Handling Temperature Bursts Reaching 464°C: Different Microbial Strategies in the Sisters Peak Hydrothermal Chimney

Mirjam Perner,[a] Giorgio Gonnella,[b] Stefan Kurtz,[b] Julie LaRoche[c]*

Molecular Biology of Microbial Consortia, University of Hamburg, Biocenter Klein Flottbek, Hamburg, Germany[a]; Center for Bioinformatics, University of Hamburg, Hamburg, Germany[b]; Marine Biogeochemistry, Leibniz Institute of Marine Science, GEOMAR, Kiel, Germany[c]

The active venting Sisters Peak (SP) chimney on the Mid-Atlantic Ridge holds the current temperature record for the hottest ever measured hydrothermal fluids (400°C, accompanied by sudden temperature bursts reaching 464°C). Given the unprecedented temperature regime, we investigated the biome of this chimney with a focus on special microbial adaptations for thermal tolerance. The SP metagenome reveals considerable differences in the taxonomic composition from those of other hydrothermal vent and subsurface samples; these could be better explained by temperature than by other available abiotic parameters. The most common species to which SP genes were assigned were thermophilic *Aciduliprofundum* sp. strain MAR08-339 (11.8%), *Hippea maritima* (3.8%), *Caldisericum exile* (1.5%), and *Caminibacter mediatlanticus* (1.4%) as well as to the mesophilic *Niastella koreensis* (2.8%). A statistical analysis of associations between taxonomic and functional gene assignments revealed specific overrepresented functional categories: for *Aciduliprofundum*, protein biosynthesis, nucleotide metabolism, and energy metabolism genes; for *Hippea* and *Caminibacter*, cell motility and/or DNA replication and repair system genes; and for *Niastella*, cell wall and membrane biogenesis genes. Cultured representatives of these organisms inhabit different thermal niches; i.e., *Aciduliprofundum* has an optimal growth temperature of 70°C, *Hippea* and *Caminibacter* have optimal growth temperatures around 55°C, and *Niastella* grows between 10 and 37°C. Therefore, we posit that the different enrichment profiles of functional categories reflect distinct microbial strategies to deal with the different impacts of the local sudden temperature bursts in disparate regions of the chimney.

Hydrothermal vents represent hot spots of life in the otherwise poorly inhabited deep sea. This life is based on energy-rich inorganic compounds, e.g., hydrogen or sulfide, which are transported with the reduced hydrothermal fluids from inner earth to more habitable grounds (1). These reduced inorganic substrates can be oxidized by a phylogenetically diverse group of microorganisms, and the yielded energy can fuel autotrophic CO_2 fixation (1). Although nourishment provided by vent fluids supports the settlement and successful inhabitance of microbial communities, the hydrothermal vent environment can also be associated with extreme conditions—such as high temperature or toxicity of specific compounds—which may well act as a deterrent to microbial colonization. High temperatures provide challenges for microbes, and multiple strategies have been developed to deal with elevated temperature (2). Different microbial strategies for dealing with thermal damage exist and include efficient repair systems, reverse gyrase, multiple copies of chromosomes, increased G+C contents in RNAs, abundant posttranscriptional modification of RNAs, and increased thermostability of proteins (2). Besides temperature, different concentrations of specific chemical compounds in vent fluid can also profoundly impact the microbial community (3–7). Clues to specific adaptations for coping with the extreme conditions have been found in metagenomes from hydrothermal samples (8–10). They include a high degree of mismatch repair and homologous recombination systems as well as abundant genes for chemotaxis and flagellar assembly (8). Membrane transport and multidrug-resistant efflux pumps were judged important for the 0.1- to 0.2-μm microbial size fraction from hydrothermal fluids in the Mariana Trough (9). Other strategies for successfully colonizing hydrothermally influenced biotopes include adaptability, versatile metabolisms, and virulence traits (11).

The Sisters Peak (SP) venting site at 5°S along the Mid-Atlantic Ridge (MAR) is unique in that it emits hot, phase-separated hydrothermal fluids, and some of the currently hottest known hydrothermal fluids have been measured here (maximum temperature, 464°C) (12). To gain insights into the variety of microbial strategies that different groups of microorganisms employ to deal with the extreme conditions at SP, we performed a differential analysis of the functional annotation of genes assigned to the most common taxa in the metagenome. We found that distinct functional profiles are associated with different taxonomic groups. We linked these observations to temperature preferences known for cultured members of these taxa, leading us to the hypothesis that the observed differences reflected the spatial thermal heterogeneity. Thus, we postulate that the different functional profiles observed in this chimney cross-sectional sample reflect different microbial strategies by which microorganisms in SP handle the range and variability of temperature regimes at the site, from hotter inner chimney regions to cooler outer chimney provinces.

MATERIALS AND METHODS

Sampling site and DNA extraction. A flank of the SP vent massive sulfide chimney, located at 4°48′S along the MAR, was sampled by a remote operated vehicle (ROV 6000; GEOMAR, Kiel, Germany) during the MAR-SUED V cruise (March-April 2009) with the RV *Meteor*. After recovery, the chimney sample (274 ROV 1B) was stored immediately at −80°C until further processing. For DNA extraction, 246 g of the entire sulfide chimney cross section was ground, and DNA was extracted using a phenol-chloroform extraction method (13). DNA extraction yielded between 20 and 120 ng μl^{-1} of DNA in a small volume and a low ratio of absorbance at 260 nm to that at 280 nm.

Amplification and pyrosequencing. To obtain sufficient material for pyrosequencing of the appendant metagenome, we performed multiple displacement amplification (MDA) with φ29 DNA polymerase (REPLI-g kit; Qiagen, Germany) according to the manufacturer's instructions. MDA was employed in parallel samples each with 2.5 μl (50 to 100 ng μl^{-1}) of bulk DNA vent chimney starting material, which resulted in 2.5 μg μl^{-1} of pooled, purified DNA (260/280 ratio, 1.8). Random shotgun pyrosequencing using titanium chemistry of the material subjected to MDA was performed on a 454 Life Sciences GS FLX system platform and yielded 340 Mb of DNA sequences, consisting of 875,069 reads with an average sequence length of 389 nucleotides (see Table S1 in the supplemental material). Pyrosequenced reads from the metagenomes of an active venting chimney at Juan de Fuca (JdF) (8), from hydrothermal fluids emitted in the Mariana Trough (MT) (9), and from deep subsurface sediments in the Brazos-Trinity Basin (BTB) (14) were downloaded from the NCBI SRA database (JdF, SRR029255; MT, SRR016610; and BTB, SRR023396). A detailed summary of preprocessing results for the SP, JdF, MT, and BTB samples is given in Table S1.

Preprocessing of raw reads. The raw reads of all data sets, i.e., from SP, JdF, MT, and BTB, were preprocessed to reduce low-quality regions, contaminations, and artifacts (see Table S1). An overview of the upstream analysis pipeline is available in Fig. S1A in the supplemental material. First, emulsion PCR (emPCR) artifacts in form of duplicates were removed using cd-hit-454 version 4.5.4 (15). In the only fosmid-based data set (JdF), contaminations were eliminated using SeqClean (http://sourceforge.net/projects/seqclean/) with the options -l 50 and -y 8, using as references the sequences of pCC1FOS (GenBank accession number EU140751.1) and the *Escherichia coli* DH10B chromosome (NCBI RefSeq number NC_010473.1). After these preliminary steps, the quality of each read set was determined using FastQC version 0.10.1 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). This analysis detected regions of skewed composition and overrepresented k-mers. Thus, we used Prinseq version 0.20.2 (16) to eliminate a 5-bp prefix from each of the JdF and BT reads, to cut all reads to a read set-specific maximum length (SP, 600 bp; MT, 260 bp; JdF; 280 bp; and BT, 290 bp), and to filter out all reads with a length of less than 50 bp or of low complexity (entropy > 70). Furthermore, from the resulting reads, the ends of low-quality were trimmed by applying the seqtk trimfq tool (https://github.com/lh3/seqtk) with an error threshold of 0.0158 (corresponding to a Phred score of 18).

Clustering and assembly of preprocessed reads. The preprocessed reads of SP, JdF, MT, and BTB were clustered by cd-hit-est version 4.5.4 (17), using parameters adapted to 454 reads (identity threshold, 98%; mismatch score, −1; and gap opening score, −3). This step was performed for the following reasons: first, it reduces the redundancy (see "Handling of possible methodological biases" below); second, it allows successful assembly of the MT sample, which could not be accomplished without this step (9). The sets of cluster representatives were assembled using Newbler version 2.8 (18), with a minimal match length of 50 bp and a minimal match identity of 98%. The final assemblies used in all subsequent analysis consisted of the set of all contigs plus all unassembled reads marked by Newbler as singleton or outlier.

Gene prediction and taxonomic and functional assignment. The identification of protein-coding genes in the assemblies was predicted using FragGeneScan version 1.16 (19), with the options -complete set to 0 and -train set to 454_10. All subsequent analyses were performed at the gene level; an overview of the downstream analysis pipeline is available in Fig. S1B in the supplemental material.

The assignment of protein-coding genes to taxonomic groups was performed using MEGAN version 5.0 (20), based on the alignments to the NCBI NR database (21), performed by blastp version 2.2.27+ (22). The lowest common ancestor (LCA) parameters of MEGAN were set as follows: minimal support, 3; minimal score, 50; maximal E value, 10^{-3}; and top percentage, 10. Hierarchical plots of the taxonomic assignments were prepared using Krona version 2.4 (23).

Using the functional annotation feature of MEGAN (24), we assigned protein-coding genes to categories in SEED (25), KEGG Orthology (26), and COG (27). For the SEED analysis, the built-in mappings of MEGAN were used. For KEGG Orthology and COG, we integrated the built-in mappings with data from the Uniprot database (28). An updated KEGG functional hierarchy tree was constructed using data from the KEGG Orthology database (26). Textual descriptions for each of the COG/KOG/NOG identifiers were obtained from the eggNOG database (27).

rRNA genes were identified in the assembly using version 1.0 of the Meta-RNA(H3) tool (29). Alignment of the predicted 16S and 23S genes to the SILVA database (30) was performed using the SINA web server (31). Thereby, taxonomical classification of the genes was obtained by the lowest common ancestor algorithm, using "search and classify" with the following parameters: minimal identity, 0.9, and neighbors per query sequence, 10.

The number of species of protein-coding genes in the assemblies was predicted using FragGeneScan version 1.16 (19), with the options -complete set to 0 and -train set to 454_10. All subsequent analyses were performed at the gene level; an overview of the downstream analysis pipeline is available in Fig. S1B in the supplemental material.

Binning of the SP assembly. The binning tool Metawatt, version 1.7 (32), was applied to the SP assembly, to obtain an estimate of the number of species, independent from the taxonomic assignments by MEGAN. We used the parameter set "high confidence" and limited the analysis to the first step available in the tool, which is based only on sequence composition statistics; the subsequent refinement step of Metawatt does not apply to our analysis, as it is based on coverage information.

