker 3 (up to 14% of all sequences) and Calypso (2-8% of all sequences), roughly matching the distribution of MAG-1144, while another ASV (differing from the first by four bases) was only abundant in Sombrero fluids, similar to MAG-1580 (**Figure 2**).

Both *Desulfotomaculum* MAGs encode a complete or nearly complete glycolysis pathway and at least two genes of the TCA cycle (malate dehydrogenase and isocitrate dehydrogenase). In addition, *Desulfotomaculum* MAG-1144 has succinate dehydrogenase (SdhABC) and the beta subunit of fumarate dehydratase. *Desulfotomaculum* MAG-1580 encodes pyruvate ferredoxin oxidoreductase (PorABCD) and pyruvate formate-lyase (PflD). Both MAGs have a cation/acetate symporter (ActP), and MAG-1580 encodes a phosphonate transporter (PhnCDE).

Both MAGs have incomplete Wood-Ljungdahl pathways that lack the key enzyme CODH/ACS. Both MAGs also have monomeric CO dehydrogenase (CooS) and aerobic formate dehydrogenase (FdoG). No [NiFe] hydrogenases were detected, and only one subunit of [FeFe]

hydrogenase (HndC) was present in the *Desulfotomaculum* MAGs. This hydrogenase is capable of H<sub>2</sub> oxidation with reduction of NADP in some organisms (Kpebe et al., 2018), but the presence of only one subunit in multiple Lost City MAGs (**Supplemental Table S5**) is curious and has unknown implications for the ability of these organisms to metabolize H<sub>2</sub>.

The two *Desulfotomaculum* MAGs, in addition to the *Thermodesulfovibionales* MAG, are the only MAGs reported here that encode dissimilatory sulfite reductase (DsrAB). *Desulfotomaculum* dsrAB sequences were most abundant in Sombrero fluids, while *Thermodesulfovibionales* dsrAB sequences were most abundant in Marker 3 and Calypso fluids. The two *Desulfotomaculum* MAGs also have the key genes required for nitrogen fixation (nifHDK).

#### ***Natronincolaceae* MAGs**

Two MAGs were classified as family *Natronincolaceae* within the Clostridia and shared 81-89% ANI with a MAG reconstructed from seafloor borehole fluids at North Pond (Tully et al., 2018). Other genera within the family *Natronincolaceae* include *Alkaliphilus* and *Serpentinicella*, which have been isolated from the Prony Bay hydrothermal field (Mei et al., 2016; Postec et al., 2021). The coverage of the *Natronincolaceae* MAGs was primarily in Sombrero fluids. The two MAGs shared 80% ANI but showed a few potentially important differences in their genomic inventories.

*Natronincolaceae* MAG-2163 (2163 CDS; 90% complete with 3% redundancy) has one of the largest genomes in this study, and it has an incomplete glycolysis pathway and incomplete TCA cycle. It encodes at least three steps of the Wood-Ljungdahl pathway, monomeric CODH (CooS), the Rnf complex, and the electron carriers HdrABC and MvhD.



In contrast, *Natronincolaceae* MAG-1138 (1138 CDS; 75% complete with 0% redundancy) has a smaller genome, a nearly complete glycolysis pathway, and no TCA cycle. It includes two genes associated with the Wood-Ljungdahl pathway, but not CODH, Rnf, Hdr, or MvhD. It does include acetate kinase (AckA) and phosphotransacetylase (Pta). It lacks any genes for ATP synthase, suggesting an obligate fermentative lifestyle.

*Natronincolaceae* MAG-2163 has [FeFe] hydrogenase (HndBCD), while MAG-1138 only has the HndC subunit. Both *Natronincolaceae* MAGs have at least one subunit of pyruvate dehydrogenase and pyruvate ferredoxin oxidoreductase, though they differ in which subunits they include (**Supplemental Table S5**).

### ***Dehalococcoidia* MAGs**

The most abundant *Dehalococcoidia* MAG (844 CDS; 73% complete with 0% redundancy) was classified as family SpSt-899 within order SZUA-161. It was one of the highest coverage MAGs in Calypso fluids as well as the Sombrero and Marker 2 metatranscriptomes. Unlike the other two *Dehalococcoidia* MAGs described below, glycolysis and the TCA cycle are incomplete, and there are no genes required for the degradation of large organic compounds. However, the presence of pyruvate:ferredoxin oxidoreductase, pyruvate formate-lyase, and an oxalate/formate antiporter suggest the ability to ferment low-molecular-weight organic compounds. In addition, the MAG has two steps of the Wood-Ljungdahl pathway (FchA and MetF), the Rnf complex, and a complete set of genes encoding the key enzyme CODH/ACS. No hydrogenases or formate dehydrogenases are present. Curiously, *Dehalococcoidia* MAG-844 has V(A)-type ATP synthase, which is typically associated with archaea, but has also been observed in Chloroflexi and Parcubacteria in another serpentinite-hosted spring (Suzuki et al., 2017).

