Biogeography and ecology of the rare and abundant microbial lineages in deep-sea hydrothermal vents

Rika E. Anderson[1,2], Mitchell L. Sogin[3], and John A. Baross[1]

[1]School of Oceanography and Astrobiology Program, University of Washington, Seattle, WA
[2]Institute for Genomic Biology, University of Illinois at Urbana-Champaign, IL
[3]Josephine Bay Paul Center, Marine Biological Laboratory, Woods Hole, MA

Corresponding author:
Rika Anderson
Institute for Genomic Biology
University of Illinois at Urbana-Champaign
1206 West Gregory Drive
MC-195
Urbana, IL 61801
(206) 353-2019
rikander@illinois.edu

Running title: Rare and abundant lineages in deep-sea hydrothermal vents

Abstract

Environmental gradients generate countless ecological niches in deep-sea hydrothermal vent systems, which foster diverse microbial communities. The majority of distinct microbial lineages in these communities occur in very low abundance. However, the ecological role and distribution of rare and abundant lineages, particularly in deep, hot subsurface environments, remains unclear. Here, we use 16S rRNA tag sequencing to describe biogeographic patterning and microbial community structure of both rare and abundant archaea and bacteria in hydrothermal vent systems. We show that while rare archaeal lineages and almost all bacterial lineages displayed geographically restricted community structuring patterns, the abundant lineages of archaeal communities displayed a much more cosmopolitan distribution. Finally, analysis of one high-volume, high-temperature fluid sample representative of the deep hot biosphere described a unique microbial community that differed from microbial populations in diffuse flow fluid or sulfide samples, yet the rare thermophilic archaeal groups showed similarities to those that occur in sulfides. These results suggest that while most archaeal and bacterial lineages in vents are rare and display a highly regional distribution, a small percentage of lineages, particularly within the archaeal domain, are successful at widespread dispersal and colonization.

Introduction

In almost all ecosystems investigated to date, a minority of microbial lineages dominates the community, while many low-abundance lineages account for most of the total community diversity. These rare lineages make up the archaeal and bacterial “rare

biosphere," a term first coined by Sogin *et al.* (2006). Despite the ubiquity of rare lineages, the ecological role and distribution of the rare biosphere in a given community remains unclear.

Analysis of the spatial distribution of rare lineages can shed light on their ecological role. Generally speaking, microbial spatial distribution can be controlled by environmental selection, historical events such as dispersal limitation, or both (Martiny et al. 2006). A commonly discussed hypothesis is that high population densities of microorganisms facilitate rapid and widespread dispersal, but that contemporary environments select for particular microbial assemblages—essentially, that "everything is everywhere, but the environment selects" (Baas Becking 1934). It has been argued that if microorganisms are freely dispersed, rare strains should be fairly cosmopolitan (Pedrós-Alió 2006). In a study illustrating the enormous dispersal capabilities of microbes, Gibbons *et al.* (2013) found that most microbial lineages identified in the International Census of Marine Microbes dataset could be found at a single deeply-sequenced site in the Western English Channel. However, a global analysis of bacterial distribution revealed that microbial lineages had a bipolar distribution, being confined by hemisphere more than expected under a null model (Sul *et al.* 2013), suggesting that marine microorganisms are subject to a certain degree of dispersal limitation.

Moreover, studies examining the spatial distributions of rare versus abundant strains have suggested that rare strains are not cosmopolitan, but are subject to environmental and ecological pressures similar to those of the abundant strains. A study of rare and abundant strains in the Arctic Ocean found that rare operational taxonomic units (OTUs) exhibited similar geographic patterns to those of the abundant OTUs

(Galand et al. 2009), demonstrating that rare strains in the Arctic Ocean are not widely dispersed. Similarly, Vergin *et al.* (2013) showed that both rare and abundant strains are subject to spatiotemporal patterns such as seasonality and stratification.

With diverse but isolated habitats and a global distribution, hydrothermal vent systems present a compelling test case for these questions regarding the distributions of rare and abundant strains. In these systems, a wide variety of ecological niches develop by mixing of high-temperature, reduced, metal-enriched hydrothermal fluid with cooler, oxidized seawater both above and below the seafloor. The high-temperature centers of structures are inhabited by microbial communities that tend to be much more archaeal-dominated than the diffuse flow communities that result from the mixing of high-temperature fluid and background seawater (Schrenk *et al.* 2003; Takai & Horikoshi 1999; Takai *et al.* 2001; Slobodkin *et al.* 2001; Kelley *et al.* 2002). Samples of fluids from diffuse flow vents reveal richly diverse microbial communities with members that range from deep subsurface hyperthermophiles to mesophiles and psychrophiles entrained from deep seawater (Huber *et al.* 2003, 2002, 2007; Deming & Baross 1993). Moreover, sampling of diffuse flow fluids suggests that a portion of the microbial community found in these fluids draws from a deep subsurface habitat hosting thermophilic, anaerobic archaeal and bacterial communities; effectively, hydrothermal systems provide a “window” to the deep biosphere (Summit & Baross 2001; Deming & Baross 1993). Deep sequencing of archaea and bacteria from seamount diffuse fluids revealed thousands of bacterial and archaeal lineages, the majority of which occur in very low abundances (Huber *et al.* 2007; Sogin *et al.* 2006). Some of these rare lineages may be abundant in other niches of the vent system.

97 The question of geographic isolation versus geochemical selection in
98 hydrothermal systems has been investigated previously, but without consistent trends.
99 While previous work at hydrothermal systems has indicated that geographically distinct
100 vents generally host phylogenetically distinct populations (Holden *et al.* 2001; Huber *et*
101 *al.* 2006; Flores *et al.* 2012; Opatkiewicz *et al.* 2009), only some studies have seen clear
102 correlations between chemistry and community composition (Huber et al. 2006; Huber et
103 al. 2003). Other studies have found evidence of geographic isolation, but with little
104 correlation to fluid chemistry (Opatkiewicz et al. 2009; Huber et al. 2010). Here, we seek
105 to address two primary questions regarding the distribution of rare and abundant lineages
106 in hydrothermal systems. First, across large spatial scales, do rare and abundant strains
107 exhibit the same patterns of spatial structuring, or do we see differences that might point
108 to different ecological roles? While some have suggested that the rare biosphere
109 encompasses a persistent, cosmopolitan seed bank throughout the ocean (Pedrós-Alió
110 2006), longstanding theories of spatial ecology suggest that rare species should be fairly
111 restricted, while more abundant species have a more widespread distribution (Brown
112 1984). We aim to determine which pattern prevails in hydrothermal systems. Second,
113 since hydrothermal systems encompass many different habitat types linked by fluid flux,
114 we seek to determine whether strains that are rare in certain habitats of hydrothermal
115 systems are abundant in others.
116 We investigated all publicly available DNA sequences of the v6 region of 16S
117 rRNA gene (pyrotags) from hydrothermal vent samples in the VAMPS database, which
118 includes samples from diffuse flow fluids and vent sulfides worldwide, and combined
119 this evaluation with analysis of a single, large-volume high-temperature sample, sampled

from Hulk vent on the Juan de Fuca Ridge, that provides the best representation currently available of the deep hot subsurface biosphere in vent systems. In this way, we sampled multiple hydrothermal vent niches across a wide spatial sample area. We show that while most archaeal and bacterial strains were observed only in particular hydrothermal regions due to either restricted dispersal or ecological selection, certain archaeal strains were widespread and therefore appear to be able to disperse and colonize new niches efficiently. Moreover, analysis of the high-temperature fluid sample in combination with other sample types indicates that strains that are rare in some niches are abundant in others.

Materials and Methods

Sample site description and sampling procedures

We obtained all 16S v6 pyrotag datasets used in this study, with the exception of those newly acquired from a single high-temperature sample, from the publicly available VAMPS database (www.vamps.mbl.edu). Table S1 presents a list of all samples and associated metadata; Figure S1 depicts a map of the sample sites.

We collected the high-temperature fluid sample at Hulk vent in August 2009 aboard the *R/V Atlantis.* Hulk is a large sulfide chimney located at 47˚ 57.00’ N, 129˚ 5.81’ W on the Main Endeavour Field on the Juan de Fuca Ridge, a spreading center located about 200 miles from the coast of Washington and Oregon (Figure S1). We deployed a custom-built barrel sampler using *DSV Alvin* to collect 170 L of high-temperature diffuse flow fluid from the base of the sulfide structure. The average temperature of the sample, as determined from its silica and magnesium concentrations, was approximately 125˚C (Anderson *et al.* 2011). This sample most likely represents a

143 wide range of niches resulting from mixtures of fluid at various temperatures, since the
144 sample funnel was placed on top of a colony of tube worms at approximately 20˚C, and
145 most likely sampled fluid from a nearby conduit measured to be about 300˚C. On deck,
146 we put samples on ice prior to filtering through three 0.02 μm Steripaks (Millipore,
147 USA). DNA extraction procedures are described in detail in Anderson *et al.* (2013).
148 Briefly, we freeze-thawed one of the Steripaks three times, then added DNA extraction
149 buffer (0.1M Tris-HCl, 0.2M Na-EDTA, 0.1M NaH_2PO_4, 1.5M NaCl, and 1%
150 cetyltrimethylammonium bromide), 50 mg ml^{-1} lysozyme, 1% proteinase K, and 20%
151 SDS solution to the filter. We extracted DNA using phenol:chloroform:isoamyl alcohol
152 and chloroform:isoamyl alcohol.
153 Other diffuse flow samples obtained from the VAMPS database were collected by
154 J. Huber from eight different seamounts at Axial Seamount, the Mariana Arc, and Loihi
155 Seamount, depicted in Figure S1. Axial Seamount is an active volcano located on the
156 Juan de Fuca Ridge about 300 miles off the coast of Oregon. The caldera is about 700 m
157 above the level of the ridge, and is bordered on three sides by a boundary fault. Several
158 areas of active venting are located within the caldera. Other diffuse flow fluid samples
159 derive from several seamounts along the Mariana Arc, located in the Western Pacific
160 Ocean from about 12 to 24˚N. These seamounts, including NW Eifuku, Daikoku, Nikko,
161 NW Rota, and E Diamante, were located along the active front of the Mariana Arc, with
162 the exception of Forecast, which may have greater influence from the backarc spreading
163 axis (Huber *et al.* 2010; Embley *et al.* 2004). Analysis of *Epsilonproteobacteria* in these
164 samples was described in Huber *et al.* 2010. Loihi Seamount is an active submarine

volcano located above the Hawaiian hotspot. Sampling and DNA extraction methods for these diffuse flow samples were discussed in Huber *et al.* (2010).

All sulfide samples collected from the VAMPS database were collected by A.-L. Reysenbach from the Lau Basin, which is a back-arc basin formed by the subduction of the Pacific plate below the Australian plate. Sulfide samples were collected and processed as described in Flores *et al.* 2011 and 2012. Samples were sequenced as part of the International Census of Marine Microbes (ICoMM) initiative.

Sequencing

We constructed and sequenced v4–v6 and v6 amplicon libraries for the Hulk sample at the Josephine Bay Paul Center at the Marine Biological Laboratory on a Roche 454 GSFLX Titanium platform using the techniques described in Huber *et al.* (2007) and Sogin *et al.* (2006). All sequences are publicly available on the VAMPS website (http://vamps.mbl.edu) under dataset names REA_HDF_Av6v4, REA_HDF_Bv6v4 and REA_HDF_Bv6 for the Hulk sample. We used the v6 bacterial dataset for this study to enable direct comparison between v6 datasets, and trimmed the archaeal v6v4 datasets when comparing against other v6 samples. We compared this sample with sequences from projects ICM_ALR_Av6, ICM_ALR_Bv6, KCK_SMT_Av6, and KCK_SMT_Bv6, which we obtained from the VAMPS database.