Fragment recruitment to selected genomes. Sequences and annotations of the following reference genomes of the most frequent species according to the taxonomic assignments were obtained from the NCBI RefSeq database (21): *Aciduliprofundum boonei* T469 (NC_013926), *Aciduliprofundum* sp. strain MAR08-339 (NC_019942), *Caldisericum exile* AZM16c01 (NC_017096), *Hippea maritima* DSM 10411 (NC_015318), and *Niastella koreensis* GR20-10 (NC_016609). Sequences in the SP assembly were aligned to the reference genomes using blastp and tblastx, version 2.2.27+ (22). We considered only results with a maximum E value of 10^{-5} and a minimum identity of 75%. For tblastx, the genetic code parameters for query and subject were set to 11 (bacterial, archaeal, and plant plastid code). KEGG Orthology functional annotations and their mappings to COG identifications (IDs) were obtained from KEGG (26). Genomic islands in the reference genomes were computed using IslandViewer (33). Circular plots of the genome annotations and SP fragment recruitment were created using Circos version 0.64 (34).

Analysis of the variability of *Aciduliprofundum* sequences. To estimate the degree of sequence variability in *Aciduliprofundum* species in the SP metagenome, we further analyzed the results of fragment recruitment of SP contigs to the genomes of *Aciduliprofundum boonei* T469 and *Aciduliprofundum* sp. MAR08-339. For each alignment set (blastn and tblastx), a set R of ranges of genome coordinates was computed, where each range maximal includes only positions aligned to the same, possibly empty, set of query positions (i.e., all positions in the range either are uncovered or are covered by the same set of substrings of the contigs). The number of variants for each genomic range in R was computed as the size of the set of different contigs substrings aligned to the genomic range.
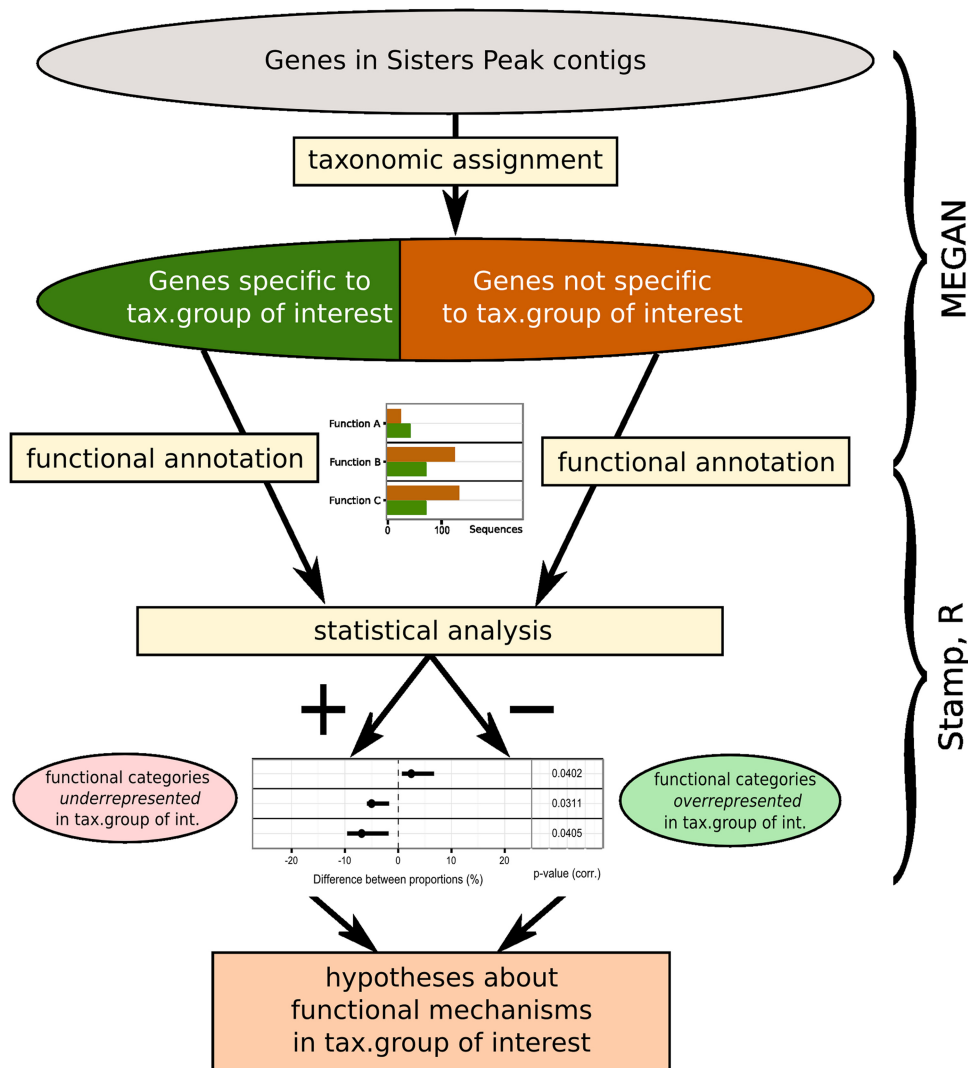
FIG 1 Overview of our method for the differential analysis of functions associated with specific taxa.

Analysis of predicted *Aciduliprofundum* sp. MAR08-339 genomic islands. Putative extensions of the genomic islands predicted by Island-Viewer in the *Aciduliprofundum* sp. MAR08-339 reference genome were defined by computing ranges of the genome without blastn hits in the SP metagenome and completely including the predicted islands. To investigate the function and possible origin of the islands, we aligned the products of annotated island genes to the NCBI NR database (21) using blastp version 2.2.27+ (22), with a maximum E value of 10^{-5}.

Analysis of alpha and beta diversity. Operational taxonomic unit (OTU) tables based on the genes taxonomic assignments at different taxonomic rank levels were exported from MEGAN and used for alpha and beta diversity analyses. Using Qiime version 1.6 (35), we computed the number of OTUs and the Shannon diversity index (using log base 2) at each taxonomic rank level and prepared rarefaction curves. Using the R library Vegan (36), we computed pairwise Bray-Curtis dissimilarity indices based on the assignments at the species level and performed nonmetric multidimensional scaling using the metaMDS function. Species abundances normalized to 1,000 total counts per site were computed, and species scores were plotted as bubbles representing the sum of normalized abundances at the four sites (bubble size, log scale) and which portion of the sum comes from the count at SP (bubble color).

Comparisons of metagenome functional profiles. Using STAMP version 2.0.0 (37), we prepared scatter plots for the comparison of KEGG, SEED, and COG functional profiles of SP to those of other metagenomes (JdF, MT, and BTB). Relative frequencies of the assignments of genes to functional categories were used, i.e., normalized by the total number of genes predicted in the metagenome (see Table S1 in the supplemental material).

Functional profiles of specific taxonomic groups. We investigated the presence of functions enriched or depleted in specific taxonomic groups (for a graphical overview of the methods, see Fig. 1). For each taxonomic group, using MEGAN, we partitioned the set of genes as follows: those assigned to any node in the taxonomic tree representing the taxonomic group of interest and the subtree below them and all other genes. Taxonomic groups analyzed included the five genera (*Aciduliprofundum*, *Hippea*, *Niastella*, *Caldisericum*, and *Caminibacter*) to which the most common species in SP belong and a further group denominated "uncultured organisms," which consisted of all taxa whose name, in the NCBI taxonomy tree provided with MEGAN, contains the words uncultured and/or environmental.

Using STAMP version 2.0.0 (37), for each taxonomy-based partition of the genes, we computed the proportions of the total number of genes in the two subsets assigned to each functional category in SEED, KEGG, and COG. The significance of differences between the proportions was evaluated using a two-sided Fisher exact text (38); thereby the false-discovery rate was controlled by applying the Benjamini-Hochberg procedure (39),

and confidence intervals of the difference of proportions were computed using the Newcombe-Wilson method ([40]). The results were plotted using the R library ggplot2 ([41]).

Handling of possible methodological biases. The SP, JdF, MT, and BTB data sets all consist of sequencing reads acquired with 454 pyrosequencing (for further details, see Table S1 in the supplemental material). Although other metagenomic data from hydrothermal habitats also exist, e.g., from *Alvinella pompejana* epibionts at the East Pacific Rise ([11]) and from a biofilm of the Lost City chimney ([42]), we decided against incorporating these data sets into our analyses because they were obtained by Sanger sequencing, so a comparison would be biased. While SP, MT, and BTB environmental samples were all processed by MDA, the JdF data were generated by cloning the DNA in fosmids. The different handling of the DNA might reduce the accuracy of the comparison of the data sets; this was in part compensated for by introducing an additional cleaning step in the preprocessing of the JdF data set. MDA is an indispensable tool enabling the sequencing of samples for which only small amounts of DNA can be extracted. However, it has also been shown that MDA can introduce significant bias in metagenomic analyses (for examples, see references [43], [44], and [45]). In particular, nonuniform amplification of different regions of the original material has been reported. The procedure applied to the JdF sample, based on the preparation of a fosmid library, also results in artifactual coverage differences.

Our analysis strategy takes these possible coverage biases into account: we estimated the abundance of a species by determining the number of different sequences assigned to a species, ignoring redundant sequences. This was achieved by clustering and assembling the sequencing reads. The main disadvantage of ignoring the redundancy information may lie in limited accuracy in the estimation of the abundance of very common taxa. In fact, due to sampling effects (low coverage sequencing of complex community), the number of different sequences observed for such taxa would also depend on other factors besides abundance, such as genome size and degree of intraspecific genome variability. Nevertheless, in metagenomic surveys with limited coverage, such as those compared for this study, none of the genomes are sequenced with high redundancy. Thus, we expect the number of different reads sampled from the genome of a species to remain approximately proportional to the species abundance. Therefore, we consider our method as the best available choice for the analysis of data derived from samples that require MDA.

A possible concern about assembled metagenomes is the presence of chimeric contigs. In order to reduce this risk, the assembly was computed using conservative parameters when determining overlaps between the reads. Furthermore, taxonomic and functional assignments were derived not from the contigs directly but rather from predicted protein fragments. Thus, artifacts from erroneously joined reads are expected to have only a limited local impact. As the numbers of predicted protein fragments varied considerably among the data sets (see Table S1 in the supplemental material), comparative analyses involving different data sets were performed using relative frequencies, normalized by the total counts for each data set.

Another source of potential bias may result from the presence of multiple nonoverlapping short fragments of the same gene, which would be counted separately. This may lead to overestimation of the prevalence of the function associated with the gene. However, this kind of error, arising from the low sampling rate in less abundant genomes, is not specific to our analysis method, and no computational method can address this bias, as the lack of a significant overlap does not allow recognition that multiple fragments belong to the same gene.

Nucleotide sequence accession number. The sequence data (raw reads) have been deposited in the National Center for Biotechnology Information (NCBI) SRA under accession number SRA047926.

RESULTS AND DISCUSSION

We investigated the microbiome associated with the active venting chimney at Sisters Peak (SP); a major point of interest for this site is represented by its strikingly high temperature, with a base level of 400°C and fluid bursts of even higher temperatures (measured up to 464°C). The results presented in this study are based on a shotgun metagenomic analysis of a cross-sectional sample of the SP chimney. In the following sections, we first discuss the results of the sequencing and preliminary data processing steps and then present the results of functional and taxonomic characterization of the metagenome. This includes a comparison with available metagenome data from other vent and subsurface samples, a detailed description of the taxonomic characterization of the SP microbial community, and a differential functional analysis, which relates the taxonomic and functional annotations of the genes, to better understand the underlying microbial strategies.