Two additional *Dehalococcoidia* MAGs belong to the SAR202 cluster, one of which was classified by GTDB as order SAR202 and one as order UBA3495. Each MAG has 98% ANI with previously published marine Chloroflexi MAGs (**Supplemental Table S4**). *Dehalococcoidia* MAG-2669 (2669 CDS; 86% complete with 4% redundancy) has moderately high coverage in all chimney fluids except for Marker 3. *Dehalococcoidia* MAG-2875 (2875 CDS; 87% complete with 1% redundancy) was less abundant than MAG-2669 in all samples but otherwise exhibited a similar distribution pattern. Both MAGs encoded nearly complete glycolysis pathways and TCA cycles, and they have a variety of genes associated with the oxidation of various organic molecules, consistent with previous studies of marine Chloroflexi (Liu et al., 2021; Landry et al.). Both MAGs encode cytochrome c oxidase (CoxA), which represents the only evidence for aerobic respiration in any of the final, refined MAGs. *Dehalococcoidia* MAG-2875 has both (aerobic) pyruvate dehydrogenase and (anaerobic) pyruvate:ferredoxin oxidoreductase, as reported for other SAR202 genomes (Mehrshad et al., 2018), while MAG-2669 has neither. Neither MAG encodes any FDH or hydrogenases. MAG-2669 is predicted to encode sulfite reductase (Sir), adenylylsulfate reductase (AprAB), and sulfate adenylyltransferase (Sat), but dissimilatory sulfite reductase (DsrAB) is not present. Both MAGs have genes associated with the metabolism of organosulfur compounds, similar to those reported by (Mehrshad et al., 2018) for deep-sea SAR202 genomes (**Supplemental Table S5**).

All three *Dehalococcoidia* MAGs include genes for glycine reductase, thioredoxin, and selenocysteine synthesis. GrdB (beta subunit of glycine reductase) sequences from the two Lost City *Dehalococcoidia* MAGs that belong to the SAR202 marine cluster (MAG-2669 and MAG-2875) are distinct from the GrdB of *Dehalococcoidia* MAG-844 and from the GrdB of all other Lost City MAGs (**Supplemental Figure S11**).

## **NPL-UPA2 MAGs**

673 Three MAGs classified as candidate phylum NPL-UPA2 (new name *Candidatus Horikoshi*  
674 bacteria proposed by (Suzuki, Nealson & Ishii, 2018) were recovered, each only 55-72%  
675 complete and with distinct coverage patterns among the hydrothermal fluid samples. No ASVs  
676 or unclassified reads were classified as NPL-UPA2, as this group was not yet represented in the  
677 SILVA or Kaiju databases. NPL-UPA2 MAG-914 (914 CDS; 62% complete with 0%  
678 redundancy) and MAG-1083 (1083 CDS; 55% complete with 3% redundancy) were most  
679 abundant in Calypso fluids and nearly absent in all other locations. MAG-718 (718 CDS; 72%  
680 complete with 3% redundancy) was most abundant in Marker 3 fluids and exhibited a nearly  
681 inverse abundance distribution compared to the other two MAGs. All three NPL-UPA2 MAGs  
682 have incomplete Wood-Ljungdahl pathways that lack the first few steps and begin with  
683 methylene-THF dehydrogenase (FolD), and all three MAGs encode V(A)-type ATP synthase  
684 and the Rnf complex, as reported by (Suzuki, Nealson & Ishii, 2018).

685

686 In most other respects the two NPL-UPA2 MAGs that are prominent in Calypso fluids differ from  
687 the MAG that is most abundant in Marker 3 fluids. The Calypso MAGs encode multiple subunits  
688 of the CODH/ACS enzyme, while MAG-718 has only the methyltransferase subunit (AcsE).  
689 MAG-1083 also includes acetate kinase (AckA), phosphotransacetylase (Pta), and a  
690 cation/acetate symporter (ActP). In contrast, MAG-718 is the only NPL-UPA2 MAG that  
691 encodes pyruvate formate-lyase (PflD), the oxalate/formate antiporter, and carbonic anhydrase.

692

693 Evidence for hydrogenases in the NPL-UPA2 MAGs is lacking. Membrane-bound hydrogenase  
694 (MbhL) was identified in two of the MAGs by GhostKoala, but Prokka annotated these  
695 sequences as NAD(P)H-quinone oxidoreductase and formate hydrogenlyase. Neither sequence  
696 could be placed in the mbhL phylogeny of **Figure 5**. MAG-1083 includes one subunit of F<sub>420</sub>-  
697 reducing hydrogenase (FrhB) and one subunit of [FeFe] hydrogenase (HndD), but the functional  
698 roles of these predicted proteins are unclear.

699

700 Unlike the NPL-UPA2 MAG reported by (Suzuki, Nealson & Ishii, 2018), none of the Lost City  
701 NPL-UPA2 MAGs have formate dehydrogenase nor the electron carriers Hdr and Etf.