Bioinformatic analysis

We trimmed and quality filtered all reads through the VAMPS pipeline using the quality control parameters outlined in Huse et al. (2007). We performed taxonomic

188 analyses of each sample using the GAST process in VAMPS (Huse *et al.* 2008). We
189 further screened, filtered and trimmed all sequences as a batch set using mothur (Schloss
190 *et al.* 2009). Trimming of sequences removed the v4–v5 region of the Hulk sequences,
191 leaving behind only the v6 region to facilitate direct comparison. We aligned both
192 archaeal and bacterial sequences against the SILVA database (Quast *et al.* 2013) through
193 the mothur pipeline. We clustered sequences into operational taxonomic units (OTUs)
194 using average-neighbor hierarchical clustering to the 0.03 level using mothur, again
195 treating all sequences from all samples as a batch set for OTU clustering. We tested
196 different clustering cutoffs, including unique, 0.2, and 0.4 distance cutoffs, and observed
197 no qualitative differences in our results. We calculated diversity indices (rarefaction
198 curves and Shannon and Simpson evenness) using mothur. For comparison between
199 samples, we constructed distance matrices in mothur using the Bray-Curtis calculator of
200 community membership and structure. We also assessed community structure with
201 distance matrices calculated using the Unifrac method after creating a phylogenetic tree
202 for all sequences in each sample using clearcut within the mothur package (Schloss et al.
203 2009). To normalize between different sample sizes, we randomly subsampled the data
204 1000 times (to the size of the smallest sample) when comparing between datasets. We
205 generated cluster dendrograms from these distance matrices using PRIMERv6 (Clarke &
206 Gorley 2006).
207 For the rare vs. abundant OTU analysis, we separated sequences from OTUs that
208 were considered to be abundant in each sample (representing equal to or greater than 1%
209 of all sequences in the sample) from those considered to be rare (representing equal to or
210 less than 0.1% of all sequences in the sample). We carried out analysis of similarity

(ANOSIM) tests using PRIMERv6 to determine whether there were assemblage differences between groups of samples specified according to geographic location. We conducted nine hundred ninety nine permutations of the test for each ANOSIM analysis, using a resemblance matrix of Bray-Curtis dissimilarity as determined in mothur. To test for community similarity distance decay, we conducted Mantel tests of mothur-generated Bray-Curtis community dissimilarity matrices against distance matrices of geographic distance using the "fossil" library with the statistical software package R (R Core Team, 2013). We used R to generate heatmaps with OTU relative abundance data generated in mothur.

To create phylogenetic trees of sequences from the *Thermococcales* and *Methanococcales*, we identified all OTUs belonging to either of these groups according to the SILVA taxonomic classification conducted in mothur, and selected a reference sequence from each OTU. We created reference data sets with full-length 16S sequences from the SILVA database (Quast *et al.* 2013); we aligned both the reference sequences and sample sequences in the SILVA aligner (Pruesse *et al.* 2012). We created a base tree from the reference sequences in RAxML (Stamatakis 2006) using a rapid bootstrap analysis to search for the best maximum likelihood tree with 100 alternative runs on distinct starting trees. We used EPA (Berger *et al.* 2011) within the RAxML package to insert the short v6 sample sequences into the base tree. For tree construction, we used the GTR+ optimization of substitution rates and the GAMMA model of rate heterogeneity.

We conducted comparisons between v4–v6 sequences in the Hulk sample and uncultured crenarchaeal sequences with USEARCH v6 (Edgar 2010) using the usearch_global command.

Results

Comparative community structure of diffuse flow and sulfide structures

Taxonomic classification of bacterial v6 sequences for all samples revealed high abundances of both the *Epsilonproteobacteria* and *Gammaproteobacteria* groups in most samples for both diffuse flow and sulfide samples (Figure 1A). Among the archaea, sulfide samples exhibited high abundances of *Archaeoglobi* and *Halobacteria*, whereas Marine Group I and *Thermoplasmata* dominated most diffuse flow samples (Figure 1B). The high-temperature Hulk sample, in contrast, was unique in its high abundance of *Thermococcales*.

OTUs at a 3% distance produced a total of 3711 OTUs in the archaea and 22029 OTUs in the bacteria for all samples. In almost all samples, bacterial communities had higher richness than the archaeal communities, as depicted in the rarefaction curves for samples from both domains (Figure S2). None of these rarefaction curves reached an asymptote, indicating that none of the datasets captured the total diversity of the sample. Simpson and Shannon evenness indices showed that bacterial communities had higher evenness than archaeal communities (Table S2), which was significant for the Simpson evenness test (t-test, p=7.8E-18).

Clustering of samples according to community similarity

Cluster dendrograms indicated the degree of similarity between samples based on the relative abundance of OTUs. For the bacteria, some clustering according to seamount was apparent among the diffuse flow samples (Figure 2A). For the archaea, there was

much higher similarity between diffuse flow samples and between sulfide samples than for the bacteria (Figure 3A). Archaeal communities in diffuse flow samples were less likely to cluster by location. In most cases, the sulfide structures clustered separately from the diffuse flow samples. The high-temperature Hulk sample grouped with the other diffuse flow samples, though at very low similarity.

We used ANOSIM analyses to test the hypothesis that archaeal and bacterial samples clustered according to geographic location. We grouped diffuse flow samples according to seamount, while sulfide samples, which were all collected in the Lau Basin, were grouped together. We grouped the high temperature Hulk sample separately from other samples. Location designations are shown in the legends of Figures 2 and 3. To test whether clustering of OTUs by location was significant, we conducted ANOSIM analyses for bacteria and archaea. ANOSIM analysis indicated that clustering according to geographic location was significant for the bacteria ($p \leq 0.1\%$), but not for the archaea (Table 1). Comparison of samples using Unifrac, a phylogenetic rather than taxonomic-based metric of beta diversity, provided similar results (Figures S3, S4; Table S3). To test the hypothesis that OTUs are dispersal-limited, we conducted Mantel tests of the correlation between OTU dissimilarity and geographic distance. Our analyses found a significant positive correlation (Table 1).

Biogeography and distribution of rare and abundant OTUs in hydrothermal systems

Analyses of all OTUs together cannot identify differences in ecological patterning between the rare and abundant OTUs. Therefore, we separated the rare and abundant OTUs within each sample to determine whether they exhibit different biogeographic and

community structuring patterns. For this analysis, we considered rare OTUs to be those OTUs representing less than or equal to 0.1% of the sequences in the sample; abundant OTUs were considered to be those OTUs representing greater than or equal to 1% of all the sequences in the sample. This scoring follows definitions of rare and abundant groups previously established by Pedros-Alió (2006) and Fuhrman (2009). This scoring leaves out ambiguous sequences between 0.1-1% abundance, which cannot be not easily defined as either "rare" or "abundant." We also tested different percentage cutoffs for the definition of "rare" and "abundant," testing a range between 0.5 to 5% for abundant strains, and 0.05 to 0.3% for rare strains. We repeated the analyses described below and the patterns were consistent, indicating that the trends described here are not sensitive to strict cutoffs.

The taxonomic identification of rare and abundant OTUs did not differ drastically from each other, though certain taxonomic groups had a greater tendency to contain abundant OTUs, specifically *Gammaproteobacteria* and *Epsilonproteobacteria* for bacteria and *Thermoplasmata* and Marine Group I for archaea (Figures S5 and S6).

We identified the rare and abundant OTUs in each sample and clustered the samples according to community similarity as before to determine whether the rare and abundant OTUs exhibited similar patterns. In the bacteria, community structuring among locations was similar, though not identical, between the rare and the abundant OTUs (Figure 2B, C). Overall, rare OTUs showed more dissimilarity from sample to sample (averaging about 80% dissimilarity) compared to abundant OTUs (averaging 60–70% dissimilarity). ANOSIM tests of bacterial clustering indicated that clustering according to seamount was significant in all cases (Table 1). Mantel tests indicated a significant

303 positive correlation between community dissimilarity and geographic distance for all
304 cases (Table 1). Clustering with the Unifrac method again yielded similar results;
305 however, ANOSIM analysis of geographic clustering for abundant bacterial strains
306 without sulfides was not significant (Table S3) and abundant bacterial strains showed
307 greater similarity among samples than among rare bacterial strains (Figure S3B, C).
308 For archaeal lineages, different patterns appeared. As with bacteria, abundant
309 OTUs showed much higher similarity between samples than rare OTUs. Separating the
310 rare and abundant OTUs revealed that the lack of biogeographic patterning for all archaea
311 was driven almost entirely by the abundant OTUs (Figure 3B, C). In contrast to the
312 bacterial case, the only archaeal OTUs to cluster significantly by geographic location
313 were the rare OTUs (Table 1). An extremely low R statistic and high p-value from the
314 ANOSIM analyses indicated that abundant archaeal OTUs showed almost no tendency to
315 group according to geographic location, especially when considering only diffuse flow
316 samples. Mantel tests of community dissimilarity versus geographic distance showed the
317 same pattern (Table 1). Unifrac analyses indicated similar trends, though all archaea
318 together demonstrated statistically significant biogeographic clustering (Table S3, Figure
319 S4).
320 Our analyses indicated that 69% of the bacterial OTUs and 66% of archaeal
321 OTUs were found only in one sample, most of which were rare lineages. Nevertheless,
322 removal of OTU singletons yielded the same statistical results. While a more
323 comprehensive sampling effort might have detected these rare lineages in a greater
324 number of samples, the community structure of rare OTUs was significantly consistent
325 within each region (Table 1). Random sampling from the environment in this way would

not yield a statistically significant biogeographic pattern of detection if the rare biosphere were truly cosmopolitan (Pedrós-Alió 2006; Galand et al. 2009), and indicates different patterns of spatial distributions between rare and abundant OTUs among the archaea, but not among the bacteria.

Persistence of OTUs across samples

As suggested in Figure 3, the widely dispersed archaeal OTUs tended to be those that were more abundant within samples. To determine whether archaeal strains shifted in relative abundance from site to site, we visualized OTU abundance in the heatmap depicted in in Figure 4, depicting the relative abundance of archaeal OTUs from the high-temperature Hulk sample across all other samples analyzed here. Generally, abundant OTUs in Hulk were more likely to be abundant or at least present in other samples, while rare OTUs were more likely to be rare or undetected across samples. OTU 3147, for example, a member of the Marine Group I crenarchaea, was abundant in almost all diffuse flow samples. OTUs that were abundant in diffuse flow were absent or rare in sulfide samples, and vice versa. An exception to this trend was the most abundant archaeal OTU in the Hulk sample, OTU 3645, a *Thermococcus* sequence that comprised 63% of the sample. This OTU was rare in most other samples examined. However, this OTU reached a maximum of 9% in one sulfide sample (sulf_20), from an entirely different ecological niche.

Phylogenetics of the Thermococcales and Methanococcales

Clustering samples into OTUs does not give an indication of phylogenetic relatedness between sequences and across samples, yet understanding phylogenetic relatedness can provide insight into similarity and gene flow between samples. To determine the extent to which rare strains bloomed or dispersed between niches within hydrothermal systems, we focused on the phylogenetics of specific strains. *Thermococcales* are a general indicator of high-temperature fluids and were dominant in the Hulk sample. Thus we created a phylogenetic tree of sequences falling within the order *Thermococcales* to investigate relationships between samples that might be based on temperature, especially the high-temperature Hulk sample and the sulfide samples (Figure 5). Both the sulfide samples and the Hulk sample had high diversity within the *Thermococcales* order, with several OTUs falling into many different clades in the tree. It was much more common for sulfide and Hulk sequences to group together into the same OTU or branch (at 3% distance) than it was for diffuse flow fluid sequences to group with sequences from sulfides or Hulk. Fewer *Thermococcales* OTUs were found in diffuse flow samples overall; those that were present tended to cluster into a few clades on the tree, particularly within the *Palaeococcus* genus, or to group on branches with no cultured representatives.

Similar results were found in a phylogenetic tree of OTUs falling in the *Methanococcales* order, though these OTUs were found with greater frequency in diffuse flow samples (Figure S7). The separation between sulfide and diffuse flow OTUs was more distinct in this tree. The OTU from the Hulk sample fell within a clade shared with other sulfide OTUs, despite being geographically closer to Axial Volcano, where most of the *Methanococcales* OTUs were found.