Sequencing and annotation of the SP metagenome. To obtain enough high-quality DNA material for sequencing, we subjected the DNA extracted from the sample to multiple displacement amplification (MDA). To make the data analysis more robust to MDA-related concerns, we developed a new analysis approach tailored to our data (see Materials and Methods). The results of the different processing steps leading to the list of predicted genes are reported in Table S1 in the supplemental material. Pyrosequencing of the SP metagenome yielded 875,069 sequencing reads, with a total of 340 Mb of sequences. The nonredundant set, consisting of contigs and nonredundant unassembled reads, included 23,933 sequences with a total length of 8 Mb, from which 24,055 complete and partial protein-coding genes were predicted.

To place the SP metagenomic information into context with other metagenomic data sets, we compared it to the metagenomes of two hydrothermal vent samples, namely, an active venting chimney from JdF ([8]) and hydrothermal fluids from the MT ([9]), as well as from deep marine subsurface sediments (8 meters below seafloor [mbsf]) of the BTB ([14]). To ensure comparability of the results, we reanalyzed these metagenome data sets (see Table S1), using the same data analysis protocol as employed for SP. The subsequent analyses were based on the alignment of the gene products to the NCBI NR database. Significant blastp hits (E value < 0.001) were found for 12,112 predicted SP genes (50.4%). This proportion was slightly higher than the proportions for the MT (43.9%) and BTB (44.8%) data sets but considerably lower than for JdF (69.8%). A possible reason for the higher proportion for JdF is the different procedure for preparation of the JdF sample. However, it may also indicate the presence of a higher proportion of known microbial groups at JdF relative to the other sites.

Taxonomic and functional assignments of the predicted genes from SP, JdF, MT, and BTB were based on alignments to the NR database using the lowest common ancestor algorithm implemented in MEGAN ([20]). The results are summarized in Table S2 in the supplemental material. For all data sets for which gene products had at least one blastp hit in the NR database, a taxonomic assignment was possible in most cases (for SP, 94.3%), while the functional annotation had a lower degree of success (for SP; SEED, 40.0%; KEGG, 58.4%; and COG, 73.0%). The better performance of the functional annotation with KEGG and COG than with SEED can be in part attributed to our update of the built-in mapping files provided by MEGAN, with functional annotations from UniProt ([28]), eggNOG ([27]), and KEGG Orthology ([26]). This update affected KEGG and COG but not SEED.

Placing SP into context with other vent and subsurface metagenomes. The metagenome of the SP chimney was considerably different from the metagenomes from the other environments.
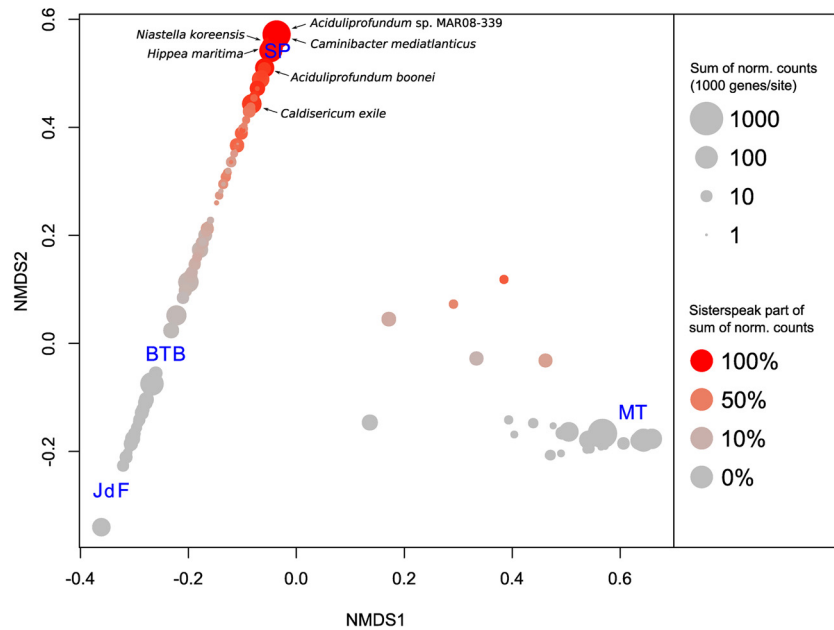
FIG 2 Nonmetric multidimensional scaling plot for the tested metagenomes from SP, JdF, MT, and BTB, using the Bray-Curtis dissimilarity index, computed from taxonomic assignments at the species level. Species scores are represented as bubbles, sized by the sum of normalized counts (at 1,000 counts/site) in the four sites, and colored using a gradient from red to gray, where the red intensity represents the proportion of the SP counts.

We analyzed the alpha diversity of the four sites by computing rarefaction curves based on the taxonomic assignments, summarized at different taxonomic rank levels (see Fig. S2 in the supplemental material). The results suggest that the diversity of the community in the SP chimney is considerably lower than in the JdF chimney and in the BTB subsurface sediments (on the species, genus, family, and order levels). This may be a consequence of local SP parameters selecting against a high microbial diversity. Furthermore, the composition of the community in SP was clearly distinct from that of the other analyzed sites (Fig. 2; Table 1). In the following, we attempt to link these differences to known abiotic parameters for the four sites.

One feature that so far has been observed only in the SP environment is sudden temperature bursts, reaching 464°C (12). In the SP microbiome, genes related to archaeal DNA replication were found in higher proportions than in the other metagenomes, but the same does not hold for genes related to bacterial DNA replication (Fig. 3A to C). Hence, distinct DNA replication genes are probably more important in the microhabitats colonized by the SP archaea than in those in which bacteria thrive. Comparing cultured taxa of archaea and bacteria to which the SP genes were assigned (Fig. 4), we noticed that most of the archaeal cardinal growth temperatures are considerably hotter than those of bacteria (46–50). Since genes for DNA replication were enriched only for archaea, and the archaea identified in the SP sample likely colonize hotter regimes than the detected bacteria, this may be an adaptation to sudden high temperature bursts. Hence, a more diverse range of enzymes for DNA replication may be a strategy to deal with sudden heat bursts that have been encountered for the SP chimney. Further indicators of cellular mechanisms reacting to increased rates of thermal damage in the SP metagenome relative to JdF, MT, and BTB are the enrichments of DNA sequences encoding aminoacyl-tRNAs, purine and pyrimidine (nucleotide) metabolisms, and ribosomes in the metagenome (Fig. 3D to F). Aminoacyl-tRNAs are fundamental elements in protein biosynthesis and are particularly sensitive to high temperatures (2). The maintenance of nucleic acids' primary structure is problematic at high temperature, since they are subjected to deamination, depurination, and single- and double-strand breaks (2). Enrichment of DNA encoding ribosomal structures may be a consequence of a dramatic increase in protein degradation at high temperatures.

Besides temperature, other factors expected to impact the micro-

TABLE 1 Absolute and relative counts (numbers of taxonomically assigned genes) of the six most abundant SP species at SP, JdF, MT, and BTB