702

### 703 **Patescibacteria MAGs**

704 Several MAGs classified as candidate phylum Patescibacteria were represented by two classes:  
705 Paceibacteria and Gracilibacteria (according to GTDB taxonomy). Paceibacteria MAGs (55-86%  
706 completion with 0-3% redundancy) had small genomes (as low as 307 kb with 63% estimated  
707 completion) with low GC content (24-45%). Gracilibacteria MAGs (61-83% completion with 0-  
708 1% redundancy) had even lower GC content (22-34%) and somewhat larger genomes but with  
709 very low annotation success. As few as 8% of all coding sequences in Gracilibacteria MAGs  
710 yielded GhostKoala results. Two of the Paceibacteria MAGs were most abundant in Marker 3  
711 fluids, while Gracilibacteria MAGs were notably absent in Marker 3 fluids and were more  
712 abundant in Camel Humps and Sombrero fluids (**Figure 3; Supplemental Table S4**).

713

714 In general, Paceibacteria and Gracilibacteria MAGs included genes for the biosynthesis of key  
715 cellular components and basic information processing, but genes specific to catabolic pathways  
716 were rare. The few exceptions included a lactate transporter (LctP) that was present in the two  
717 highest-coverage Paceibacteria MAGs, and one of these MAGs also has lactate dehydrogenase  
718 (LdhA). In addition, acetate kinase (AckA), acylphosphatase (AcyP), and phosphoenolpyruvate  
719 synthase (PpsA) were present in one or more Paceibacteria MAGs (**Supplemental Table S5**).  
720 Propionyl-CoA carboxylase (PccB) was included in one Gracilibacteria MAG. All Paceibacteria  
721 MAGs, one of the Gracilibacteria MAGs, and almost all other MAGs in this study have a  
722 substrate-binding protein associated with the peptide/nickel transport system (K02035;  
723 **Supplemental Table S6**).

724

Curiously, methylene tetrahydrofolate dehydrogenase (FolD), which is part of the Wood-Ljungdahl pathway, was encoded by three Paceibacteria MAGs and one Gracilibacteria MAG. FolD was also observed in 17 MAGs from The Cedars, a serpentinite-hosted spring in California, that were classified as candidate phylum OD1 (Suzuki et al., 2017), now included within Paceibacteria in GTDB.

As mentioned in the main text, ATP synthase was completely absent in three of the Paceibacteria MAGs, and one Gracilibacteria MAG included only a single subunit. One Paceibacteria MAG encodes a V(A)-type ATP synthase instead of the F-type ATP synthase present in the other Paceibacteria and Gracilibacteria MAGs. MAGs that were classified as candidate phylum OD1 from The Cedars also encoded V(A)-type ATP synthase or lacked any ATP synthase at all (Suzuki et al., 2017).

### **WOR-3 MAG**

One of the highest-coverage MAGs in Marker 3 and Sombrero fluids was classified as candidate phylum WOR-3, which was previously identified in methane-rich marine sediments (Baker et al., 2015) and was proposed to be renamed as Candidatus Stahlbacteria (Dombrowski et al., 2017). The WOR-3 MAG (59% completion with 0% redundancy) has a partial glycolysis pathway, perhaps explainable by the incompleteness of the MAG, and no TCA cycle. Other genes possibly indicative of organic carbon catabolism include those predicted to encode one subunit of pyruvate dehydrogenase (PdhD), pyruvate formate-lyase (PflD), formate dehydrogenase (FdoG), and a cation/acetate symporter (ActP). Like the NPL-UPA2 MAGs, one Paceibacteria MAG, and the archaeal MAGs, the WOR-3 MAG encodes a V(A)-type ATP synthase.

The WOR-3 formate dehydrogenase shares a maximum of only ~31% amino acid identities with any other sequences in the NCBI NR and JGI IMG “all isolates” databases. This divergent FDH-like sequence in the WOR-3 MAG is distinct from the divergent FDH-like sequence described above for other Lost City MAGs (**Figure 6**), and in both cases, additional research is required to establish whether their functions are indeed associated with formate metabolism.

## References

- Abe K, Ruan Z-S, Maloney PC. 1996. Cloning, Sequencing, and Expression in *Escherichia coli* of OxIT, the Oxalate:Formate Exchange Protein of *Oxalobacter formigenes*(\*). *Journal of Biological Chemistry* 271:6789–6793. DOI: 10.1074/jbc.271.12.6789.
- Aquino KA, Früh-Green GL, Rickli J, Bernasconi SM, Lang SQ, Lilley MD, Butterfield DA. In Revision. Multi-stage evolution of the Lost City hydrothermal vent fluids. *Geochimica et Cosmochimica Acta*.
- Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, Dehal P, Ware D, Perez F, Canon S, Sneddon MW, Henderson ML, Riehl WJ, Murphy-Olson D, Chan SY, Kamimura RT, Kumari S, Drake MM, Brettin TS, Glass EM, Chivian D, Gunter D, Weston DJ, Allen BH, Baumohl J, Best AA, Bowen B, Brenner SE, Bun CC, Chandonia J-M, Chia J-M, Colasanti R, Conrad N, Davis JJ, Davison BH, DeJongh M, Devoid S, Dietrich E, Dubchak I, Edirisinghe JN, Fang G, Faria JP, Frybarger PM, Gerlach W, Gerstein M, Greiner A, Gurtowski J, Haun HL, He F, Jain R, Joachimiak MP, Keegan KP, Kondo S, Kumar V, Land ML, Meyer F, Mills M, Novichkov PS, Oh T, Olsen GJ, Olson R, Parrello B, Pasternak S, Pearson E, Poon SS, Price GA, Ramakrishnan S, Ranjan P, Ronald PC, Schatz MC, Seaver SMD, Shukla M, Sutormin RA, Syed MH, Thomason J, Tintle NL, Wang D, Xia F, Yoo H, Yoo S, Yu D. 2018. KBase: The United States Department of Energy Systems Biology Knowledgebase. *Nature Biotechnology* 36:566–569. DOI: 10.1038/nbt.4163.