Given the phylogenetic similarities between sulfide samples and the high-temperature sample, we also sought to determine whether abundant sequences in sulfides matched rare sequences in the Hulk high-temperature sample. Because previous work has indicated that uncultured crenarchaea dominate the interior of sulfide structures (Schrenk *et al.* 2003), we conducted global sequence comparisons of the Hulk high-temperature v4–v6 region against a database of uncultured crenarchaea identified from sulfide clone libraries. Two sequences were found that matched previously identified crenarchaea in sulfides at 99–100% identity: a *Desulfurococcales* lineage from a white smoker spire on the East Pacific Rise (Kormas *et al.* 2006), and a *Pyrodictium* lineage identified in an in-situ growth chamber deployed within a sulfide structure (Nercessian *et al.* 2003).

Discussion

Domain differences in the ecology of rare and abundant lineages of the bacteria and archaea

Null ecological models predict a positive correlation between abundance and distribution: that is, the more highly abundant species will be found in more locations (Brown 1984). With abundance aggregated among many OTUs, this leaves many rare and few abundant. The abundance of a given OTU is distributed across space, leaving many sites without rare taxa, and many sites where abundant taxa are present. Global studies of microbial biogeography have confirmed this trend (Nemergut et al. 2011). On the large scale, our results here confirm that abundant strains tend to be more widespread in vent systems. Indeed, for both the archaeal and bacterial domains, there was a greater tendency for community structuring to be similar across sites for the abundant strains.

However, in contrast to previous studies, our results indicate a striking difference in ecological patterning between the archaeal and bacterial domains, and point to a widespread, global range for specific abundant archaeal phylotypes.

For bacteria, microbial community structuring according to geographic location applied to both the rare and abundant lineages. In all cases, bacterial community structure was significantly similar within seamounts but distinct from vents farther afield. The rare archaeal strains displayed a similar pattern. While we cannot distinctively determine whether this biogeographic pattern can be attributed to environmental selection or historical causes, the distance effect observed through the Mantel tests suggests that dispersal limitation plays a strong role in forming the biogeographic patterns observed for bacterial and rare archaeal strains. Moreover, since location was found to be a stronger indicator of community similarity than chemistry in previous work (Opatkiewicz et al. 2009; Huber et al. 2010), dispersal limitation or other historical causes may be a stronger driver of community structuring than environmental selection in hydrothermal systems.

However, a key difference from previous studies indicating community structuring by vent field is the finding that many abundant archaeal strains were observed at high abundance at sites spread across the globe. The results suggest an ecological pattern in which the majority of archaeal OTUs are rare and biogeographically restricted, but a few abundant archaeal OTUs dominate and are widespread. Many of the most abundant and widespread archaeal OTUs in diffuse fluids belonged to Marine Groups I and II, which are native to deep seawater and therefore might more easily travel in ocean currents from one vent system to the next. These particular lineages of Marine Groups I

and II may have been ecotypes that gained a fitness advantage through some means, such as horizontal gene transfer, that allowed them to proliferate rapidly in vent habitats.

However, it is unclear why certain abundant archaeal OTUs appear to be so widely dispersed, while abundant bacterial OTUs are more biogeographically restricted. The perceived ubiquity of certain archaeal lineages may be linked to the observation that archaea generally exhibit lower richness compared to bacteria in various environments globally (Aller & Kemp 2008). Low archaeal diversity may result from slow evolution of the 16S gene in archaea relative to the bacterial 16S gene. It is also possible that current primer designs do not adequately detect the true extent of archaeal diversity. In this case, the taxonomic resolution used here cannot detect certain biogeographic patterns, a potential problem that has been discussed previously (Hanson *et al.* 2012) and may be rectified with new primers that target different regions of the 16S gene. Moreover, a likely contributing factor is that 16S-based studies do not reflect the diversity encoded on the rest of the archaeal genome. Previous work by Holden *et al.* (2001) has indicated that *Thermococcales* isolates exhibit genetic and phenotypic diversity that correlates with geographic location and habitat, yet is not reflected in the highly conserved 16S gene. While a certain 16S sequence may be ubiquitous, individual phenotypes within that OTU may be confined to particular regions. Extensive intra-species genetic diversity through gene transfer and recombination has been observed in natural archaeal populations (Allen *et al.* 2007), and physiological variation within the genome has been observed in Marine Group II, for example, which encodes proteorhodopsins in the photic zone, but not deeper in the water column (Frigaard *et al.* 2006). The archaeal pangenome can therefore be quite extensive. This distinction may pertain especially to thermophilic archaea, given

that high rates of horizontal gene transfer have been observed among thermophiles (Koonin *et al.* 2001; Beiko *et al.* 2005). Thus, while the same OTUs were observed across multiple vent sites, it is possible that there was a range of physiological variation within those OTUs from one site to the next that was not reflected by the 16S v6 sequence. Further research involving comparisons of full genome sequences will provide insight into this possibility.

Community structuring within hydrothermal niches and the deep subsurface

The gradients within deep-sea hydrothermal systems, created by the mixing of hydrothermal fluid and deep seawater, establish multiple ecological niches that foster high microbial diversity. Specific examination of the Hulk high-temperature sample provides further insight into the dynamics of rare and abundant strains within vent environments, as well as the extent of their dispersal between ecological niches. While the community structure of the Hulk high-temperature sample was generally more similar to diffuse fluid samples than to sulfide samples, specific OTUs were shared among the sulfide samples and the Hulk sample, particularly among the rare archaeal communities. This included the thermophilic archaeal groups *Thermococcales* and *Methanococcales,* as well as uncultured crenarchaea. Microbial communities in sulfides are exposed to more focused hydrothermal fluid flow, and thus higher temperatures, than microbial communities in diffuse flow. Therefore the similarities in archaeal groups between the high-temperature fluid sample and the sulfides most likely reflect selection according to temperature, and may also point to a common source for these archaeal lineages, in which thermophilic archaea in the hot subsurface are flushed into sulfide structures or fluids

462 emerging at the seafloor. This scenario depicts the microbial community in the deep hot
463 subsurface as occupying a unique niche, distinct from that of the diffuse flow and sulfide
464 microbial communities found downstream in the fluid flow, but one which we can
465 glimpse through the rare biosphere.

466 Taken together, these results provide a picture of microbial colonization and
467 dispersal both within and between vent systems. Within vents, we observe the abundant
468 strains of the deep, hot biosphere emerging as rare strains in sulfide structures, possibly
469 pointing to a deep, hot habitat for specific rare strains in sulfide structures. However, the
470 strains that were most successful at widespread dispersal and colonization appeared to be
471 specific archaeal strains that were not native to the deep biosphere but found mostly in
472 diffuse flow fluids, and appeared to be more capable of traveling deep-sea currents to
473 colonize new vent systems. Future work can reveal the extent of genome heterogeneity
474 within OTU groupings, and may reveal why certain archaeal strains appeared to be such
475 successful dispersers.

476

477

478 **Acknowledgements**

479 The authors gratefully acknowledge Chief Scientist Jim Holden, David Butterfield, and
480 the captain and crew of the R/V *Atlantis* for help with sample collection at sea. We would
481 like to thank W. Brazelton for stimulating discussions that informed our analyses, and we
482 thank J. Deming for edits that greatly improved the manuscript. J. Huber and A.-L.
483 Reysenbach kindly provided background information regarding samples used in this

484 analysis. Funding for this work was provided by a grant from the NASA Astrobiology
485 Institute to the Carnegie Institution of Washington.

References

Allen EE, Tyson GW, Whitaker RJ, Detter JC, Richardson PM & Banfield JF (2007) Genome dynamics in a natural archaeal population. *Proc. Natl. Acad. Sci. U. S. A.* 104: 1883–1888.

Aller JY & Kemp PF (2008) Are Archaea inherently less diverse than Bacteria in the same environments? *FEMS Microbiol. Ecol.* 65: 74–87.

Anderson RE, Beltrán MT, Hallam SJ & Baross JA (2013) Microbial community structure across fluid gradients in the Juan de Fuca Ridge hydrothermal system. *FEMS Microbiol. Ecol.* 83: 324–339.

Anderson RE, Brazelton WJ & Baross JA (2011) Using CRISPRs as a metagenomic tool to identify microbial hosts of a diffuse flow hydrothermal vent viral assemblage. *FEMS Microbiol. Ecol.* 77: 120–133.

Baas Becking L (1934) *Geobiologie of Inleiding Tot de Milieukunde [Geobiology or Introduction to the Science of the Environment]*. W.P. Van Stockum & Zoon, The Hague, Netherlands.

Beiko RG, Harlow TJ & Ragan M a (2005) Highways of gene sharing in prokaryotes. *Proc. Natl. Acad. Sci. U. S. A.* 102: 14332–14337.

Berger SA, Krompass D & Stamatakis A (2011) Performance, accuracy, and Web server for evolutionary placement of short sequence reads under maximum likelihood. *Syst. Biol.* 60: 291–302.

Brown JH (1984) On the Relationship between Abundance and Distribution of Species. *Am. Nat.* 124: 255–279.

Clarke KR & Gorley RN (2006) PRIMER v6: User Manual/tutorial. Primer-E Ltd Plymouth. *text*.

Deming J & Baross J (1993) Deep-sea smokers: Windows to a subsurface biosphere? *Geochim. Cosmochim. Acta* 57: 3219–3230.

Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26: 2460–2461.

Embley RW, Baker ET, Chadwick WW, Lupton JE, Resing JA, Massoth GJ & Nakamura K (2004) Explorations of Mariana Arc volcanoes reveal new hydrothermal systems. *Eos, Trans. Am. Geophys. Union* 85: 37.

Flores GE et al. (2011) Microbial community structure of hydrothermal deposits from geochemically different vent fields along the Mid-Atlantic Ridge. *Environ. Microbiol.* 13: 2158–2171.

Flores GE, Shakya M, Meneghin J, Yang ZK, Seewald JS, Wheat GC, Podar M & Reysenbach A-L (2012) Inter-field variability in the microbial communities of hydrothermal vent deposits from a back-arc basin. *Geobiology* 10: 333–346.

Frigaard N-U, Martinez A, Mincer TJ & DeLong EF (2006) Proteorhodopsin lateral gene transfer between marine planktonic Bacteria and Archaea. *Nature* 439: 847–850.

Fuhrman JA (2009) Microbial community structure and its functional implications. *Nature* 459: 193–199.

Galand PE, Casamayor EO, Kirchman DL & Lovejoy C (2009) Ecology of the rare microbial biosphere of the Arctic Ocean. *Proc. Natl. Acad. Sci. U. S. A.* 106: 22427–22432.

Gibbons SM, Caporaso JG, Pirrung M, Field D, Knight R & Gilbert JA (2013) Evidence for a persistent microbial seed bank throughout the global ocean. *Proc. Natl. Acad. Sci. U. S. A.* 110: 4651–4655.

Hanson CA, Fuhrman JA, Horner-Devine MC & Martiny JBH (2012) Beyond biogeographic patterns: processes shaping the microbial landscape. *Nat. Rev. Microbiol.* 10: 497–506.

Holden JF, Takai K, Summit M, Bolton S, Zyskowski J & Baross JA (2001) Diversity among three novel groups of hyperthermophilic deep-sea Thermococcus species from three sites in the northeastern Pacific Ocean. *FEMS Microbiol. Ecol.* 36: 51–60.

Huber J a, Johnson HP, Butterfield D a & Baross J a (2006) Microbial life in ridge flank crustal fluids. *Environ. Microbiol.* 8: 88–99.

Huber JA, Butterfield DA & Baross JA (2003) Bacterial diversity in a subseafloor habitat following a deep-sea volcanic eruption. *FEMS Microbiol. Ecol.* 43: 393–409.

Huber JA, Butterfield DA & Baross JA (2006) Diversity and distribution of subseafloor Thermococcales populations in diffuse hydrothermal vents at an active deep-sea volcano in the northeast Pacific Ocean. *J. Geophys. Res.* 111: G04016.

Huber JA, Butterfield DA & Baross JA (2002) Temporal changes in archaeal diversity and chemistry in a mid-ocean ridge subseafloor habitat. *Appl. Environ. Microbiol.* 68: 1585–1594.