| Organism(s)	SP		JdF		MT		BTB	
	Count	%	Count	%	Count	%	Count	%
All	11,419	100	65,899	100	6,597	100	17,434	100
Aciduliprofundum sp. MAR08-339	1,341	11.74	0	0.00	0	0.00	0	0.00
Hippea maritima	429	3.76	0	0.00	0	0.00	3	0.02
Niastella koreensis	315	2.76	5	0.01	0	0.00	0	0.00
Aciduliprofundum boonei	183	1.60	0	0.00	0	0.00	6	0.03
Caldisericum exile	172	1.51	3	0.00	0	0.00	22	0.13
Caminibacter mediatlanticus	162	1.42	0	0.00	0	0.00	0	0.00
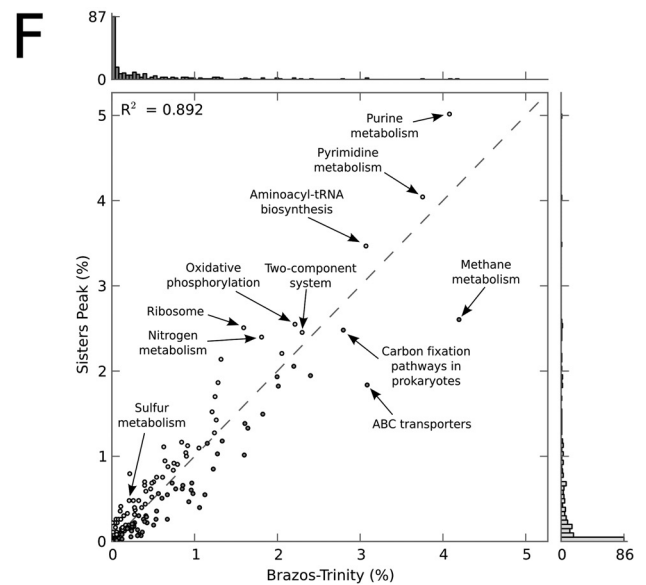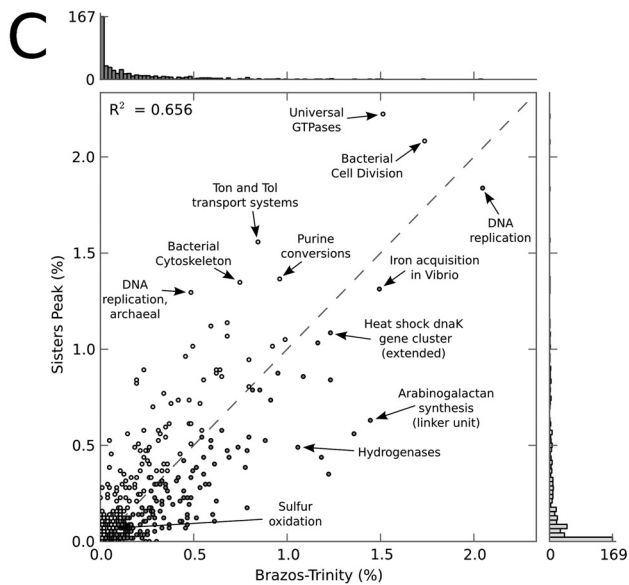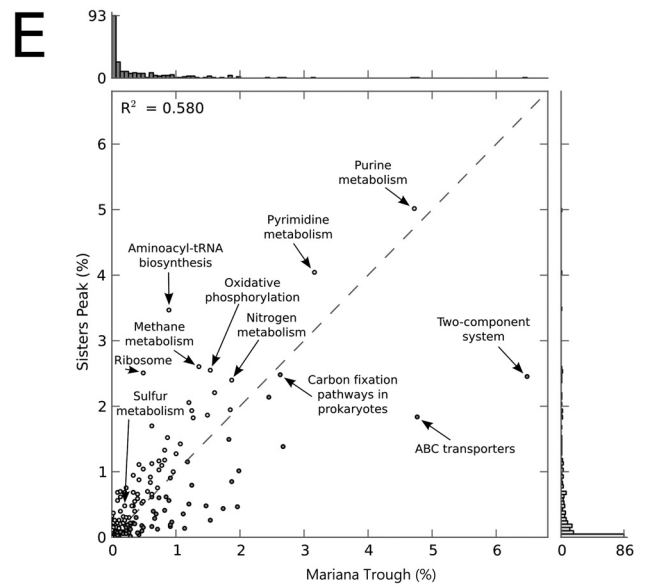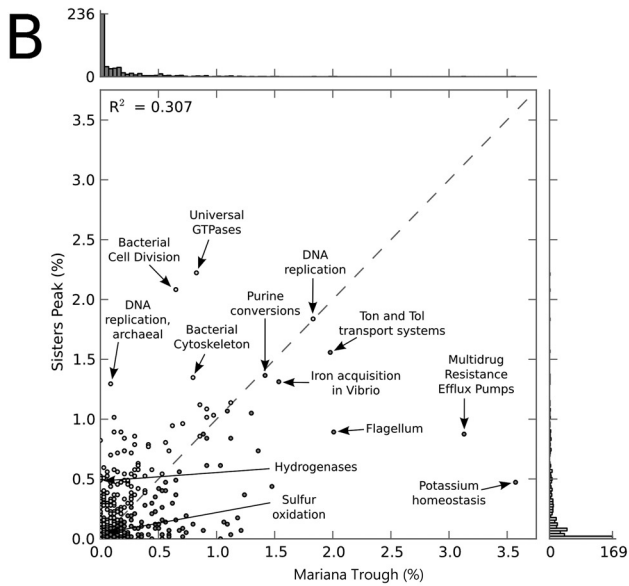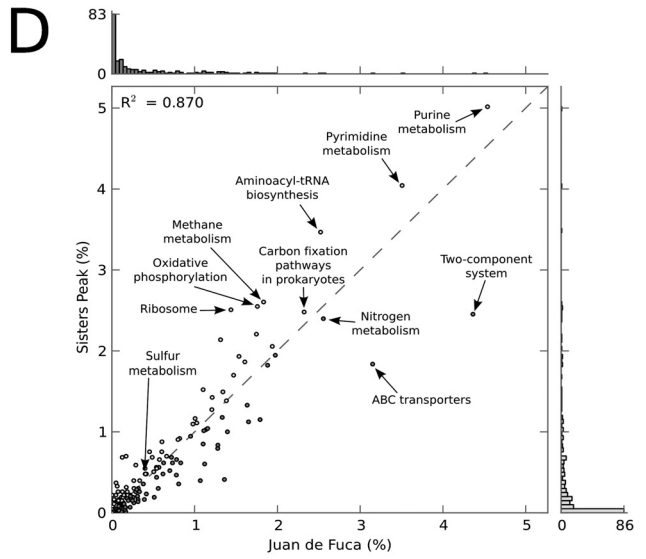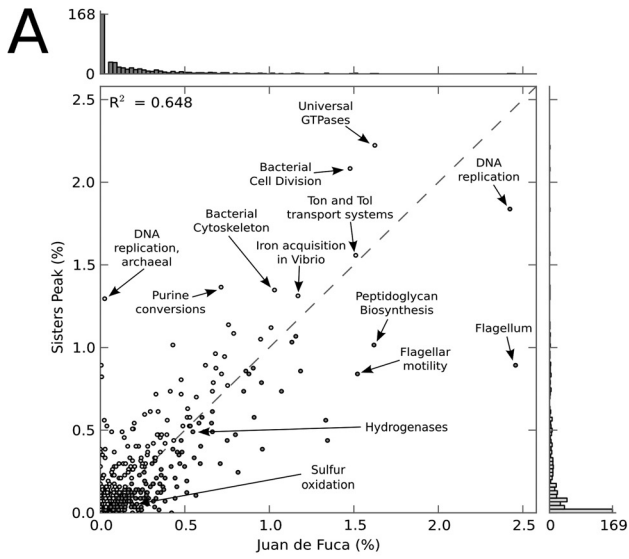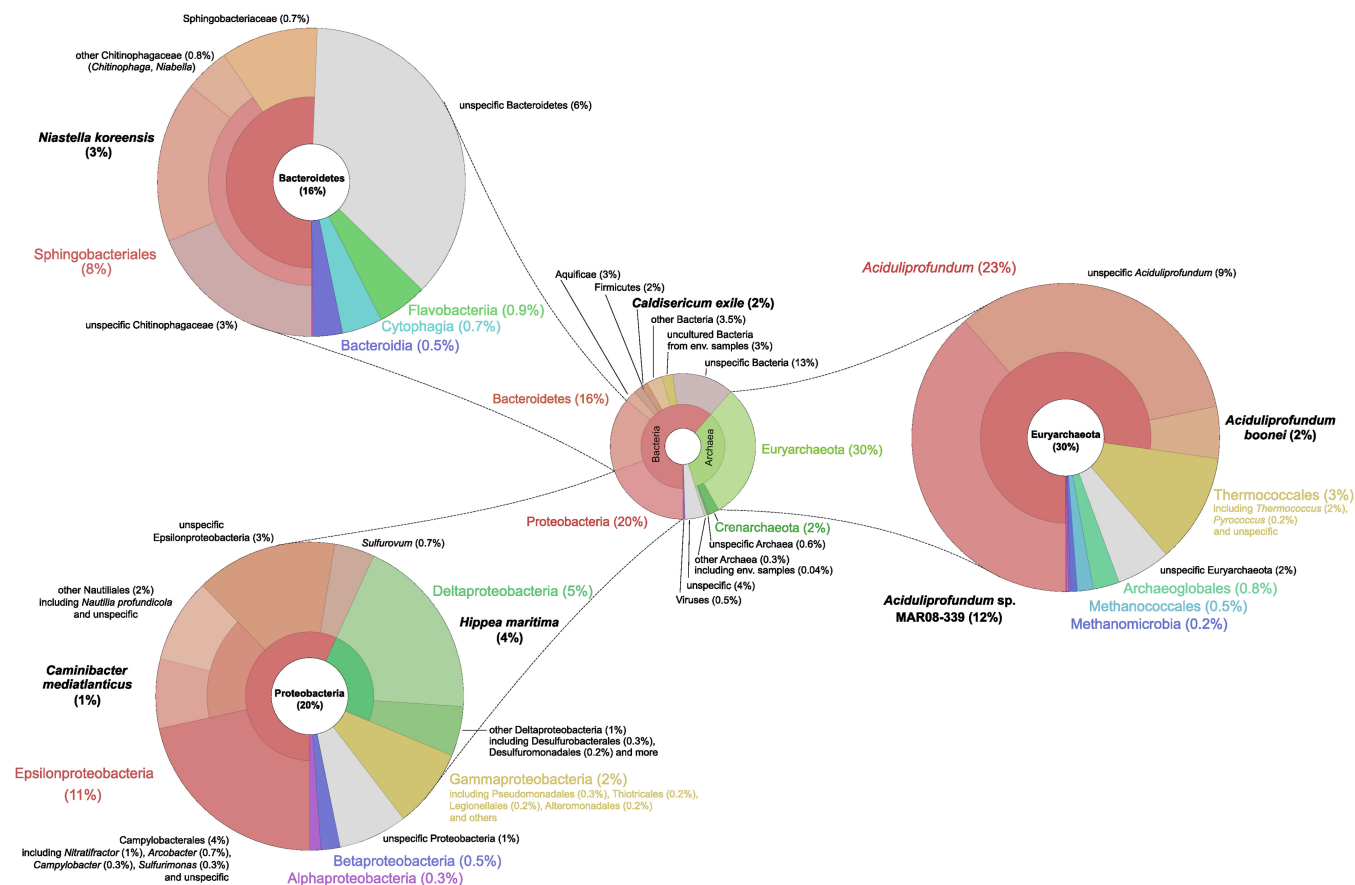
FIG 4 Community composition of the SP metagenome, according to the taxonomic assignments by MEGAN. In all four pies, the relative values (percentage) refer to the total number of taxonomic assigned predicted genes.

bial community in venting habitats include the abundance of inorganic electron donors, like hydrogen, and sulfide or other energy sources like methane for chemosynthetic growth (6, 51, 52). We have summarized the fluid chemical data that are available for the different metagenomic sample areas in Table 2. Hydrogen concentrations in SP hydrothermal fluids (H_2 in hydrothermal endmember, 1.6 mM [R. Seifert, personal communication]) are higher than in most other basalt-hosted vent systems, including those along the JdF ridge (H_2 in hydrothermal endmembers, ≤0.96 mM) (52, 53). A higher diversity among genes encoding hydrogen-metabolizing functions (hydrogenases) has been detected before in locations with elevated hydrogen (5). However, we could not identify a higher diversity of genes related to hydrogen metabolism in the SP than in the JdF chimney (Fig. 3A). Unfortunately, no hydrogen concentrations are available for the samples from MT and BTB to which we could relate the respective metagenomic information. Sulfide concentrations in JdF vents can be considerably higher than what was determined for endmember SP hydrothermal fluids, but this does not appear to be reflected in a considerable overrepresentation of sulfur oxidation or metabolism genes in the JdF chimney (Fig. 3A and D). Despite lacking sulfur data

for the sites at MT and BTB, from which the metagenomes derived, it is highly unlikely that concentrations of reduced sulfur compounds are comparable at SP and these different habitats (8, 9, 14). Hence, it is somewhat surprising that no distinct over- or underrepresentation of sulfur-oxidizing and -metabolizing genes in general could be recognized in the SP metagenome relative to the MT and BTB microbiomes (Fig. 3B, C, E, and F). Given the very low methane concentrations in SP hydrothermal fluids (CH_4 in hydrothermal endmember, 0.05 mM [Seifert, personal communication]) relative to what has been recognized in fluids emanating from JdF vents (CH_4 in hydrothermal endmember, ≤1.7 mM) (53), the overrepresentation of genes encoding methane metabolism in the SP metagenome rather than in the JdF microbiome is unexpected (Fig. 3D). This suggests that other factors besides methane availability alone govern the community structure at these locations. This is supported by an overrepresentation of genes encoding methane-metabolizing functions in the metagenome of BTB sediments relative to the SP metagenome (Fig. 3F), which does not coincide with higher methane concentrations in the depth interval from which the metagenomic BTB data set derived (54). We are not aware of any methane concentration data

FIG 3 Scatter plots comparing the functional assignments to SEED Subsystems (A to C) and KEGG Pathways (D to F) for SP and JdF (A and D), SP and MT (B and E), and SP and BTB (C and F). The axes represent the proportion of assignments to each functional category in the two metagenomes compared. On each axis a binned histogram of the assignments is depicted (light gray, SP; dark gray, other metagenome). The coefficient of determination (R^2) is reported on each plot.

TABLE 2 Abiotic properties for metagenomic sampling areas[a]

Sample site	Sample type	Temp (°C)	Concn (mM) of:			Reference(s)
			Hydrogen	Methane	Sulfide	
Marine hydrothermal vent samples						
Sisters Peak (southern Mid-Atlantic Ridge)	Sulfide chimney	464	1.6	0.05	8.2–10.5	12, 64; Seifert, personal communication
Juan de Fuca (Eastern Pacific Ridge)[b]	Sulfide chimney	316	≤0.96	≤1.7	≤25	8, 52, 53
Mariana Trough (western Pacific backarc basin)	Hydrothermal fluids	106	No data	No data	No data	9
Marine nonvent samples						
Brazos-Trinity Basin (8 mbsf) (Gulf of Mexico)	Sediment	No data	No data	<0.001	No data	14, 54

[a] SP and JdF concentrations are for hydrothermal fluid endmembers; i.e., magnesium = 0. In case of methane concentrations for BTB, measured values are displayed because these liquids cannot be calculated into endmembers.