776 Baker BJ, Lazar CS, Teske AP, Dick GJ. 2015. Genomic resolution of linkages in carbon,  
777 nitrogen, and sulfur cycling among widespread estuary sediment bacteria. *Microbiome*  
778 3:14. DOI: 10.1186/s40168-015-0077-6.

779 Beulig F, Røy H, McGlynn SE, Jørgensen BB. 2019. Cryptic CH<sub>4</sub> cycling in the sulfate–methane  
780 transition of marine sediments apparently mediated by ANME-1 archaea. *The ISME*  
781 *Journal* 13:250–262. DOI: 10.1038/s41396-018-0273-z.

782 Biegel E, Schmidt S, González JM, Müller V. 2011. Biochemistry, evolution and physiological  
783 function of the Rnf complex, a novel ion-motive electron transport complex in  
784 prokaryotes. *Cellular and Molecular Life Sciences* 68:613–634. DOI: 10.1007/s00018-  
785 010-0555-8.

786 Brazelton WJ, Baross JA. 2009. Abundant transposases encoded by the metagenome of a  
787 hydrothermal chimney biofilm. *The ISME Journal* 3:1420–1424. DOI:  
788 10.1038/ismej.2009.79.

789 Brazelton WJ, Baross JA. 2010. Metagenomic Comparison of Two *Thiomicrospira* Lineages  
790 Inhabiting Contrasting Deep-Sea Hydrothermal Environments. *PLoS ONE* 5:e13530.  
791 DOI: 10.1371/journal.pone.0013530.

792 Brazelton WJ, Ludwig KA, Sogin ML, Andreishcheva EN, Kelley DS, Shen C-C, Edwards RL,  
793 Baross JA. 2010. Archaea and bacteria with surprising microdiversity show shifts in  
794 dominance over 1,000-year time scales in hydrothermal chimneys. *Proceedings of the*  
795 *National Academy of Sciences* 107:1612–1617. DOI: 10.1073/pnas.0905369107.

796 Brazelton WJ, Mehta MP, Kelley DS, Baross JA. 2011. Physiological Differentiation within a  
797 Single-Species Biofilm Fueled by Serpentinization. *mBio* 2:e00127-11. DOI:  
798 10.1128/mBio.00127-11.

799 Brazelton WJ, Schrenk MO, Kelley DS, Baross JA. 2006. Methane- and Sulfur-Metabolizing  
800 Microbial Communities Dominate the Lost City Hydrothermal Field Ecosystem. *Applied*  
801 *and Environmental Microbiology* 72:6257–6270. DOI: 10.1128/AEM.00574-06.

802 Brazelton WJ, Thornton CN, Hyer A, Twing KI, Longino AA, Lang SQ, Lilley MD, Früh-Green  
803 GL, Schrenk MO. 2017. Metagenomic identification of active methanogens and  
804 methanotrophs in serpentinite springs of the Voltri Massif, Italy. *PeerJ* 5:e2945. DOI:  
805 10.7717/peerj.2945.

806 Buchfink B, Reuter K, Drost HG. 2021. Sensitive protein alignments at tree-of-life scale using  
807 DIAMOND. *Nature Methods* 18:366–368. DOI: 10.1038/s41592-021-01101-x.

808 Bushnell B, Rood J, Singer E. 2017. BBMerge – Accurate paired shotgun read merging via  
809 overlap. *PLoS ONE* 12:1–15. DOI: 10.1371/journal.pone.0185056.

810 Butterfield DA, Massoth GJ. 1994. Geochemistry of north Cleft segment vent fluids: temporal  
811 changes in chlorinity and their possible relation to recent volcanism. *Journal of*  
812 *Geophysical Research* 99:4951–4968. DOI: 10.1029/93JB02798.

813 Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. 2016. DADA2:  
814 High-resolution sample inference from Illumina amplicon data. *Nature Methods* 13:581–  
815 583. DOI: 10.1038/nmeth.3869.