Huber JA, Cantin H V, Huse SM, Welch DBM, Sogin ML & Butterfield DA (2010) Isolated communities of Epsilonproteobacteria in hydrothermal vent fluids of the Mariana Arc seamounts. *FEMS Microbiol. Ecol.* 73: 538–549.

Huber JA, Mark Welch DB, Morrison HG, Huse SM, Neal PR, Butterfield DA & Sogin ML (2007) Microbial population structures in the deep marine biosphere. *Science (80-.).* 318: 97–100.

Huse SM, Dethlefsen L, Huber JA, Mark Welch D, Welch DM, Relman DA & Sogin ML (2008) Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing. *PLoS Genet.* 4: e1000255.

Huse SM, Huber JA, Morrison HG, Sogin ML & Mark Welch DB (2007) Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biol.* 8: R143.

Kelley DS, Baross JA & Delaney JR (2002) Volcanoes, fluids, and life at mid-ocean ridge spreading centers. *Annu. Rev. Earth Planet. Sci.* 30: 385–491.

Koonin E V., Makarova KS & Aravind L (2001) Horizontal gene transfer in prokaryotes: Quantification and classification. *Annu. Rev. Microbiol.* 55: 709–742.

Kormas KA, Tivey MK, Von Damm K & Teske A (2006) Bacterial and archaeal phylotypes associated with distinct mineralogical layers of a white smoker spire from a deep-sea hydrothermal vent site (9˚N, East Pacific Rise). *Environ. Microbiol.* 8: 909–920.

Martiny JBH et al. (2006) Microbial biogeography: putting microorganisms on the map. *Nat. Rev. Microbiol.* 4: 102–112.

Nemergut DR et al. (2011) Global patterns in the biogeography of bacterial taxa. *Environ. Microbiol.* 13: 135–144.

Nercessian O, Reysenbach A-L, Prieur D & Jeanthon C (2003) Archaeal diversity associated with in situ samplers deployed on hydrothermal vents on the East Pacific Rise (13˚N). *Environ. Microbiol.* 5: 492–502.

Opatkiewicz AD, Butterfield DA & Baross JA (2009) Individual hydrothermal vents at Axial Seamount harbor distinct subseafloor microbial communities. *FEMS Microbiol. Ecol.* 70: 81–92.

Pedrós-Alió C (2006) Marine microbial diversity: can it be determined? *Trends Microbiol.* 14: 257–263.

Pruesse E, Peplies J & Glöckner FO (2012) SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* 28: 1823–1829.

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: D590–6.

R Core T (2013) R: A Language and Environment for Statistical Computing.

Schloss PD et al. (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75: 7537–7541.

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.

Slobodkin A, Campbell B, Cary SC, Bonch-Osmolovskaya E & Jeanthon C (2001) Evidence for the presence of thermophilic Fe(III)-reducing microorganisms in deep-sea hydrothermal vents at 13 degrees N (East Pacific Rise). *FEMS Microbiol. Ecol.* 36: 235–243.

Sogin ML, Morrison HG, Huber JA, Mark Welch D, Huse SM, Neal PR, Arrieta JM & Herndl GJ (2006) Microbial diversity in the deep sea and the underexplored “rare biosphere”. *Proc. Natl. Acad. Sci. U. S. A.* 103: 12115–12120.

Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22: 2688–2690.

Sul WJ, Oliver TA, Ducklow HW, Amaral-Zettler LA & Sogin ML (2013) Marine bacteria exhibit a bipolar distribution. *Proc. Natl. Acad. Sci. U. S. A.* 110: 2342–2347.

Summit M & Baross JA (2001) A novel microbial habitat in the mid-ocean ridge subseafloor. *Proc. Natl. Acad. Sci. U. S. A.* 98: 2158–2163.

Takai K & Horikoshi K (1999) Genetic diversity of archaea in deep-sea hydrothermal vent environments. *Genetics* 152: 1285–1297.

Takai K, Komatsu T, Inagaki F & Horikoshi K (2001) Distribution of archaea in a black smoker chimney structure. *Appl. Environ. Microbiol.* 67: 3618–3629.

Vergin K, Done B, Carlson C & Giovannoni S (2013) Spatiotemporal distributions of rare bacterioplankton populations indicate adaptive strategies in the oligotrophic ocean. *Aquat. Microb. Ecol.* 71: 1–13.

Figure Legends

Figure 1. Bar charts of bacterial (A) and archaeal (B) taxonomy for all samples. Taxonomy was assigned in VAMPS by the GAST process (Huse *et al.* 2008). Hulk archaeal sample is classified based on v4–v6 sequence; all others are classified based on v6 sequence.

Figure 2. Cluster dendrograms of diffuse flow and sulfide bacterial samples. Cluster dendrograms were created with group average method using the Bray-Curtis dissimilarity index. Operational taxonomic units are defined at the 3% distance for these analyses: A) analysis including all OTUs in each sample; B) analysis including only abundant OTUs (representing 1% or more of all sequences in each sample); and C) analysis including only rare OTUs (representing 0.1% or less of all sequences in each sample). Background samples are marked by asterisks.

Figure 3. Cluster dendrograms of diffuse flow and sulfide archaeal samples. Cluster dendrograms were created with group average method using the Bray-Curtis dissimilarity index. Operational taxonomic units are defined at the 3% distance for these analyses: A) analysis including all OTUs in each sample; B) analysis including only abundant OTUs (representing 1% or more of all sequences in each sample); and C) analysis including only rare OTUs (representing 0.1% or less of all sequences in each sample). Background samples are marked by asterisks.

Figure 4. Heatmap depicting the relative abundance of archaeal OTUs found in Hulk vent compared to other samples. OTUs are ordered according to their abundance in Hulk. OTUs falling roughly at or below the “Rare in Hulk” marker on the heatmap were present at 0.1% abundance or lower in the Hulk sample. White colors indicate that the OTU was not found in a given sample. Background samples are marked with asterisks.

Figure 5. Phylogenetic tree of *Thermococcales* based on 16S rRNA gene sequences, with pyrotag sequences added to the reference tree. Red dots indicate a sequence found in the high-temperature Hulk sample; blue dots, sequences found in diffuse flow samples; and green dots, sequences found in sulfide samples. Collapsed wedges are annotated with the

number of sequences in each cluster that was found in each respective environment. Evolutionary history was inferred using a rapid bootstrap, maximum likelihood method with 100 alternative runs on distinct starting trees, using the GTR+ optimization of substitution rates and the GAMMA model of heterogeneity in RAxML (Stamatakis 2006). The Evolutionary Placement Algorithm (Berger *et al.* 2011) was used to insert short reads into the reference tree. Bootstrap values for the reference tree are labeled where they are over 50.

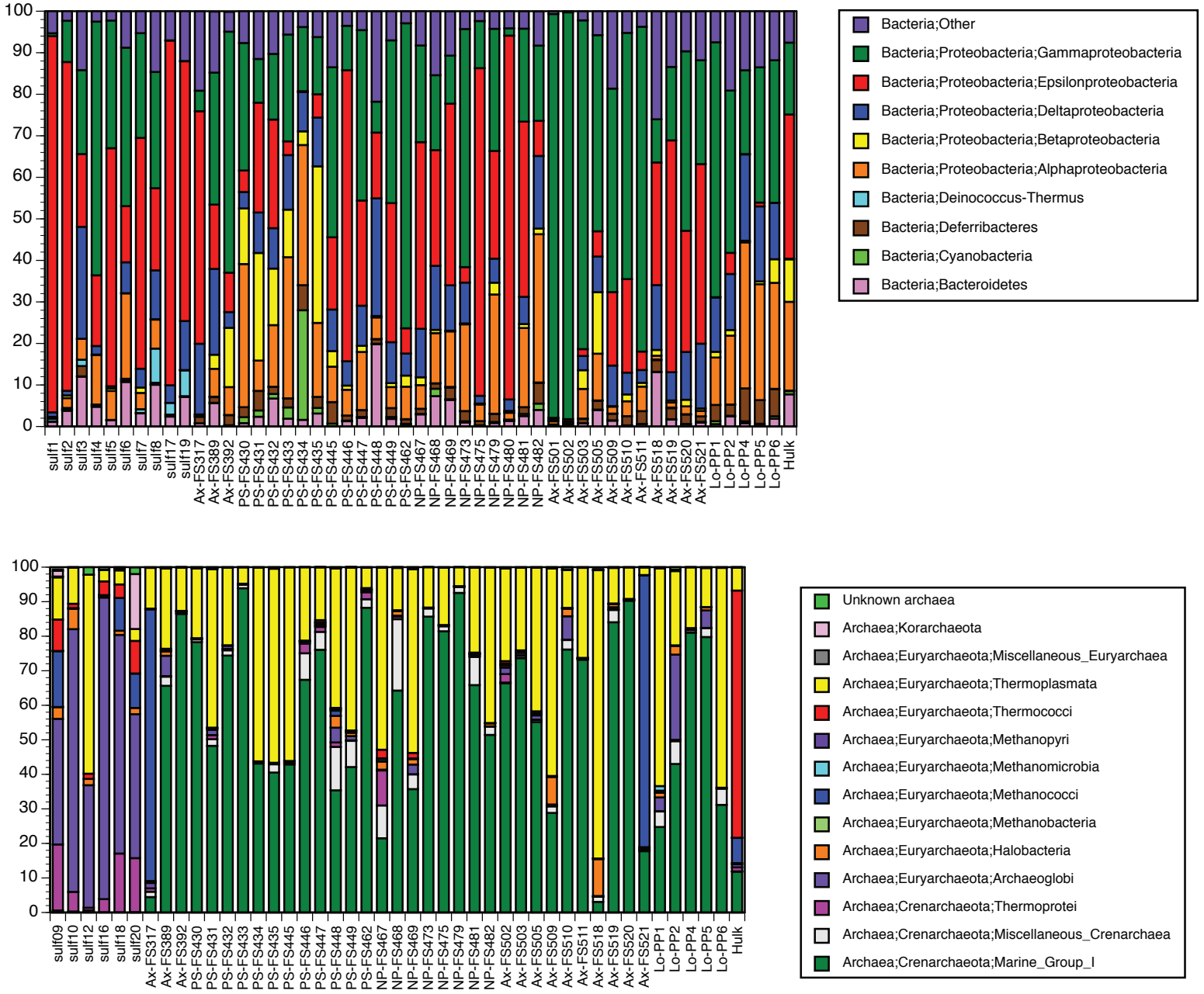


Figure 1. Bar charts of bacterial (A) and archaeal (B) taxonomy for all samples. Taxonomy was assigned in VAMPS by the GAST process (Huse et al. 2008). Hulk archaeal sample is classified based on v4–v6 sequence; all others are classified based on v6 sequence.