[b] Fluid chemistry is not from the chimney the metagenomic analyses was performed with but from a hot venting, chemically comparable site in the JdF vent field.

available for the fluids that were collected for the MT metagenome study (9). Hence, we cannot establish a connection between the observed overrepresentation of SP genes encoding methane metabolism compared to those from the MT sample (Fig. 3E).

Although some of the chemical data are not available for comparison, it appears that the tested electron donors for which data are available are not the parameters that segregate SP from the other analyzed environments. However, the extreme temperature situation may be a vital factor selecting for the SP community. While it is unlikely that the organisms directly reside in the outlet of the SP chimney, where 464°C fluid pulses emanate, the effects of such sudden temperature bursts likely reach throughout the chimney. Hence, not only organisms residing in the hotter regions but also those colonizing cooler chimney regions may be impacted by a sudden temperature rise. Adaptations of the microbes from the multiple temperature regimes would probably evoke different strategies to deal with the short-term rapidly changing conditions.

The SP chimney biome. One of the basic questions which typically arise when analyzing metagenome data is, how many different genomes are in there? An estimation is in many cases possible by using a binning algorithm. However, most current binning algorithms, e.g., MetaCluster (55), rely on coverage information in order to separate the sequences, an option which is undesirable for samples treated with MDA. In other methods, e.g., Scimm (56), the number of bins is an input parameter, and therefore, these methods cannot be employed to estimate the number of species. A binning algorithm which does not require redundancy information or a prior estimation of the number of species is implemented in the tool Metawatt (32). Using the "high confidence" settings, it divides the Sisters Peak assembly into 47 bins. This number is significantly lower than the number (140) of species-level taxonomic assignments performed by MEGAN (see Fig. S2 in the supplemental material). However, for many species, MEGAN assigns only a very few genes: e.g., only 46 species have more than 8 genes assigned (0.2% of the species-level assignments). Thus, the discrepancy likely arises from the fact that not enough sequences are available to collect composition statistics for reliable *de novo* binning.

Although the SP chimney sample does not appear to be as diverse as other hydrothermal vent or sediment subsurface samples (see Fig. S2 in the supplemental material), the identified microorganisms (Fig. 4) are commonly found in hydrothermal venting habitats (57–59). The majority of the protein-coding genes of the investigated SP metagenome were assigned to bacteria (61%)

(Fig. 4). Genes assigned to archaea and viruses accounted for 34% and 0.5%, respectively (Fig. 4). Approximately 3% were assigned to environmental sequences. The bacterial genes were mostly grouped into *Proteobacteria* (30%), with *Epsilonproteobacteria* representing the dominant fraction (11%), and *Bacteroidetes* (16%) (Fig. 4). Most of the archaeal genes resembled those of *Euryarchaeota* (30%), with representatives of *Aciduliprofundum* (23%) and *Thermococcales* (3%) being the prevailing organisms (Fig. 4). The six most abundant microbial species represented by genes in the SP metagenome were *Aciduliprofundum* sp. MAR08-339 (*Euryarchaeota*) (12%), *Hippea maritima* (*Deltaproteobacteria*) (4%), *Niastella koreensis* (*Bacteroidetes*) (3%), *A. boonei* (*Euryarchaeota*) (2%), *Caldisericum exile* (*Caldisericia*) (2%), and *Caminibacter mediatlanticus* (*Epsilonproteobacteria*) (1.4%) (Fig. 4; Table 1). It is noteworthy that these six species were not detected or had a very low coverage in the metagenomes of JdF, MT, and BTB (Fig. 2; Table 1). Besides *Niastella*, all these affiliates have been isolated from hydrothermal vents or hot springs (47, 48, 50, 60). While *Niastella* is typically cultivated from soil (61), gene sequences related to it were also identified in the JdF metagenome (Table 1).

Aciduliprofundum species are likely a dominant component of the SP chimney biome (Fig. 2 and 4). A further confirmation of this was obtained by taxonomic classification of 16S and 23S rRNA genes predicted in the metagenome (see Table S3 in the supplemental material): the most common assignments (24%) were to the DHVE2 group, which includes *Aciduliprofundum*. A strong dominance of *Aciduliprofundum* coincides with previous findings which demonstrate that affiliates of this group can account for up to 15% of the 16S rRNA gene sequences in hydrothermal habitats (50, 59). This genus appears to be widespread in hydrothermal vent environments, and the only published representative cultured so far has been described as an obligate thermoacidophilic sulfur- and iron-reducing heterotroph (ferments peptides as a primary metabolic pathway) (50, 59), making it perfectly well adapted to deal with the extreme conditions in hydrothermal venting chimneys. An analysis of the variability found in SP metagenomic sequences recruited to the genomes of *Aciduliprofundum* sp. MAR08-339 and *A. boonei* (see Fig. S3 in the supplemental material) showed that several variants of the same genomic regions were often found; these may give rise to different phenotypes within the same species. Furthermore, SP metagenome sequences were aligned to a significant portion of the *Aciduliprofundum* sp. MAR08-339 genome (blastn, 32.4%;

tblastx, 42,2% [see Fig. S4A in the supplemental material]) and, to a lower degree, to the *A. boonei* genome (blastn, 2.8%; tblastx, 24.3% [see Fig. S4B]). This supports the assumption that *Aciduliprofundum* species are highly abundant in the sampled habitat. Despite this generally good coverage of the *Aciduliprofundum* sp. MAR08-339 genome, some large regions were absent or only a low level of similarity could be found in the SP metagenome sequences (see Fig. S4A and Table S4). For example, the blastn alignment of the SP assembly to the *Aciduliprofundum* sp. MAR08-339 reference genome revealed that 30 regions larger than 5 kbp were not covered by any hit. Although in some cases, this will be the effect of random sampling, it is interesting to observe that five of these large absent regions were predicted genomic islands (see Fig. S4A). Alignments to the NCBI NR database (see Table S4) showed that some of the genes in these predicted genomic islands also lack significant hits in the annotated genes of *A. boonei*, the closest cultured and sequenced representative. Among these, for some genes no significant similarity in the SP metagenome could be found (e.g., AciM339_0807 to AciM339_0809 [see Table S4]), while in some other cases (e.g., AciM339_0800 to AciM339_0805, AciM339_1221 to AciM339_1224, or AciM339_1350 to AciM339_1353), significant protein-level similarity was found for the SP metagenome by tblastx alignments. Although other explanations cannot be excluded, regions missing both in *A. boonei* and in the SP metagenome indicate that these DNA sequences have been taken up only by certain *Aciduliprofundum* specimens. Among the genes absent in *A. boonei* but with tblastx hits in SP, alignments to the NCBI NR database showed significant similarities to, e.g., mesophilic and thermophilic methanogens, other thermophilic archaea, the hyperthermophilic bacterium *Thermotoga*, and mesophilic ammonia-oxidizing or sulfate-reducing bacteria (see Table S4).

Besides the genomes of *Aciduliprofundum*, other genomes with a high coverage in the SP metagenome were those of *Hippea*, *Caldisericum*, and *Caminibacter*. These bacteria are also thermophiles and have been described as anaerobes capable of reducing sulfur or thiosulfate (46–48, 60). *Hippea* and *Caminibacter* can grow on hydrogen, but only *Caminibacter* can grow autotrophically (47, 60). *C. exile* is the first cultured representative of the originally known candidate division OP5, and it has been described as a chemoheterotrophic scavenger which contributes to the sulfur cycle (48, 62). *N. koreensis* is the only organism among the most abundant species of the SP metagenome that is commonly associated with moderate temperatures, and a heterotrophic lifestyle has been shown for this organism (61). Like for the *Aciduliprofundum*-related SP genes, mapping of the SP metagenome to *H. maritima*, *C. exile*, and *N. koreensis* demonstrated relatively high sequence variability for certain genes (see Fig. S4C to E in the supplemental material). Since the SP metagenome did not cover these genomes as extensively as the *Aciduliprofundum* genomes, it is difficult to point to gene regions that might be missing in the SP genomes. Given that for *C. mediatlanticus* no complete genome sequence is available, we excluded it from this analysis.

In summary, with the exception of *Niastella*, the most abundant organisms in the SP chimney appear to be thermophilic anaerobes, which can grow chemolithoautotrophically or chemolithoheterotrophically on hydrogen or ferment organic substrates and are thus likely associated with the hotter, more reduced SP chimney realms. However, while these microbial physiologies match the conditions expected to pertain to the SP chimney, these

conditions can also be found in other active sulfide venting chimneys (63) and are not sufficient to explain the high coverage of the above-mentioned species in the SP chimney microbiome.

Functions associated with the biome of the most abundant species at SP. To understand if particular microbial strategies could be associated with different organisms, we compared the functional annotations, in terms of assignments in the KEGG, COG, and SEED hierarchies, of genes assigned to a specific taxonomic group with those in the rest of the data set. We analyzed the functional profiles (KEGG, COG, and SEED) of five genera in SP (*Aciduliprofundum*, *Hippea*, *Niastella*, *Caminibacter*, and *Caldisericum*) in comparison to the rest of the data set; these genera are those to which the most common species in SP belong (Fig. 4; Table 1). The analyses were based on assignments at the genus level, and not the species level, to avoid potential biases from the presence of two *Aciduliprofundum* species (*Aciduliprofundum* sp. MAR08-339 and *A. boonei*) among the most frequent taxonomical assignments.

The functional profile of SP *Aciduliprofundum* was quite different from the profiles of the other most abundant species in the SP metagenome (Fig. 5). The higher abundance of *Aciduliprofundum* genes than for other taxa resulted in a more exhaustive list of functional profiles with a statistically significant difference (Fig. 5A). Genes overrepresented in *Aciduliprofundum* included mostly genes encoding enzymes involved in protein biosynthesis, nucleotide metabolism, and energy production or conversion (Fig. 5A, KEGG subcategories and COG functional groups). This indicates their importance for a successful inhabitance of the chimney in the *Aciduliprofundum*-colonized temperature regime. Since both DNA and proteins are sensitive to high temperatures (2), enrichment of genes involved in protein biosynthesis and nucleotide metabolism may be a strategy to deal with sudden high temperature bursts. Based on *A. boonei*'s physiology and the possible high abundance of species grouped within *Aciduliprofundum*, it has been suggested that this lineage may be a key player in sulfur and iron cycling at deep-sea vents (50). In the SP chimney, highly overrepresented genes of energy production or conversion for *Aciduliprofundum*-like sequences in the SP metagenome were mostly grouped into one-carbon metabolism, carbohydrate metabolism, and electron-donating reactions.