816 Chadwick GL, Skennerton CT, Laso-Pérez R, Leu AO, Speth DR, Yu H, Morgan-Lang C,  
817 Hatzenpichler R, Goudeau D, Malmstrom R, Brazelton WJ, Woyke T, Hallam SJ, Tyson  
818 GW, Wegener G, Boetius A, Orphan VJ. 2021. *Comparative genomics reveals electron*  
819 *transfer and syntrophic mechanisms differentiating methanotrophic and methanogenic*  
820 *archaea*. DOI: 10.1101/2021.09.25.461819.

821 Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. 2020. GTDB-Tk: a toolkit to classify  
822 genomes with the Genome Taxonomy Database. *Bioinformatics* 36:1925–1927. DOI:  
823 10.1093/bioinformatics/btz848.

824 Davis NM, Proctor DM, Holmes SP, Relman DA, Callahan BJ. 2018. Simple statistical  
825 identification and removal of contaminant sequences in marker-gene and metagenomics  
826 data. *Microbiome* 6:226. DOI: 10.1186/s40168-018-0605-2.



827 Dombrowski N, Seitz KW, Teske AP, Baker BJ. 2017. Genomic insights into potential  
828 interdependencies in microbial hydrocarbon and nutrient cycling in hydrothermal  
829 sediments. *Microbiome* 5:106. DOI: 10.1186/s40168-017-0322-2.

830 Dombrowski N, Teske AP, Baker BJ. 2018. Expansive microbial metabolic versatility and  
831 biodiversity in dynamic Guaymas Basin hydrothermal sediments. *Nature*  
832 *Communications* 9:4999. DOI: 10.1038/s41467-018-07418-0.

833 Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, Delmont TO. 2015. Anvi'o:  
834 an advanced analysis and visualization platform for 'omics data. *PeerJ* 3:e1319. DOI:  
835 10.7717/peerj.1319.

836 Eren AM, Kiefl E, Shaiber A, Veseli I, Miller SE, Schechter MS, Fink I, Pan JN, Yousef M,  
837 Fogarty EC, Trigodet F, Watson AR, Esen ÖC, Moore RM, Clayssen Q, Lee MD,  
838 Kivenson V, Graham ED, Merrill BD, Karkman A, Blankenberg D, Eppley JM, Sjödin A,  
839 Scott JJ, Vázquez-Campos X, McKay LJ, McDaniel EA, Stevens SLR, Anderson RE,  
840 Fuessel J, Fernandez-Guerra A, Maignien L, Delmont TO, Willis AD. 2021. Community-  
841 led, integrated, reproducible multi-omics with anvi'o. *Nature Microbiology* 6:3–6. DOI:  
842 10.1038/s41564-020-00834-3.

843 Fones EM, Colman DR, Kraus EA, Stepanauskas R, Templeton AS, Spear JR, Boyd ES. 2021.  
844 Diversification of methanogens into hyperalkaline serpentinizing environments through  
845 adaptations to minimize oxidant limitation. *The ISME Journal* 15:1121–1135. DOI:  
846 10.1038/s41396-020-00838-1.

847 Garber AI, Nealson KH, Okamoto A, McAllister SM, Chan CS, Barco RA, Merino N. 2020.  
848 FeGenie: A Comprehensive Tool for the Identification of Iron Genes and Iron Gene  
849 Neighborhoods in Genome and Metagenome Assemblies. *Frontiers in Microbiology*  
850 11:37. DOI: 10.3389/fmicb.2020.00037.

851 Graham ED, Heidelberg JF, Tully BJ. 2017. BinSanity: unsupervised clustering of environmental  
852 microbial assemblies using coverage and affinity propagation. *PeerJ* 5:e3035. DOI:  
853 10.7717/peerj.3035.

854 Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic  
855 gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119.  
856 DOI: 10.1186/1471-2105-11-119.

857 Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A,  
858 Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong S-Y, Lopez  
859 R, Hunter S. 2014. InterProScan 5: genome-scale protein function classification.  
860 *Bioinformatics* 30:1236–1240. DOI: 10.1093/bioinformatics/btu031.

861 Kanehisa M, Sato Y, Morishima K. 2016. BlastKOALA and GhostKOALA: KEGG Tools for  
862 Functional Characterization of Genome and Metagenome Sequences. *Journal of*  
863 *Molecular Biology* 428:726–731. DOI: 10.1016/j.jmb.2015.11.006.

864 Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, Wang Z. 2019. MetaBAT 2: an adaptive  
865 binning algorithm for robust and efficient genome reconstruction from metagenome  
866 assemblies. *PeerJ* 7:e7359. DOI: 10.7717/peerj.7359.

867 Kaster A-K, Goenrich M, Seedorf H, Liesegang H, Wollherr A, Gottschalk G, Thauer RK. 2011.  
868 More Than 200 Genes Required for Methane Formation from H<sub>2</sub> and CO<sub>2</sub> and Energy  
869 Conservation Are Present in *Methanothermobacter marburgensis* and  
870 *Methanothermobacter thermautotrophicus*. *Archaea* 2011:e973848. DOI:  
871 10.1155/2011/973848.