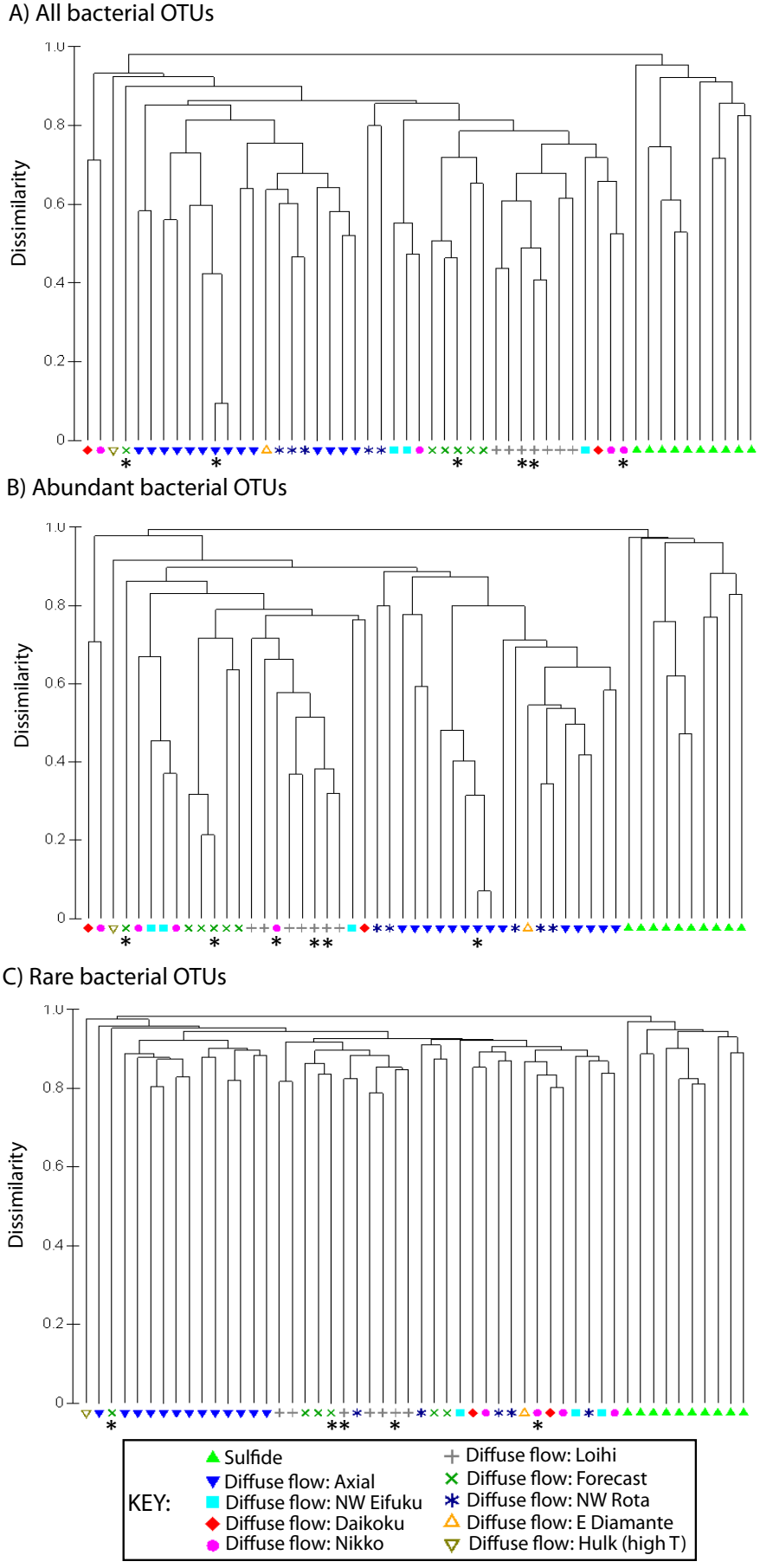
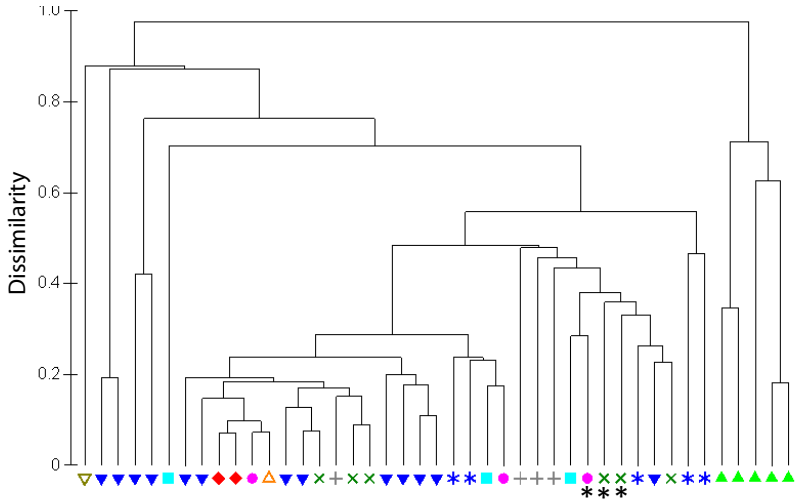
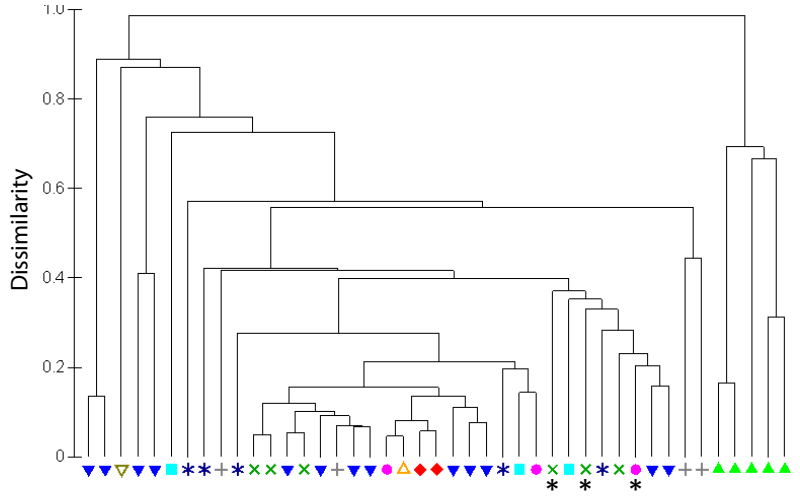


Figure 2. Cluster dendrograms of diffuse flow and sulfide bacterial samples. Cluster dendrograms were created with group average method using the Bray-Curtis dissimilarity index. Operational taxonomic units are defined at the 3% distance for these analyses: A) analysis including all OTUs in each sample; B) analysis including only abundant OTUs (representing 1% or more of all sequences in each sample); and C) analysis including only rare OTUs (representing 0.1% or less of all sequences in each sample). Background samples are marked by asterisks.

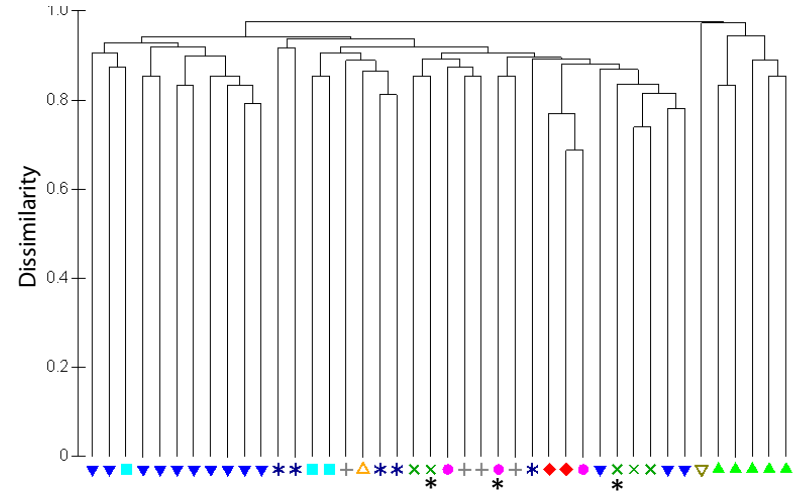
A) All archaeal OTUs



B) Abundant archaeal OTUs



C) Rare archaeal OTUs



KEY:

- Sulfide
- Diffuse flow: Axial
- Diffuse flow: NW Eifuku
- Diffuse flow: Daikoku
- Diffuse flow: Nikko
- Diffuse flow: Loihi
- Diffuse flow: Forecast
- Diffuse flow: NW Rota
- Diffuse flow: E Diamante
- Diffuse flow: Hulk (high T)

Figure 3. Cluster dendrograms of diffuse flow and sulfide archaeal samples. Cluster dendrograms were created with group average method using the Bray-Curtis dissimilarity index. Operational taxonomic units are defined at the 3% distance for these analyses: A) analysis including all OTUs in each sample; B) analysis including only abundant OTUs (representing 1% or more of all sequences in each sample); and C) analysis including only rare OTUs (representing 0.1% or less of all sequences in each sample). Background samples are marked by asterisks.

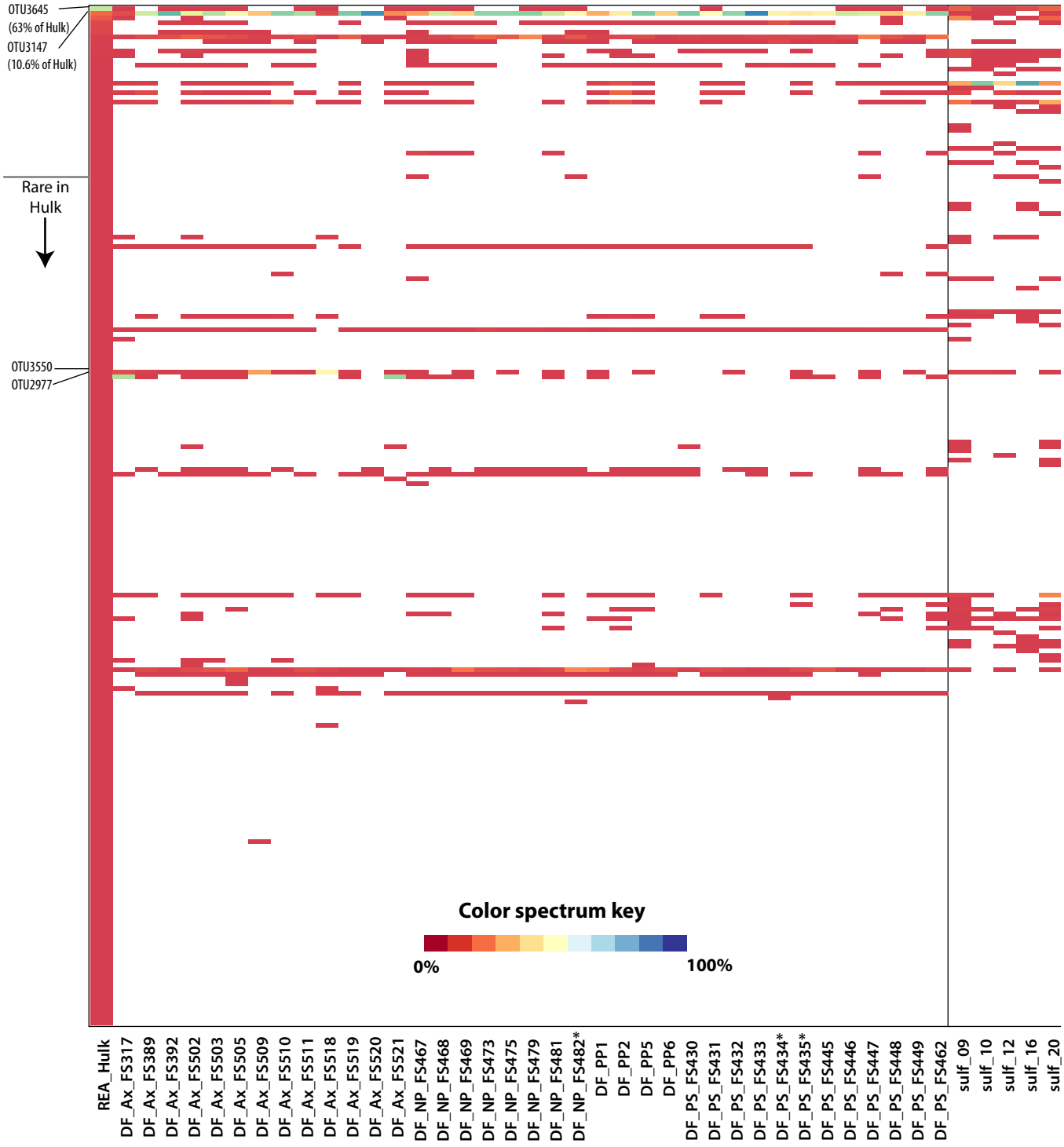


Figure 4. Heatmap depicting the relative abundance of archaeal OTUs found in Hulk vent compared to other samples. OTUs are ordered according to their abundance in Hulk. OTUs falling roughly at or below the "Rare in Hulk" marker on the heatmap were present at 0.1% abundance or lower in the Hulk sample. White colors indicate that the OTU was not found in a given sample. Background samples are marked with asterisks.