Subcategories that were overrepresented in *Hippea* and *Caminibacter* included in particular enzymes responsible for cell motility but also some involved in cellular processes and signaling as well as DNA replication and repair (Fig. 5B and C). A relatively large number of genes for cell motility and repair systems has also been found in the JdF chimney metagenome (8). In this environment, the high proportion of genes for cell mobility was considered to reflect the highly dynamic conditions in the chimney, while the prevalence of repair systems was believed to be advantageous to adapt to the highly mutagenic environmental conditions in the vent (8).

No significant over- or underrepresentation of functional categories was recognized for *Caldisericum* (Fig. 5D). Among the genes assigned to *Niastella*, those involved in the biogenesis of cell wall and membrane were overrepresented compared to the rest of the data set (Fig. 5E). Given that cultivated members of *Niastella* are associated with moderate temperatures (61), their relatives in the SP chimney are likely to colonize outer, i.e., cooler chimney regimes. As the sudden extreme high temperature bursts in the inner chimney conduit likely impact also other chimney regions, a
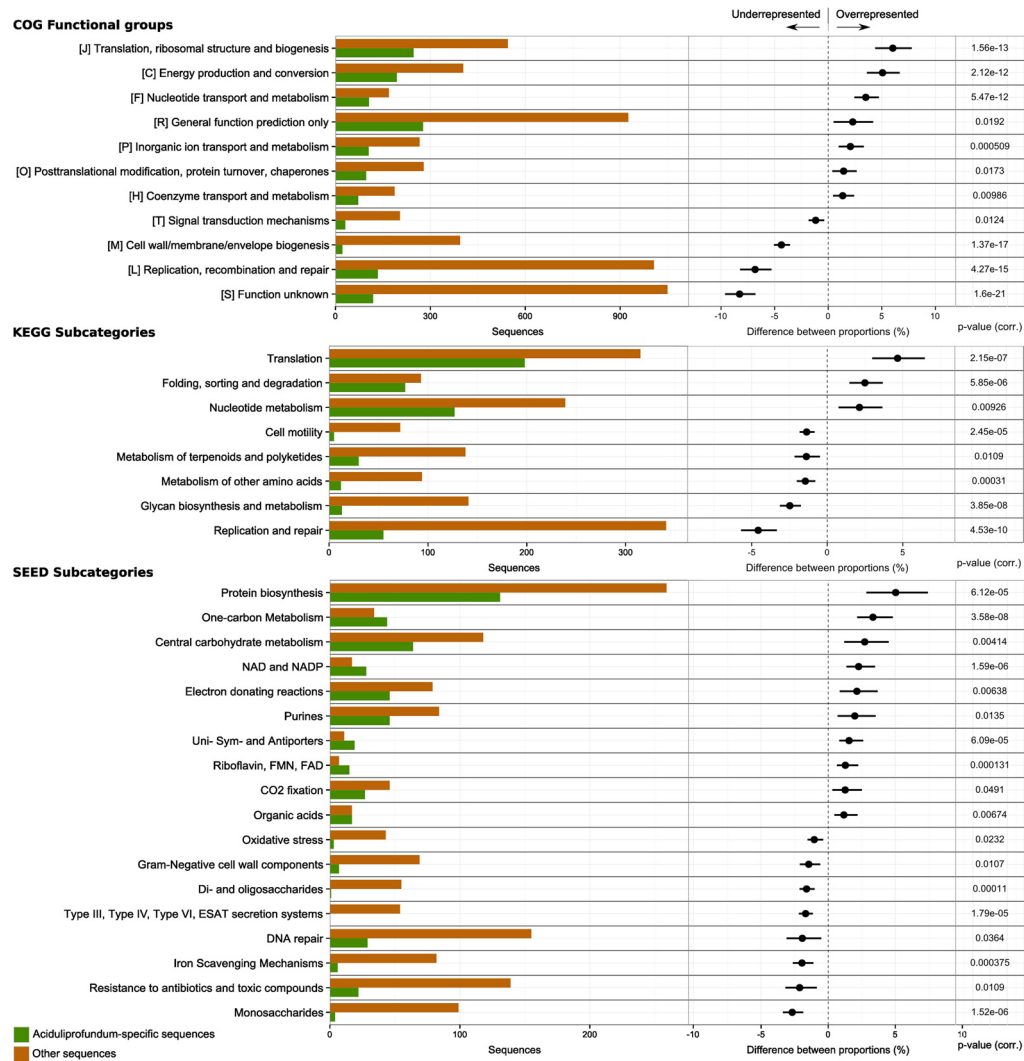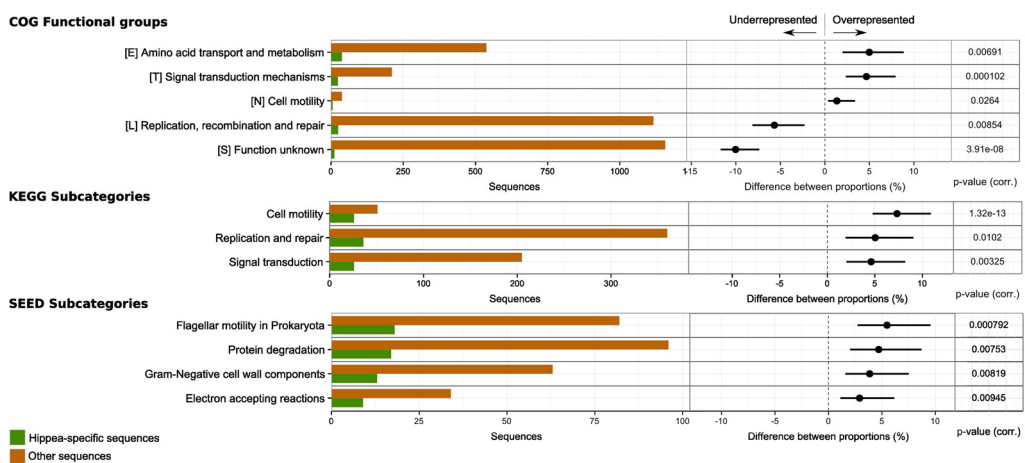
FIG 5 Differential analysis of the functional assignments for subsets of the genes assigned to taxonomic groups, compared to the rest of the SP data set. The analyses results are shown for the five most abundant genera—*Aciduliprofundum* (A), *Hippea* (B), *Caminibacter* (C), *Caldisericum* (D), and *Niastella* (E)—and for uncultured taxa (F). Each row shows absolute counts (left) and differences between relative counts (right) for a functional category. In particular, the histograms on the left report the total number of genes assigned to the functional category in the taxonomic group of interest (green) and in the rest of the data set (orange). The dots in the bar plots on the right represent the difference (taxonomic group of interest minus the rest of the data set) between the proportions of genes assigned to the functional category, relative to the total counts in the subset. The extent of the bar represents the 95% confidence intervals of the difference between proportions. Only data with a corrected *P* value of ≤0.05 (reported on the right) and a difference between proportions of ≥1.0% are reported.
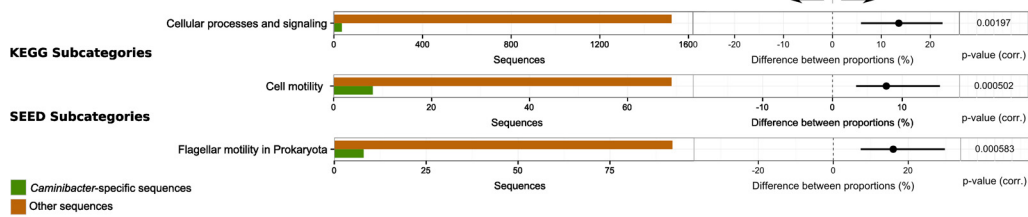
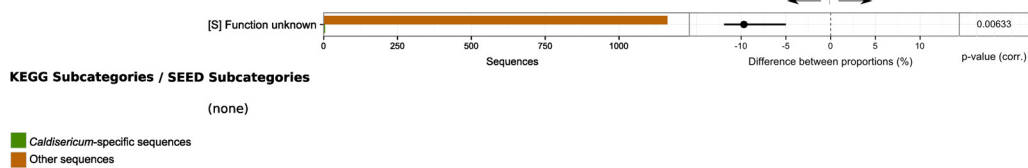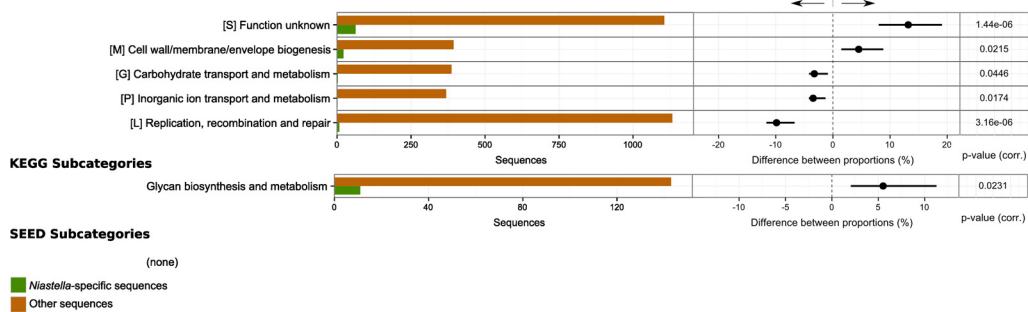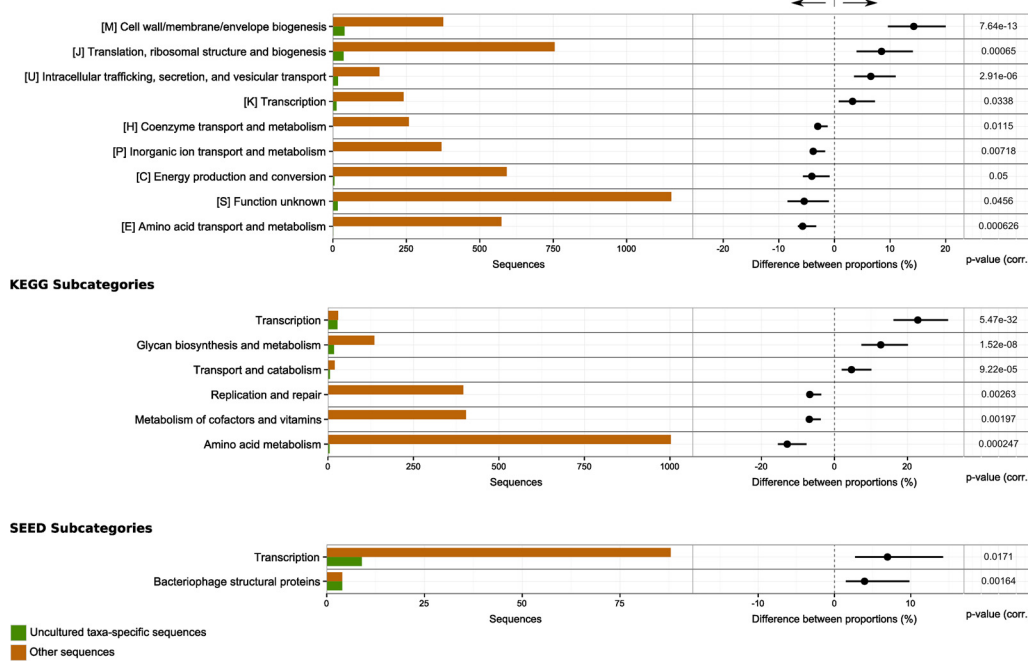FIG 5 continued

possibility is that *Niastella*'s strategy to cope with these bursts may be related to cell wall and membrane biogenesis. Alternatively, in these temperature zones, temperature variations may be subordinate in importance and other factors, e.g., those related to microbe-microbe interactions, may be more crucial for successful inhabitance.