872 Kelley DS, Karson JA, Früh-Green GL, Yoerger DR, Shank TM, Butterfield DA, Hayes JM,  
873 Schrenk MO, Olson EJ, Proskurowski G, Jakuba M, Bradley A, Larson B, Ludwig K,  
874 Glickson D, Buckman K, Bradley AS, Brazelton WJ, Roe K, Elend MJ, Delacour A,  
875 Bernasconi SM, Lilley MD, Baross JA, Summons RE, Sylva SP. 2005. A Serpentinite-

876 Hosted Ecosystem: The Lost City Hydrothermal Field. *Science* 307:1428–1434. DOI:  
877 10.1126/science.1102556.

878 Koch H, Germscheid N, Freese HM, Noriega-Ortega B, Lücking D, Berger M, Qiu G, Marzinelli  
879 EM, Campbell AH, Steinberg PD, Overmann J, Dittmar T, Simon M, Wietz M. 2020.  
880 Genomic, metabolic and phenotypic variability shapes ecological differentiation and  
881 intraspecies interactions of *Alteromonas macleodii*. *Scientific Reports* 10:809. DOI:  
882 10.1038/s41598-020-57526-5.

883 Kopylova E, Noé L, Touzet H. 2012. SortMeRNA: fast and accurate filtering of ribosomal RNAs  
884 in metatranscriptomic data. *Bioinformatics* 28:3211–3217. DOI:  
885 10.1093/bioinformatics/bts611.

886 Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. 2013. Development of a Dual-  
887 Index Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence  
888 Data on the MiSeq Illumina Sequencing Platform. *Applied and Environmental*  
889 *Microbiology* 79:5112–5120. DOI: 10.1128/AEM.01043-13.

890 Kpebe A, Benvenuti M, Guendon C, Rebai A, Fernandez V, Le Laz S, Etienne E, Guigliarelli B,  
891 García-Molina G, de Lacey AL, Baffert C, Brugna M. 2018. A new mechanistic model for  
892 an O<sub>2</sub>-protected electron-bifurcating hydrogenase, Hnd from *Desulfovibrio*  
893 *fructosovorans*. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 1859:1302–1312.  
894 DOI: 10.1016/j.bbabi.2018.09.364.

895 Landry Z, Swan BK, Herndl GJ, Stepanauskas R, Giovannoni SJ. SAR202 Genomes from the  
896 Dark Ocean Predict Pathways for the Oxidation of Recalcitrant Dissolved Organic  
897 Matter. *mBio* 8:e00413-17. DOI: 10.1128/mBio.00413-17.

898 Lang SQ, Benitez-Nelson B. 2021. Hydrothermal Organic Geochemistry (HOG) sampler for  
899 deployment on deep-sea submersibles. *Deep Sea Research Part I: Oceanographic*  
900 *Research Papers* 173:103529. DOI: 10.1016/j.dsr.2021.103529.

901 Lang SQ, Früh-Green GL, Bernasconi SM, Brazelton WJ, Schrenk MO, McGonigle JM. 2018.  
 902 Deeply-sourced formate fuels sulfate reducers but not methanogens at Lost City  
 903 hydrothermal field. *Scientific Reports* 8:755. DOI: 10.1038/s41598-017-19002-5.

904 Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods*  
 905 9:357–359. DOI: 10.1038/nmeth.1923.

906 Li D, Luo R, Liu CM, Leung CM, Ting HF, Sadakane K, Yamashita H, Lam TW. 2016. MEGAHIT  
 907 v1.0: A fast and scalable metagenome assembler driven by advanced methodologies  
 908 and community practices. *Methods* 102:3–11. DOI: 10.1016/j.ymeth.2016.02.020.

909 Liu R, Wei X, Wang L, Cao J, Song W, Wu J, Thomas T, Jin T, Wang Z, Wei W, Wei Y, Zhai H,  
 910 Yao C, Shen Z, Fang J. 2021. Novel Chloroflexi Genomes From The Deepest Ocean  
 911 Reveal Metabolic Strategies For The Adaptation To Deep-Sea Habitats. DOI:  
 912 10.21203/rs.3.rs-254541/v2.

913 Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for  
 914 RNA-seq data with DESeq2. *Genome Biology* 15:550. DOI: 10.1186/s13059-014-0550-  
 915 8.

916 Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey ARN, Potter  
 917 SC, Finn RD, Lopez R. 2019. The EMBL-EBI search and sequence analysis tools APIs  
 918 in 2019. *Nucleic acids research* 47:W636–W641. DOI: 10.1093/nar/gkz268.

919 Mand TD, Metcalf WW. 2019. Energy Conservation and Hydrogenase Function in  
 920 Methanogenic Archaea, in Particular the Genus Methanosarcina. *Microbiology and*  
 921 *Molecular Biology Reviews* 83. DOI: 10.1128/MMBR.00020-19.

922 Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads.  
 923 *EMBnet.journal* 17:10–12. DOI: 10.14806/ej.17.1.200.