Figure 5. Phylogenetic tree of Thermococcales based on 16S rRNA gene sequences, with pyrotag sequences added to the reference tree. Red dots indicate a sequence found in the high-temperature Hulk sample; blue dots, sequences found in diffuse flow samples; and green dots, sequences found in sulfide samples. Collapsed wedges are annotated with the number of sequences in each cluster that was found in each respective environment. Evolutionary history was inferred using a rapid bootstrap, maximum likelihood method with 100 alternative runs on distinct starting trees, using the GTR+ optimization of substitution rates and the GAMMA model of heterogeneity in RAxML (Stamatakis 2006). The Evolutionary Placement Algorithm (Berger et al. 2011) was used to insert short reads into the reference tree. Bootstrap values for the reference tree are labeled where they are over 50.

Supplementary Files

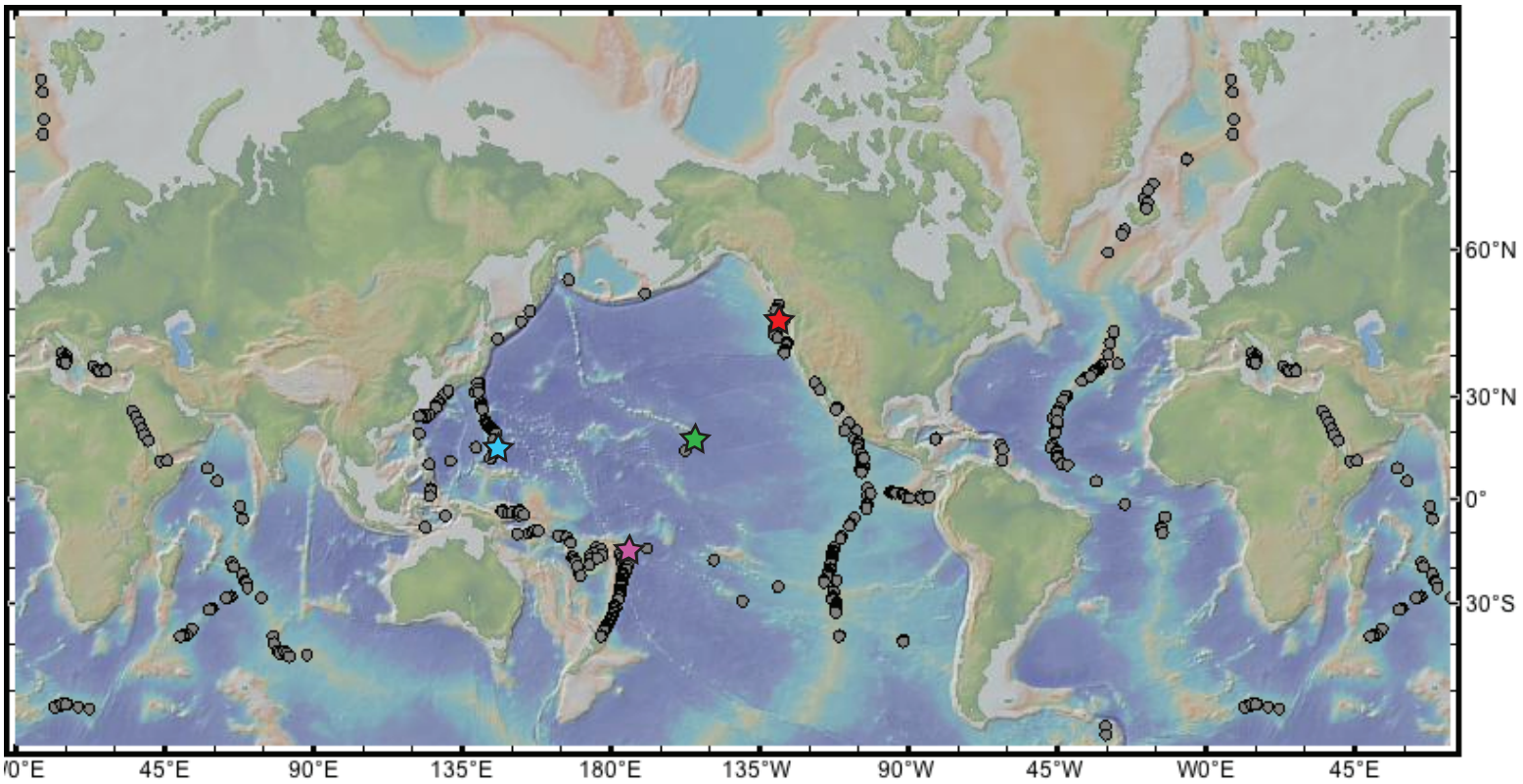


Figure S1. Approximate locations of sampling sites at hydrothermal vents worldwide. Red star: Main Endeavour Field and Axial Seamount, Juan de Fuca Ridge. Blue star: Eifuku, Daikoku, Nikko, Forecast, NW Rota, and E Diamante seamounts, Mariana Arc. Green star: Loihi Seamount, Hawaii. Purple star: Lau Basin. Map was generated using GeoMapApp (http://www.geomapapp.org/).

Table S1. Metadata for all samples used in this study. Metadata for publicly available samples were obtained from the VAMPS database.

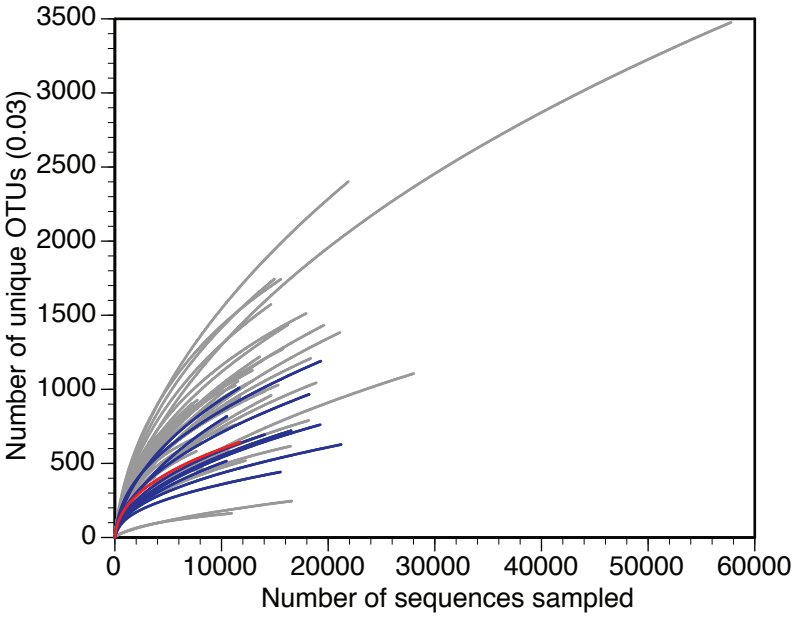
Sample name	Sample type	Depth (m)	Sampling Date	Location	Latitude	Longitude	Domain	Temp (˚C)
sulf_01	active sulfide chimney	2707	4/9/05	Lau Basin, South Pacific Ocean	–20.316686	–176.1363	Bacteria	2.712
sulf_02	active sulfide chimney	2714	4/11/05	Lau Basin, South Pacific Ocean	–20.317851	–176.13737	Bacteria	2.712
sulf_03	active sulfide chimney	2139	4/13/05	Lau Basin, South Pacific Ocean	–20.761027	–176.19081	Bacteria	2.712
sulf_04	active sulfide chimney	1908	4/16/05	Lau Basin, South Pacific Ocean	–22.180673	–176.60124	Bacteria	2.736
sulf_05	microbial mat	1918	4/22/05	Lau Basin, South Pacific Ocean	–22.180185	–176.60081	Bacteria	2.736
sulf_06	active sulfide flange	1918	4/22/05	Lau Basin, South Pacific Ocean	–22.180185	–176.60081	Bacteria	2.736
sulf_07	active sulfide flange	1875	4/24/05	Lau Basin, South Pacific Ocean	–21.989609	–176.56809	Bacteria	2.731
sulf_08	active sulfide chimney	2619	5/3/05	Lau Basin, South Pacific Ocean	–20.053045	–176.13374	Bacteria	2.706
sulf_09	active sulfide chimney	2707	4/9/05	Lau Basin, South Pacific Ocean	–20.316686	–176.1363	Archaea	2.712
sulf_10	active sulfide chimney	2714	4/11/05	Lau Basin, South Pacific Ocean	–20.317851	–176.13737	Archaea	2.712
sulf_12	active sulfide chimney	1908	4/16/05	Lau Basin, South Pacific Ocean	–22.180673	–176.60124	Archaea	2.736
sulf_16	active sulfide chimney	2619	5/3/05	Lau Basin, South Pacific Ocean	–20.053045	–176.13374	Archaea	2.706
sulf_17	active sulfide chimney-bottom	2707	4/9/05	Lau Basin, South Pacific Ocean	–20.316686	–176.1363	Bacteria	2.712
sulf_18	active sulfide chimney-bottom	2707	4/9/05	Lau Basin, South Pacific Ocean	–20.316686	–176.1363	Archaea	2.712
sulf_19	active sulfide chimney-top	2707	4/9/05	Lau Basin, South Pacific	–20.316686	–176.1363	Bacteria	2.712

				Ocean				
sulf_20	active sulfide chimney-top	2707	4/9/05	Lau Basin, South Pacific Ocean	–20.316686	–176.1363	Archaea	2.712
FS317	Hydrothermal fluids	1526	9/4/03	Axial Volcano, North Pacific Ocean	45.9227283	–129.9882383	Both	26.6
FS389	Hydrothermal fluids	1546	9/21/04	Axial Volcano, North Pacific Ocean	45.933583	–130.013583	Both	32.5
FS392	Hydrothermal fluids	1546	9/21/04	Axial Volcano, North Pacific Ocean	45.9337	—130.013617	Both	68.2
FS430	Hydrothermal fluids	1449	4/21/06	Forecast, Philippine Sea	13.394633	143.920096	Both	71
FS431	Hydrothermal fluids	1448	4/21/06	Forecast, Philippine Sea	13.394632	143.920083	Both	6
FS432	Hydrothermal fluids	1451	4/21/06	Forecast, Philippine Sea	13.39532	143.919902	Both	6.5
FS433	Hydrothermal fluids	1447	4/21/06	Forecast, Philippine Sea	13.395265	143.919873	Both	40
FS434	Background seawater	195	4/21/06	Forecast, Philippine Sea	13.3811	143.9021	Both	
FS435	Background seawater	1342.5	4/21/06	Forecast, Philippine Sea	13.3811	143.9021	Both	
FS445	Hydrothermal fluids	560	4/23/06	NW Rota, Philippine Sea	14.600912	144.775483	Both	19.7
FS446	Hydrothermal fluids	534	4/23/06	NW Rota, Philippine Sea	14.60085	144.77632	Both	48
FS447	Hydrothermal fluids	521	4/23/06	NW Rota, Philippine Sea	14.601177	144.775618	Both	29
FS448	Hydrothermal fluids	584	4/23/06	NW Rota, Philippine Sea	14.60084	144.7773	Both	25
FS449	Hydrothermal fluids	568	4/23/06	NW Rota, Philippine Sea	14.60081	144.77751	Both	15.1
FS462	Hydrothermal fluids	353	4/30/06	E Diamante, North Pacific Ocean	15.94277	145.68141	Both	22.5
FS467	Hydrothermal fluids	1612	5/4/06	Eifuku, North Pacific Ocean	21.48742	144.04163	Both	42.9
FS468	Hydrothermal fluids	1578	5/4/06	Eifuku, North Pacific Ocean	21.487248	144.042123	Both	45.1
FS469	Hydrothermal fluids	1578	5/4/06	Eifuku, North Pacific Ocean	21.487248	144.042123	Both	33.7