Functions associated with uncultured taxa. A particularly interesting category of organisms is represented by those which are still uncultured. We created a list of uncultured taxa by text mining on the NCBI taxonomy denomination (see Materials and Methods). This list gives us an opportunity to gain insights into a group of genes from uncultured species. However, the list has no pretense to be exhaustive: e.g., gene assignment to a taxa of a higher rank (e.g., phylum), which is not annotated as "uncultured" (i.e., some cultivated species in the group are known), does not guarantee that the gene belongs to a cultured taxon. Furthermore, the current sequence databases contain little information for most uncultured taxa, which makes the taxonomical and functional analysis more difficult. Notwithstanding these limitations, we further analyzed the genes specifically assigned to this group (hereafter called uncultured taxa), which made up around 3% of the complete set of taxonomically assigned genes.

In analogy to analyses of the most common genera, we performed a differential analysis of the functional annotations of genes assigned to uncultured taxa, in comparison to the rest of the data set. Similar to the situation observed for *Niastella*, sequences from the uncultured taxa were also enriched in genes encoding cell wall and membrane biogenesis (Fig. 5F). Additionally, genes encoding transcription and translation processes as well as intracellular trafficking and secretion were also overrepresented in the group (Fig. 5F). This included in particular genes encoding functions for cell division (FtsI) (COG0768) and the preprotein translocase subunit SecD (COG0342).

Conclusions. We here provide a metagenomics-based overview of the microbial community in the SP chimney and its functional potential. To identify special features of the SP biome, we compared our data with those for other pyrosequenced metagenomes from hydrothermal habitats at JdF and MT as well as from marine sediments in the BTB. Since a clear causal relation between known fluid chemistry and abundance of genes in related metabolic pathways could not be observed, the temperature bursts previously observed at the SP site, which probably affect adjacent chimney regions as well, are likely the most important factor responsible for the local microbial adaptations. The most abundant genes of thermophiles associated with the SP chimney included those with similarity to *Aciduliprofundum* sp. MAR08-339 (12%), *H. maritima* (4%), *A. boonei* (2%), *C. exile* (2%), and *C. mediatlanticus* (1%). These species were not present or were rare in the JdF, MT, and BTB metagenomes. Enrichment profiles of functional annotations of the SP metagenome genes assigned to these taxa showed considerable differences and may reflect different strategies for surviving the extreme conditions in the SP chimney. These strategies appear to be dependent on the temperature regimes that the organisms likely colonize. Among the analyzed microbial groups, *Aciduliprofundum* (optimal growth temperature [T_{opt}], 70°C) can be expected to colonize the hottest environments (50), and genes for protein biosynthesis and nucleotide and energy metabolism were enriched. In contrast, *Hippea* (T_{opt}, 52 to 54°C) and *Caminibacter* (T_{opt}, 55°C) have lower optimal growth temperatures (47, 60) and thus can be expected to inhabit still

warm but cooler regions than *Aciduliprofundum*. Overrepresentation of *Hippea* and *Caminibacter* genes encoding enzymes of cell motility, cellular processes and signaling, and DNA replication and repair may be a hallmark for these genera and the respective inhabited temperature regime at SP. *Niastella* is adapted to even cooler temperatures (10 to 37°C) and thus probably inhabits the outer SP chimney areas. Here different adaptations would be required than for the hotter regions near the inside and may be reflected by the apparent enrichment of genes for cell wall and membrane biogenesis. However, in these colder regions, other properties (e.g., microbe-microbe interaction), rather than temperature, may become more important.

REFERENCES

1. **Jannasch HW, Mottl MJ.** 1985. Geomicrobiology of deep-sea hydrothermal vents. Science **229:**717–725. http://dx.doi.org/10.1126/science.229.4715.717.
2. **Charlier D, Droogmans L.** 2005. Microbial life at high temperature, the challenges, the strategies. Cell. Mol. Life Sci. **62:**2974–2984. http://dx.doi.org/10.1007/s00018-005-5251-8.
3. **Flores GE, Shakya M, Meneghin J, Yang ZK, Seewald JS, Geoff Wheat C, Podar M, Reysenbach AL.** 2012. Inter-field variability in the microbial communities of hydrothermal vent deposits from a back-arc basin. Geobiology **10:**333–346. http://dx.doi.org/10.1111/j.1472-4669.2012.00325.x.
4. **Flores GE, Campbell JH, Kirshtein JD, Meneghin J, Podar M, Steinberg JI, Seewald JS, Tivey MK, Voytek MA, Yang ZK, Reysenbach AL.** 2011. Microbial community structure of hydrothermal deposits from geochemically different vent fields along the Mid-Atlantic Ridge. Environ. Microbiol. **13:**2158–2171. http://dx.doi.org/10.1111/j.1462-2920.2011.02463.x.
5. **Perner M, Petersen JM, Zielinski F, Gennerich H-H, Seifert R.** 2010. Geochemical constraints on the diversity and activity of H₂-oxidizing microorganisms in diffuse hydrothermal fluids from a basalt- and an ultramafic-hosted vent. FEMS Microbiol. Ecol. **74:**55–71. http://dx.doi.org/10.1111/j.1574-6941.2010.00940.x.
6. **Perner M, Hentscher M, Rychlik N, Seifert R, Strauss H, Bach W.** 2011. Driving forces behind the biotope structures in two low-temperature hydrothermal venting sites on the southern Mid-Atlantic Ridge. Environ. Microbiol. Rep. **3:**727–737. http://dx.doi.org/10.1111/j.1758-2229.2011.00291.x.
7. **Klevenz V, Sander S, Perner M, Koschinsky A.** 2012. Amelioration of free copper by hydrothermal vent microbes as a response to high copper concentrations. Chem. Ecol. **28:**405–420. http://dx.doi.org/10.1080/02757540.2012.666531.
8. **Xie W, Wang F, Guo L, Chen Z, Sievert SM, Meng J, Huang G, Li Y, Yan Q, Wu S, Wang X, Chen S, He G, Xiao X, Xu A.** 2011. Comparative metagenomics of microbial communities inhabiting deep-sea hydrothermal vent chimneys with contrasting chemistries. ISME J. **5:**414–426. http://dx.doi.org/10.1038/ismej.2010.144.
9. **Nakai R, Abe T, Takeyama H, Naganuma T.** 2011. Metagenomic analysis of 0.2-μm-passable microorganisms in deep-sea hydrothermal fluid. Mar. Biotechnol. **13:**900–908. http://dx.doi.org/10.1007/s10126-010-9351-6.
10. **Wang F, Zhou H, Meng J, Peng X, Jiang L, Sun P, Zhang C, Van Nostrand JD, Deng Y, He Z, Wu L, Zhou J, Xiao X.** 2009. GeoChip-based analysis of metabolic diversity of microbial communities at the Juan de Fuca Ridge hydrothermal vent. Proc. Natl. Acad. Sci. U. S. A. **106:**4840–4845. http://dx.doi.org/10.1073/pnas.0810418106.

11. **Grzymski JJ, Murray AE, Campbell BJ, Kaplarevic M, Gao GR, Lee C, Daniel R, Ghadiri A, Feldman RA, Cary CS.** 2008. Metagenome analysis of an extreme microbial symbiosis reveals eurythermal adaptation and metabolic flexibility. Proc. Natl. Acad. Sci. U. S. A. **105**:17516–17521. http://dx.doi.org/10.1073/pnas.0802782105.

12. **Koschinsky A, Garbe-Schönberg D, Sander S, Schmidt K, Gennerich HH, Strauss H.** 2008. Hydrothermal venting at pressure-temperature conditions above the critical point of seawater, 5°S on the Mid-Atlantic Ridge. Geology **36**:615–618. http://dx.doi.org/10.1130/G24726A.1.

13. **Streit W, Bjourson A, Cooper J, Werner D.** 1993. Application of subtraction hybridization for the development of a *Rhizobium leguminosarum* biovar phaseoli and *Rhizobium tropici* group specific DNA Probe. FEMS Microbiol. Ecol. **13**:59–67. http://dx.doi.org/10.1111/j.1574-6941.1993.tb00051.x.

14. **Biddle JF, White JR, Teske AP, House CH.** 2011. Metagenomics of the subsurface Brazos-Trinity Basin (IODP site 1320): comparison with other sediment and pyrosequenced metagenomes. ISME J. **5**:1038–1047. http://dx.doi.org/10.1038/ismej.2010.199.

15. **Niu B, Fu L, Sun S, Li W.** 2010. Artificial and natural duplicates in pyrosequencing reads of metagenomic data. BMC Bioinformatics **11**:187. http://dx.doi.org/10.1186/1471-2105-11-187.

16. **Schmieder R, Edwards R.** 2011. Quality control and preprocessing of metagenomic datasets. Bioinformatics **27**:863–864. http://dx.doi.org/10.1093/bioinformatics/btr026.

17. **Fu L, Niu B, Zhu Z, Wu S, Li W.** 2012. CD-HIT: accelerated for clustering the next generation sequencing data. Bioinformatics **28**:3150–3152. http://dx.doi.org/10.1093/bioinformatics/bts565.

18. **Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen Y-J, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer MLI, Jarvie TP, Jirage KB, Kim J-B, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM.** 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature **437**:376–380. http://dx.doi.org/10.1038/nature03959.

19. **Rho M, Tang H, Ye Y.** 2010. FragGeneScan: predicting genes in short and error-prone reads. Nucleic Acids Res. **38**:e191. http://dx.doi.org/10.1093/nar/gkq747.

20. **Huson D, Mitra S, Weber N, Ruscheweyh H, Schuster S.** 2011. Integrative analysis of environmental sequences using MEGAN4. Genome Res. **21**:1552–1560. http://dx.doi.org/10.1101/gr.120618.111.

21. **NCBI Resource Coordinators.** 2013. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. **41**:D8–D20. http://dx.doi.org/10.1093/nar/gks1189.

22. **Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25**:3389–3402. http://dx.doi.org/10.1093/nar/25.17.3389.

23. **Ondov BD, Bergman NH, Phillippy AM.** 2011. Interactive metagenomic visualization in a Web browser. BMC Bioinformatics **12**:385. http://dx.doi.org/10.1186/1471-2105-12-385.

24. **Mitra S, Rupek P, Richter DC, Urich T, Gilbert JA, Meyer F, Wilke A, Huson DH.** 2011. Functional analysis of metagenomes and metatranscriptomes using SEED and KEGG. BMC Bioinformatics **12**:S21. http://dx.doi.org/10.1186/1471-2105-12-S1-S21.