924 McGonigle JM, Lang SQ, Brazelton WJ. 2020. Genomic Evidence for Formate Metabolism by  
 925 *Chloroflexi* as the Key to Unlocking Deep Carbon in Lost City Microbial Ecosystems.  
 926 *Applied and Environmental Microbiology* 86:e02583-19. DOI: 10.1128/AEM.02583-19.

927 McMurdie PJ, Holmes S. 2013. Phyloseq: An R Package for Reproducible Interactive Analysis  
 928 and Graphics of Microbiome Census Data. *PLoS ONE* 8. DOI:  
 929 10.1371/journal.pone.0061217.

930 Mehrshad M, Rodriguez-Valera F, Amoozegar MA, López-García P, Ghai R. 2018. The  
 931 enigmatic SAR202 cluster up close: shedding light on a globally distributed dark ocean  
 932 lineage involved in sulfur cycling. *The ISME journal* 12:655–668. DOI: 10.1038/s41396-  
 933 017-0009-5.

934 Mei N, Postec A, Erauso G, Joseph M, Pelletier B, Payri C, Ollivier B, Quéméneur M. 2016.  
 935 *Serpentinicella alkaliphila* gen. nov., sp. nov., a novel alkaliphilic anaerobic bacterium  
 936 isolated from the serpentinite-hosted Prony hydrothermal field, New Caledonia.  
 937 *International Journal of Systematic and Evolutionary Microbiology* 66:4464–4470. DOI:  
 938 10.1099/ijsem.0.001375.

939 Menzel P, Ng KL, Krogh A. 2016. Fast and sensitive taxonomic classification for metagenomics  
 940 with Kaiju. *Nature Communications* 7:11257. DOI: 10.1038/ncomms11257.

941 Mirdita M, Steinegger M, Breitwieser F, Söding J, Levy Karin E. 2021. Fast and sensitive  
 942 taxonomic assignment to metagenomic contigs. *Bioinformatics* 37:3029–3031. DOI:  
 943 10.1093/bioinformatics/btab184.

944 Motamedi S, Orcutt BN, Früh-Green GL, Twing KI, Pendleton HL, Brazelton WJ. 2020. Microbial  
 945 Residents of the Atlantis Massif's Shallow Serpentinite Subsurface. *Applied and*  
 946 *Environmental Microbiology* 86:e00356-20. DOI: 10.1128/AEM.00356-20.

947 Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. 2017. metaSPAdes: a new versatile  
 948 metagenomic assembler. *Genome Research* 27:824–834. DOI: 10.1101/gr.213959.116.

949 Parks DH, Chuvochina M, Rinke C, Mussig AJ, Chaumeil P-A, Hugenholtz P. 2021. GTDB: an  
 950 ongoing census of bacterial and archaeal diversity through a phylogenetically consistent,  
 951 rank normalized and complete genome-based taxonomy. *Nucleic Acids Research*. DOI:  
 952 10.1093/nar/gkab776.

953 Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the  
954 quality of microbial genomes recovered from isolates, single cells, and metagenomes.  
955 *Genome Research* 25:1043–1055. DOI: 10.1101/gr.186072.114.

956 *Picard Toolkit*. 2019. Broad Institute.

957 Porter KG, Feig YS. 1980. The use of DAPI for identifying and counting aquatic microflora1.  
958 *Limnology and Oceanography* 25:943–948. DOI: 10.4319/lo.1980.25.5.0943.

959 Postec A, Quéméneur M, Lecoivre A, Chabert N, Joseph M, Erauso G. 2021. Alkaliphilus  
960 serpentinus sp. nov. and Alkaliphilus pronyensis sp. nov., two novel anaerobic  
961 alkaliphilic species isolated from the serpentinite-hosted Prony Bay Hydrothermal Field  
962 (New Caledonia). *Systematic and Applied Microbiology* 44:126175. DOI:  
963 10.1016/j.syapm.2020.126175.

964 Proskurowski G, Lilley MD, Seewald JS, Fru h-Green GL, Olson EJ, Lupton JE, Sylva SP,  
965 Kelley DS. 2008. Abiogenic Hydrocarbon Production at Lost City Hydrothermal Field.  
966 *Science* 319:604–607. DOI: 10.1126/science.1151194.

967 Pruesse E, Peplies J, Glöckner FO. 2012. SINA: Accurate high-throughput multiple sequence  
968 alignment of ribosomal RNA genes. *Bioinformatics* 28:1823–1829. DOI:  
969 10.1093/bioinformatics/bts252.

970 Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2013.  
971 The SILVA ribosomal RNA gene database project: Improved data processing and web-  
972 based tools. *Nucleic Acids Research* 41:590–596. DOI: 10.1093/nar/gks1219.

973 Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic  
974 features. *Bioinformatics* 26:841–842. DOI: 10.1093/bioinformatics/btq033.