FS473	Hydrothermal fluids	438	5/5/06	Daikoku, North Pacific Ocean	21.324536	144.19293	Both	15.3
FS475	Hydrothermal fluids	414	5/5/06	Daikoku, North Pacific Ocean	21.324962	144.19139	Both	45.5
FS479	Hydrothermal fluids	458	5/8/06	Nikko, North Pacific Ocean	23.081017	142.325483	Both	80.2
FS480	Hydrothermal fluids	445	5/8/06	Nikko, North Pacific Ocean	23.07913	142.326433	Bacteria	24.1
FS481	Hydrothermal fluids	413	5/8/06	Nikko, North Pacific Ocean	23.07977	142.32687	Archaea	32.6
FS482	Background seawater	344	5/8/06	Nikko, North Pacific Ocean	23.077802	142.325151	Both	14.7
FS501	Background seawater	1526	9/1/06	Axial Volcano, North Pacific Ocean	45.94667	–129.98439	Bacteria	2.4
FS502	Hydrothermal fluids	1529	9/1/06	Axial Volcano, North Pacific Ocean	45.94632	–129.98398	Archaea	83.4
FS503	Hydrothermal fluids	1530	9/1/06	Axial Volcano, North Pacific Ocean	45.94364	–29.98519	Both	
FS505	Hydrothermal fluids	1524	9/1/06	Axial Volcano, North Pacific Ocean	45.93319	–129.98223	Both	
FS509	Hydrothermal fluids	1546	9/3/06	Axial Volcano, North Pacific Ocean	45.93357	–130.01329	Both	24.6
FS510	Hydrothermal fluids	1546	9/3/06	Axial Volcano, North Pacific Ocean	45.93331	–130.01334	Both	49
FS511	Hydrothermal fluids	1546	9/3/06	Axial Volcano, North Pacific Ocean	45.93364	–130.01329	Both	96.8
FS518	Hydrothermal fluids	1546	9/3/06	Axial Volcano, North Pacific Ocean	45.93357	–130.01329	Both	
FS519	Hydrothermal fluids	1538	9/4/06	Axial Volcano, North Pacific Ocean	45.91724	–129.99299	Both	29.6
FS520	Hydrothermal fluids	1536	9/4/06	Axial Volcano, North Pacific Ocean	45.91631	–129.98916	Both	14.9
FS521	Hydrothermal fluids	1524	9/4/06	Axial Volcano, North Pacific Ocean	45.92279	–129.98838	Both	27.5
LOIHI-PP1	Hydrothermal fluids	1272	10/27/06	Loihi Seamount, North Pacific	18.900833	–155.261389	Both	

				Ocean				
LOIHI-PP2	Hydrothermal fluids	1302	10/27/06	Loihi Seamount, North Pacific Ocean	18.910278	–155.25111	Both	
LOIHI-PP4	Hydrothermal fluids	4983	11/3/06	Loihi Seamount, North Pacific Ocean	18.703056	–155.180833	Bacteria	
LOIHI-PP5	Hydrothermal fluids	1308	11/4/06	Loihi Seamount, North Pacific Ocean	18.31222	–155.26111	Both	
LOIHI-PP6	Hydrothermal fluids	4988	11/6/06	Loihi Seamount, North Pacific Ocean	18.703056	–155.180833	Both	
LOIHI-CTD03	Background seawater	1100	10/30/06	Loihi Seamount, North Pacific Ocean	18.911667	–155.26194	Bacteria	
CTDBtl12	Background seawater		4/24/06	NW Rota, Philippine Sea	14.644167	–144.56667	Bacteria	6.3
Hulk	Hydrothermal fluids	2178	6/17/2009	Juan de Fuca Ridge, North Pacific Ocean	47.9500	–129.0968	Both	125

A)



B)

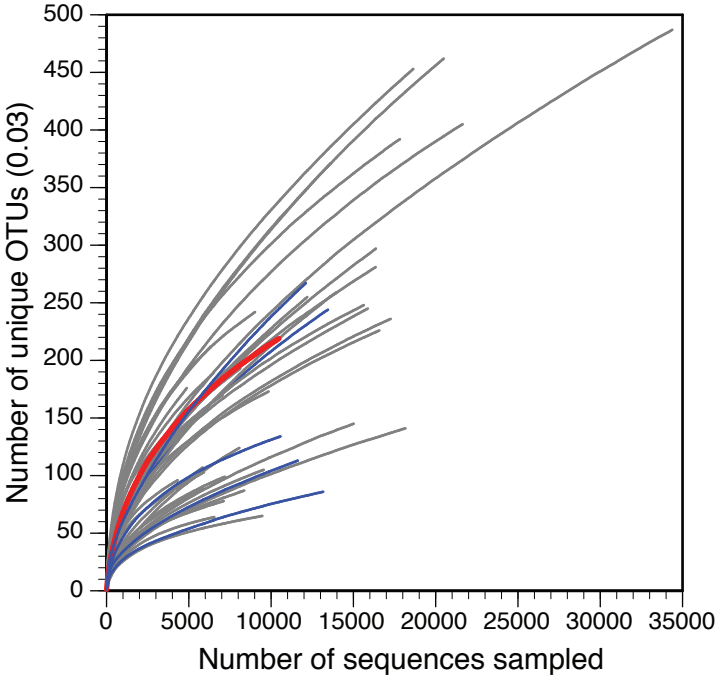


Figure S2. Rarefaction curves of A) bacterial and b) archaeal samples. Diffuse flow samples are noted in grey, sulfide samples are noted in blue, and Hulk sample is noted in red. Rarefaction curves were generated in mothur (Schloss et al., 2009).

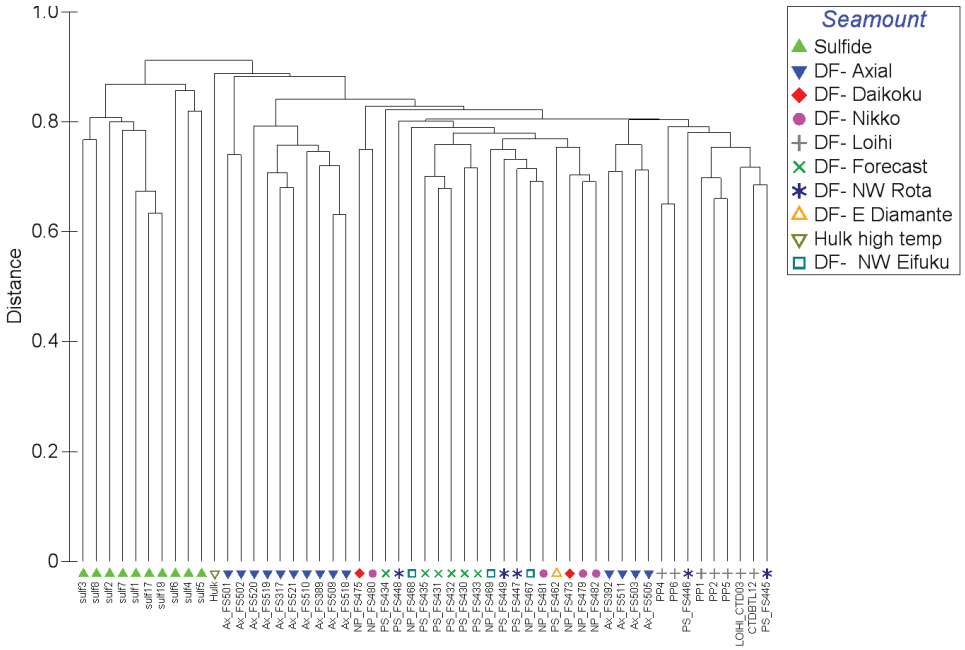
Table S2. Evenness indices for each of the samples used in this study. Both the Shannon and Simpson indices are reported for both domains. All indices were calculated in mothur (Schloss *et al*., 2009). The overall difference in evenness between archaea and bacteria is significant for the Shannon test (t-test, Shannon p-value = 7.81E-18, Simpson p-value = 0.061).

	Bacteria		**Archaea**	
Group	*Simpson*	*Shannon*	*Simpson*	*Shannon*
DF_Ax_FS317	0.037	0.73	0.0096	0.31
DF_Ax_FS389	0.012	0.66	0.011	0.38
DF_Ax_FS392	0.012	0.51	0.017	0.21
DF_Ax_FS501	0.010	0.20		
DF_Ax_FS502	0.0060	0.16	0.010	0.36
DF_Ax_FS503	0.0041	0.40	0.0092	0.29
DF_Ax_FS505	0.011	0.59	0.010	0.36
DF_Ax_FS509	0.017	0.62	0.035	0.54
DF_Ax_FS510	0.025	0.63	0.010	0.31
DF_Ax_FS511	0.021	0.55	0.020	0.29
DF_Ax_FS518	0.061	0.73	0.027	0.50
DF_Ax_FS519	0.035	0.72	0.010	0.26
DF_Ax_FS520	0.031	0.73	0.019	0.16
DF_Ax_FS521	0.037	0.71	0.015	0.22
DF_NP_FS467	0.028	0.66	0.019	0.50
DF_NP_FS468	0.011	0.64	0.010	0.35
DF_NP_FS469	0.037	0.74	0.018	0.48
DF_NP_FS473	0.0037	0.48	0.013	0.25
DF_NP_FS475	0.012	0.49	0.019	0.28
DF_NP_FS479	0.023	0.63	0.012	0.21
DF_NP_FS480	0.031	0.60		
DF_NP_FS481	0.026	0.66	0.010	0.36
DF_NP_FS482	0.022	0.71	0.014	0.39
DF_Lo_PP1	0.026	0.63	0.018	0.47
DF_Lo_PP2	0.031	0.69	0.016	0.49

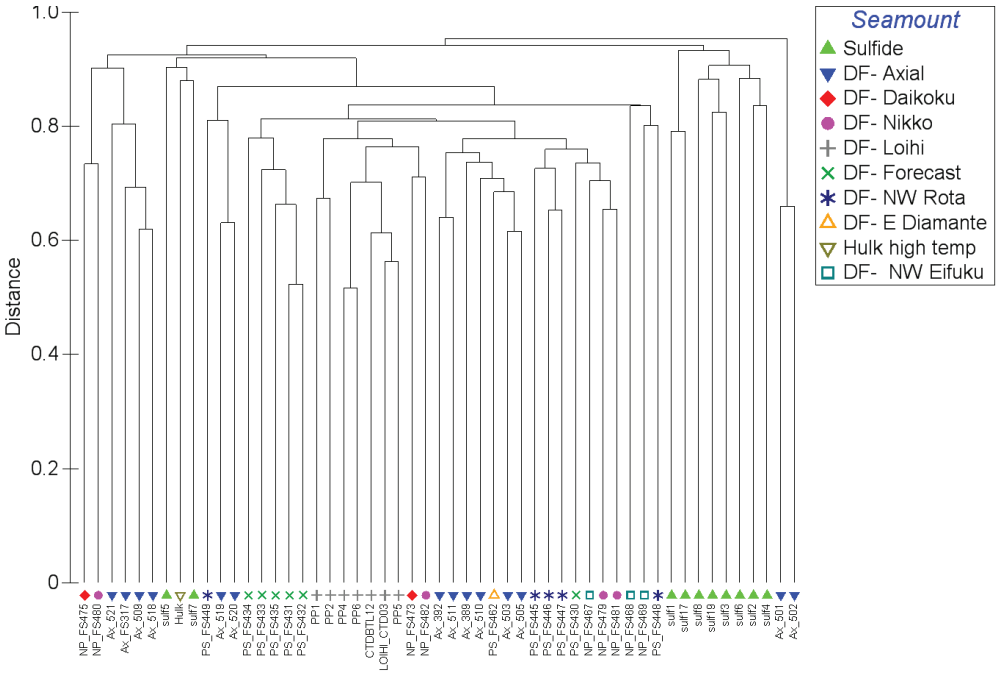
DF_Lo_PP4	0.015	0.62		
DF_Lo_PP5	0.020	0.65	0.0036	0.24
DF_Lo_PP6	0.023	0.66	0.024	0.43
DF_PS_FS430	0.079	0.74	0.022	0.26
DF_PS_FS431	0.0085	0.66	0.024	0.43
DF_PS_FS432	0.024	0.75	0.019	0.31
DF_PS_FS433	0.039	0.70	0.019	0.13
DF_PS_FS434	0.011	0.60	0.044	0.49
DF_PS_FS435	0.0098	0.60	0.018	0.40
DF_PS_FS445	0.045	0.69	0.048	0.45
DF_PS_FS446	0.012	0.47	0.023	0.36
DF_PS_FS447	0.024	0.68	0.015	0.36
DF_PS_FS448	0.034	0.72	0.016	0.50
DF_PS_FS449	0.027	0.66	0.029	0.52
DF_PS_FS462	0.0060	0.48	0.0079	0.24
Hulk high temp	0.020	0.60	0.011	0.33
sulf_01	0.0065	0.44		
sulf_02	0.023	0.57		
sulf_03	0.037	0.71		
sulf_04	0.019	0.63		
sulf_05	0.012	0.55		
sulf_06	0.011	0.60		
sulf_07	0.0071	0.46		
sulf_08	0.037	0.68		
sulf_09			0.039	0.52
sulf_10			0.015	0.27
sulf_12			0.022	0.33
sulf_16			0.016	0.18
sulf_17	0.0073	0.42		
sulf_19	0.021	0.57		
sulf_20			0.029	0.47