25. **Overbeek R, Begley T, Butler R, Choudhuri J, Chuang H-Y, Cohoon M, de Crécy-Lagard V, Diaz N, Disz T, Edwards R, Fonstein M, Frank ED, Gerdes S, Glass EM, Goesmann A, Hanson A, Iwata-Reuyl D, Jensen R, Jamshidi N, Krause L, Kubal M, Larsen N, Linke B, McHardy AC, Meyer F, Neuweger H, Olsen G, Olson R, Osterman A, Portnoy V, Pusch GD, Rodionov DA, Rückert C, Steiner J, Stevens R, Thiele I, Vassieva O, Ye Y, Zagnitko O, Vonstein V.** 2005. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. Nucleic Acids Res. **33**:5691–5702. http://dx.doi.org/10.1093/nar/gki866.

26. **Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M.** 2012. KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res. **40**:D109–D114. http://dx.doi.org/10.1093/nar/gkr988.

27. **Powell S, Szklarczyk D, Trachana K, Roth A, Kuhn M, Muller J, Arnold R, Rattei T, Letunic I, Doerks T, Jensen LJ, von Mering C, Bork P.** 2012. eggNOG v3.0: orthologous groups covering 1133 organisms at 41 different taxonomic ranges. Nucleic Acids Res. **40**:D284–D289. http://dx.doi.org/10.1093/nar/gkr1060.

28. **UniProt Consortium.** 2012. Reorganizing the protein space at the Universal Protein Resource (UniProt). Nucleic Acids Res. **40**:D71–D75. http://dx.doi.org/10.1093/nar/gkr981.

29. **Huang Y, Gilna P, Li W.** 2009. Identification of ribosomal RNA genes in metagenomic fragments. Bioinformatics **25**:1338–1340. http://dx.doi.org/10.1093/bioinformatics/btp161.

30. **Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO.** 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. **41**:D590–D596. http://dx.doi.org/10.1093/nar/gks1219.

31. **Pruesse E, Peplies J, Glöckner FO.** 2012. SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. Bioinformatics **28**:1823–1829. http://dx.doi.org/10.1093/bioinformatics/bts252.

32. **Strous M, Kraft B, Bisdorf R, Tegetmeyer HE.** 2012. The binning of metagenomic contigs for microbial physiology of mixed cultures. Front. Microbiol. **3**:410. http://dx.doi.org/10.3389/fmicb.2012.00410.

33. **Langille MGI, Brinkman FSL.** 2009. IslandViewer: an integrated interface for computational identification and visualization of genomic islands. Bioinformatics **25**:664–665. http://dx.doi.org/10.1093/bioinformatics/btp030.

34. **Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA.** 2009. Circos: an information aesthetic for comparative genomics. Genome Res. **19**:1639–1645. http://dx.doi.org/10.1101/gr.092759.109.

35. **Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R.** 2010. QIIME allows analysis of high-throughput community sequencing data. Nat. Methods **7**:335–336. http://dx.doi.org/10.1038/nmeth.f.303.

36. **Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Wagner H.** 2013. vegan: community ecology package. http://cran.r-project.org/web/packages/vegan/vegan.pdf.

37. **Parks DH, Beiko RG.** 2010. Identifying biologically relevant differences between metagenomic communities. Bioinformatics **26**:715–721. http://dx.doi.org/10.1093/bioinformatics/btq041.

38. **Rivals I, Personnaz Lo Taing L, Potier M-C.** 2007. Enrichment or depletion of a GO category within a class of genes: which test? Bioinformatics **23**:401–407. http://dx.doi.org/10.1093/bioinformatics/btl633.

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

40. **Newcombe RG.** 1998. Interval estimation for the difference between independent proportions: comparison of eleven methods. Stat. Med. **17**:873–890.

41. **Wickham H.** 2009. ggplot2: elegant graphics for data analysis. Springer, New York, NY.

42. **Brazelton WJ, Baross JA.** 2009. Abundant transposases encoded by the metagenome of a hydrothermal chimney biofilm. ISME J. **3**:1420–1424. http://dx.doi.org/10.1038/ismej.2009.79.

43. **Yilmaz S, Allgaier M, Hugenholtz P.** 2010. Multiple displacement amplification compromises quantitative analysis of metagenomes. Nat. Methods **7**:943–944. http://dx.doi.org/10.1038/nmeth1210-943.

44. **Kim KH, Bae JW.** 2011. Amplification methods bias metagenomic libraries of uncultured single-stranded and double-stranded DNA viruses. Appl. Environ. Microbiol. **77**:7663–7668. http://dx.doi.org/10.1128/AEM.00289-11.

45. **Kim KH, Chang HW, Nam YD, Roh SW, Kim MS, Sung Y, Jeon CO, Oh HM, Bae JW.** 2008. Amplification of uncultured single-stranded DNA viruses from rice paddy soil. Appl. Environ. Microbiol. **74**:5975–5985. http://dx.doi.org/10.1128/AEM.01275-08.

46. **Campbell BJ, Engel AS, Porter ML, Takai K.** 2006. The versatile ε-*proteobacteria*: key players in sulphidic habitats. Nat. Rev. Microbiol. **4**:458–468. http://dx.doi.org/10.1038/nrmicro1414.

47. **Miroshnichenko ML, Rainey FA, Rhode M, Bonch-Osmolovskaya EA.** 1999. *Hippea maritima* gen. nov., sp. nov., a new genus of thermophilic, sulfur-reducing bacterium from submarine hot vents. Int. J. Syst. Bacteriol. **49**:1033–1038. http://dx.doi.org/10.1099/00207713-49-3-1033.

48. **Mori K, Yamaguchi K, Sakiyama Y, Urabe T, Suzuki K.** 2009. *Caldi-

sericum exile gen. nov., sp nov., an anaerobic, thermophilic, filamentous bacterium of a novel bacterial phylum, *Caldiserica* phyl. nov., originally called the candidate phylum OP5, and description of *Caldisericaceae* fam. nov., *Caldisericales* ord. nov and *Caldisericia* classis nov. Int. J. Syst. Evol. Microbiol. **59:**2894–2898. http://dx.doi.org/10.1099/ijs.0.010033-0.

49. **Miroshnichenko ML, Bonch-Osmolovskaya EA.** 2006. Recent developments in the thermophilic microbiology of deep-sea hydrothermal vents. Extremophiles **10:**85–96. http://dx.doi.org/10.1007/s00792-005-0489-5.

50. **Reysenbach A-L, Liu Y, Banta AB, Beveridge T, Kirshtein JD, Schouten S, Tivey MK, Von Damm KL, Voytek MA.** 2006. A ubiquitous thermoacidophilic archaeon from deep-sea hydrothermal vents. Nature **442:**444–447. http://dx.doi.org/10.1038/nature04921.

51. **Perner M, Hansen M, Seifert R, Strauss H, Koschinsky A, Petersen S.** 2013. Linking geology, fluid chemistry and microbial activity of basalt- and ultramafic-hosted deep-sea hydrothermal vent environments. Geobiology **11:**340–355. http://dx.doi.org/10.1111/gbi.12039.

52. **Amend JP, McCollom TM, Hentscher M, Bach W.** 2011. Catabolic and anabolic energy for chemolithoautotrophs in deep-sea hydrothermal systems hosted in different rock types. Geochim. Cosmochim. Acta **75:**5736–5748. http://dx.doi.org/10.1016/j.gca.2011.07.041.

53. **Seewald J, Cruse AM, Saccocia PJ.** 2003. Aqueous volatiles in hydrothermal fluids from the Main Endeavour Vent Field, northern Juan de Fuca Ridge: temporal variability following earthquake activity. Earth Planet Sci. Lett. **216:**575–590. http://dx.doi.org/10.1016/S0012-821X(03)00543-0.

54. **Flemings PB, Behrmann JH, John CM, Expedition 308 Scientists.** 2006. Gulf of Mexico hydrogeology. Proceedings of the Integrated Ocean Drilling Program, vol 308. Integrated Ocean Drilling Program Management International, College Station, TX.

55. **Wang Y, Leung HCM, Yiu SM, Chin FYL.** 2012. MetaCluster 5.0: a two-round binning approach for metagenomic data for low-abundance species in a noisy sample. Bioinformatics **28:**i356–i362. http://dx.doi.org/10.1093/bioinformatics/bts397.

56. **Kelley DR, Salzberg SL.** 2010. Clustering metagenomic sequences with interpolated Markov models. BMC Bioinformatics **11:**544. http://dx.doi.org/10.1186/1471-2105-11-544.

57. **Perner M, Kuever J, Seifert R, Pape T, Koschinsky A, Schmidt K,** Strauss H, Imhoff JF. 2007. The influence of ultramafic rocks on microbial communities at the Logatchev hydrothermal field, located 15°N on the Mid-Atlantic Ridge. FEMS Microbiol. Ecol. **61:**97–109. http://dx.doi.org/10.1111/j.1574-6941.2007.00325.x.

58. **Perner M, Seifert R, Weber S, Koschinsky A, Schmidt K, Strauss H, Peters M, Haase K, Imhoff JF.** 2007. Microbial CO_2 fixation and sulfur cycling associated with low-temperature emissions at the Lilliput hydrothermal field, southern Mid-Atlantic Ridge (9°S). Environ. Microbiol. **9:**1186–1201. http://dx.doi.org/10.1111/j.1462-2920.2007.01241.x.

59. **Flores GE, Wagner ID, Liu Y, Reysenbach AL.** 2012. Distribution, abundance, and diversity patterns of the thermoacidophilic "deep-sea hydrothermal vent euryarchaeota 2." Front. Microbiol. **3:**47. http://dx.doi.org/10.3389/fmicb.2012.00047.

60. **Voordeckers JW, Starovoytov V, Vetriani C.** 2005. *Caminibacter mediatlanticus* sp. nov., a thermophilic, chemolithoautotrophic, nitrate-ammonifying bacterium isolated from a deep-sea hydrothermal vent on the Mid-Atlantic Ridge. Int. J. Syst. Evol. Microbiol. **55:**773–779. http://dx.doi.org/10.1099/ijs.0.63430-0.

61. **Weon HY, Kim BY, Yoo SH, Lee SY, Kwon SW, Go SJ, Stackebrandt E.** 2006. *Niastella koreensis* gen. nov., sp. nov. and *Niastella yeongjuensis* sp. nov., novel members of the phylum Bacteroidetes, isolated from soil cultivated with Korean ginseng. Int. J. Syst. Evol. Microbiol. **56:**1777–1782. http://dx.doi.org/10.1099/ijs.0.64242-0.

62. **Mori K, Sunamura M, Yanagawa K, Ishibashi J, Miyoshi Y, Iino T, Suzuki K, Urabe T.** 2008. First cultivation and ecological investigation of a bacterium affiliated with the candidate phylum OP5 from hot springs. Appl. Environ. Microbiol. **74:**6223–6229. http://dx.doi.org/10.1128/AEM.01351-08.

63. **Schrenk MO, Kelley DS, Delaney JR, Baross JA.** 2003. Incidence and diversity of microorganisms within the walls of an active deep-sea sulfide chimney. Appl. Environ. Microbiol. **69:**3580–3592. http://dx.doi.org/10.1128/AEM.69.6.3580-3592.2003.

64. **Peters M, Strauss H, Farquhar J, Ockert C, Eickmann B, Jost CL.** 2010. Sulfur cycling at the Mid-Atlantic Ridge: a multiple sulfur isotope approach. Chem. Geol. **269:**180–196. http://dx.doi.org/10.1016/j.chemgeo.2009.09.016.