975 Rempfert KR, Miller HM, Bompard N, Nothaft D, Matter JM, Kelemen P, Fierer N, Templeton  
976 AS. 2017. Geological and Geochemical Controls on Subsurface Microbial Life in the  
977 Samail Ophiolite, Oman. *Frontiers in Microbiology* 8:56. DOI: 10.3389/fmicb.2017.00056.

978 Rohland N, Reich D. 2012. Cost-effective, high-throughput DNA sequencing libraries for  
 979 multiplexed target capture. *Genome Research* 22:939–946. DOI:  
 980 10.1101/gr.128124.111.

981 Schrenk MO, Kelley DS, Bolton SA, Baross JA. 2004. Low archaeal diversity linked to  
 982 subseafloor geochemical processes at the Lost City Hydrothermal Field, Mid-Atlantic  
 983 Ridge. *Environmental Microbiology* 6:1086–1095. DOI: 10.1111/j.1462-  
 984 2920.2004.00650.x.

985 Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics (Oxford,*  
 986 *England)* 30:2068–2069. DOI: 10.1093/bioinformatics/btu153.

987 Sieber CMK, Probst AJ, Sharrar A, Thomas BC, Hess M, Tringe SG, Banfield JF. 2018.  
 988 Recovery of genomes from metagenomes via a dereplication, aggregation and scoring  
 989 strategy. *Nature Microbiology* 3:836–843. DOI: 10.1038/s41564-018-0171-1.

990 Sorokin DYu, Tourova TP, Mußmann M, Muyzer G. 2008. *Dethiobacter alkaliphilus* gen. nov.  
 991 sp. nov., and *Desulfurivibrio alkaliphilus* gen. nov. sp. nov.: two novel representatives of  
 992 reductive sulfur cycle from soda lakes. *Extremophiles* 12:431–439. DOI:  
 993 10.1007/s00792-008-0148-8.

994 Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of  
 995 large phylogenies. *Bioinformatics* 30:1312–1313. DOI: 10.1093/bioinformatics/btu033.

996 Suzuki S, Ishii S, Hoshino T, Rietze A, Tenney A, Morrill PL, Inagaki F, Kuenen JG, Nealson  
 997 KH. 2017. Unusual metabolic diversity of hyperalkaliphilic microbial communities  
 998 associated with subterranean serpentinization at The Cedars. *The ISME Journal*  
 999 11:2584–2598. DOI: 10.1038/ismej.2017.111.

1000 Suzuki S, Nealson KH, Ishii S. 2018. Genomic and in-situ Transcriptomic Characterization of the  
 1001 Candidate Phylum NPL-UPL2 From Highly Alkaline Highly Reducing Serpentinized  
 1002 Groundwater. *Frontiers in Microbiology* 9:3141. DOI: 10.3389/fmicb.2018.03141.

1003 Thauer RK, Kaster A-K, Goenrich M, Schick M, Hiromoto T, Shima S. 2010. Hydrogenases from  
 1004 Methanogenic Archaea, Nickel, a Novel Cofactor, and H<sub>2</sub> Storage. *Annual Review of*  
 1005 *Biochemistry* 79:507–536. DOI: 10.1146/annurev.biochem.030508.152103.  
 1006 Thornton CN, Tanner WD, VanDerslice JA, Brazelton WJ. 2020. Localized effect of treated  
 1007 wastewater effluent on the resistome of an urban watershed. *GigaScience* 9. DOI:  
 1008 10.1093/gigascience/giaa125.  
 1009 Tully BJ, Wheat CG, Glazer BT, Huber JA. 2018. A dynamic microbial community with high  
 1010 functional redundancy inhabits the cold, oxic subseafloor aquifer. *The ISME journal*  
 1011 12:1–16. DOI: 10.1038/ismej.2017.187.  
 1012 Wu Y-W, Simmons BA, Singer SW. 2016. MaxBin 2.0: an automated binning algorithm to  
 1013 recover genomes from multiple metagenomic datasets. *Bioinformatics* 32:605–607. DOI:  
 1014 10.1093/bioinformatics/btv638.  
 1015 Yilmaz P, Parfrey LW, Yarza P, Gerken J, Priesse E, Quast C, Schweer T, Peplies J, Ludwig  
 1016 W, Glöckner FO. 2014. The SILVA and “All-species Living Tree Project (LTP)”  
 1017 taxonomic frameworks. *Nucleic Acids Research* 42:D643–D648. DOI:  
 1018 10.1093/nar/gkt1209.  
 1019 Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, Busk PK, Xu Y, Yin Y. 2018. dbCAN2: a  
 1020 meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids*  
 1021 *Research* 46:W95–W101. DOI: 10.1093/nar/gky418.  
 1022 Zhou Z, Liu Y, Pan J, Cron BR, Toner BM, Anantharaman K, Breier JA, Dick GJ, Li M. 2020.  
 1023 Gammaproteobacteria mediating utilization of methyl-, sulfur- and petroleum organic  
 1024 compounds in deep ocean hydrothermal plumes. *The ISME journal* 14:3136–3148. DOI:  
 1025 10.1038/s41396-020-00745-5.  
 1026