CTDBTL12	0.048	0.72		
LOIHI_CTD03	0.016	0.58		
Average	**0.023**	**0.61**	**0.019**	**0.35**

A) All Bacterial OTUs



B) Abundant Bacterial OTUs



C) Rare Bacterial OTUs

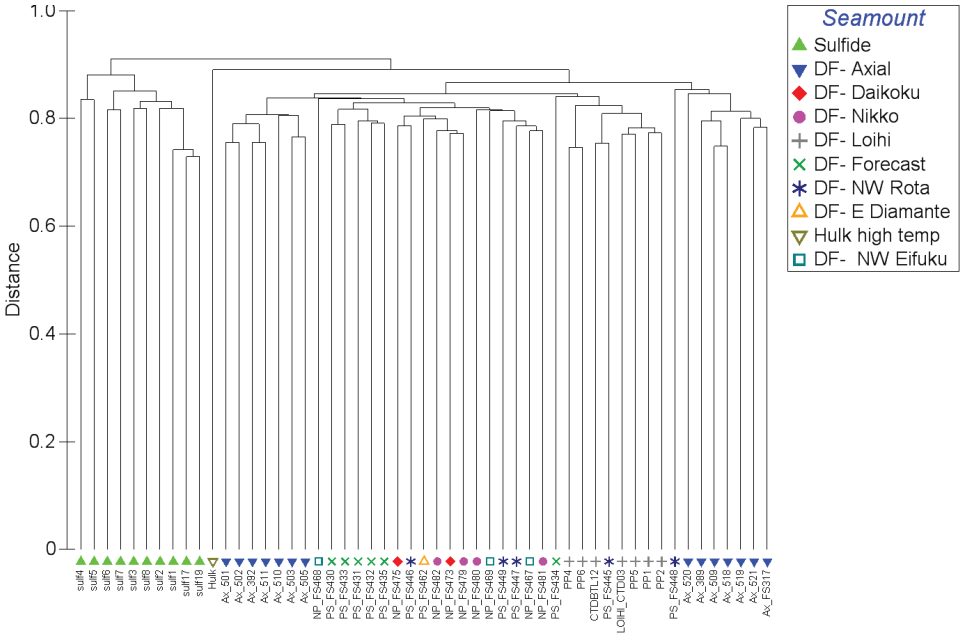
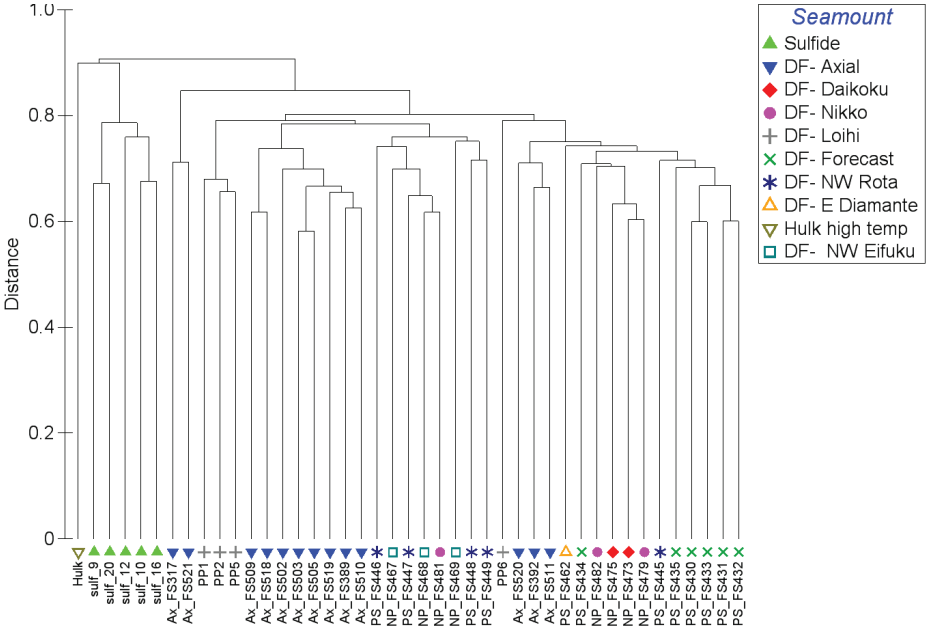
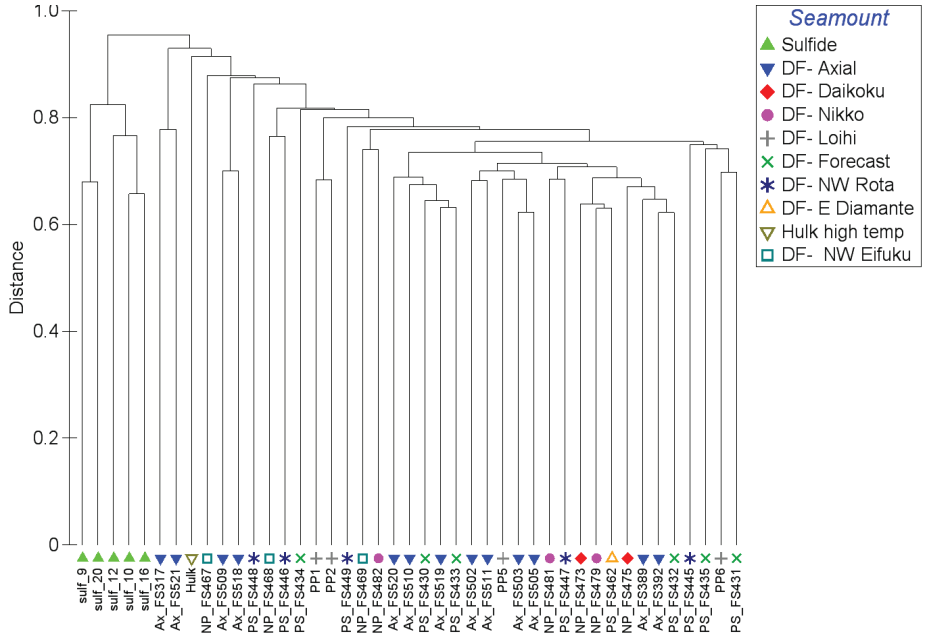


Figure S3. Cluster dendrograms of diffuse flow and sulfide bacterial samples. Cluster dendrograms were created with group average method using distance matrices calculated using the Unifrac method after creating a phylogenetic tree for all sequences in each sample using clearcut within the mothur package (Schloss et al., 2009). For comparison between samples, we randomly subsampled the dataset 1000 times to the number of sequences contained within the smallest sample. We used the same previously-defined sequences within the abundant and rare groupings for this analysis. A) analysis including all OTUs in each sample; B) analysis including only abundant OTUs (representing 1% or more of all sequences in each sample); and C) analysis including only rare OTUs (representing 0.1% or less of all sequences in each sample). Background samples are marked by asterisks, and were collected either with a Niskin rosette on a CTD (labeled "CTD__") or were collected by the hydrothermal fluid sampler on the bottom during travel between or away from vents (labeled "FS__"). Samples are labeled according to fluid sample number, seamount, and region: NP = North Pacific, Ax = Axial Seamount, PS = Philippine Sea, Lo = Loihi Seamount.

A) All Archaeal OTUs



B) Abundant Archaeal OTUs



C) Rare Archaeal OTUs

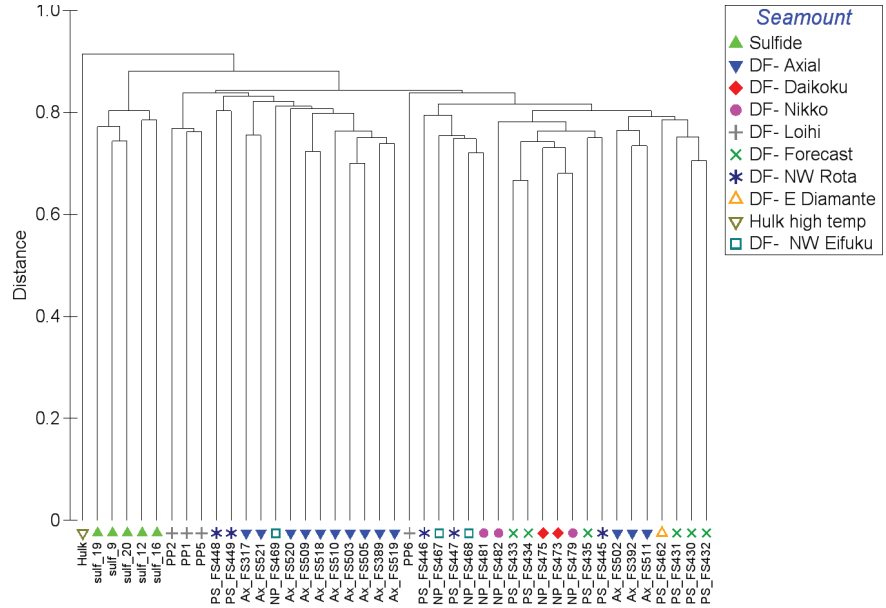


Figure S4. Cluster dendrograms of diffuse flow and sulfide archaeal samples. Cluster dendrograms were created with group average method using distance matrices calculated using the Unifrac method after creating a phylogenetic tree for all sequences in each sample using clearcut within the mothur package (Schloss et al., 2009). For comparison between samples, we randomly subsampled the dataset 1000 times to the number of sequences contained within the smallest sample. We used the same previously-defined sequences within the abundant and rare groupings for this analysis. A) analysis including all OTUs in each sample; B) analysis including only abundant OTUs (representing 1% or more of all sequences in each sample); and C) analysis including only rare OTUs (representing 0.1% or less of all sequences in each sample). Background samples are marked by asterisks, and were collected either with a Niskin rosette on a CTD (labeled "CTD__") or were collected by the hydrothermal fluid sampler on the bottom during travel between or away from vents (labeled "FS__"). Samples are labeled according to fluid sample number, seamount, and region: NP = North Pacific, Ax = Axial Seamount, PS = Philippine Sea, Lo = Loihi Seamount.

Table S3. ANOSIM results for bacterial and archaeal datasets, grouped according to environment as listed in Figures 1 and 2. ANOSIM was conducted as a one-way analysis on a distance matrix calculated with the Unifrac method among samples. Trees of all sequences in each sample were constructed using clearcut in the mothur package (Schloss et al., 2009), and distance matrices were calculated according to the Unifrac method (Lozupone & Knight, 2005) in the mothur package. ANOSIM analysis was conducted using the PRIMERv6 software package (Clarke & Gorley, 2006). We conducted nine hundred ninety nine permutations of the test for each ANOSIM analysis. A test is considered significant if $p \leq 0.001$. Clustering was significant even after removing sulfides from the analysis, indicating that differentiation occurred not only due to sample type but also due to sample location. Statistical analyses remained the same even after OTU singletons (only appearing once in a sample) were removed (data not shown).

Domain	**Grouping**	**With sulfides**			**Without sulfides**		
		R statistic	*p-value*	*Significant?*	*R statistic*	*p-value*	*Significant?*
Bacteria	All	0.534	<0.001	Yes	0.443	<0.001	Yes
	Abundant	0.307	<0.001	Yes	0.253	0.002	No
	Rare	0.588	<0.001	Yes	0.498	<0.001	Yes
Archaea	All	0.587	<0.001	Yes	0.468	<0.001	Yes
	Abundant	0.307	0.003	No	0.092	0.18	No
	Rare	0.577	<0.001	Yes	0.462	<0.001	Yes

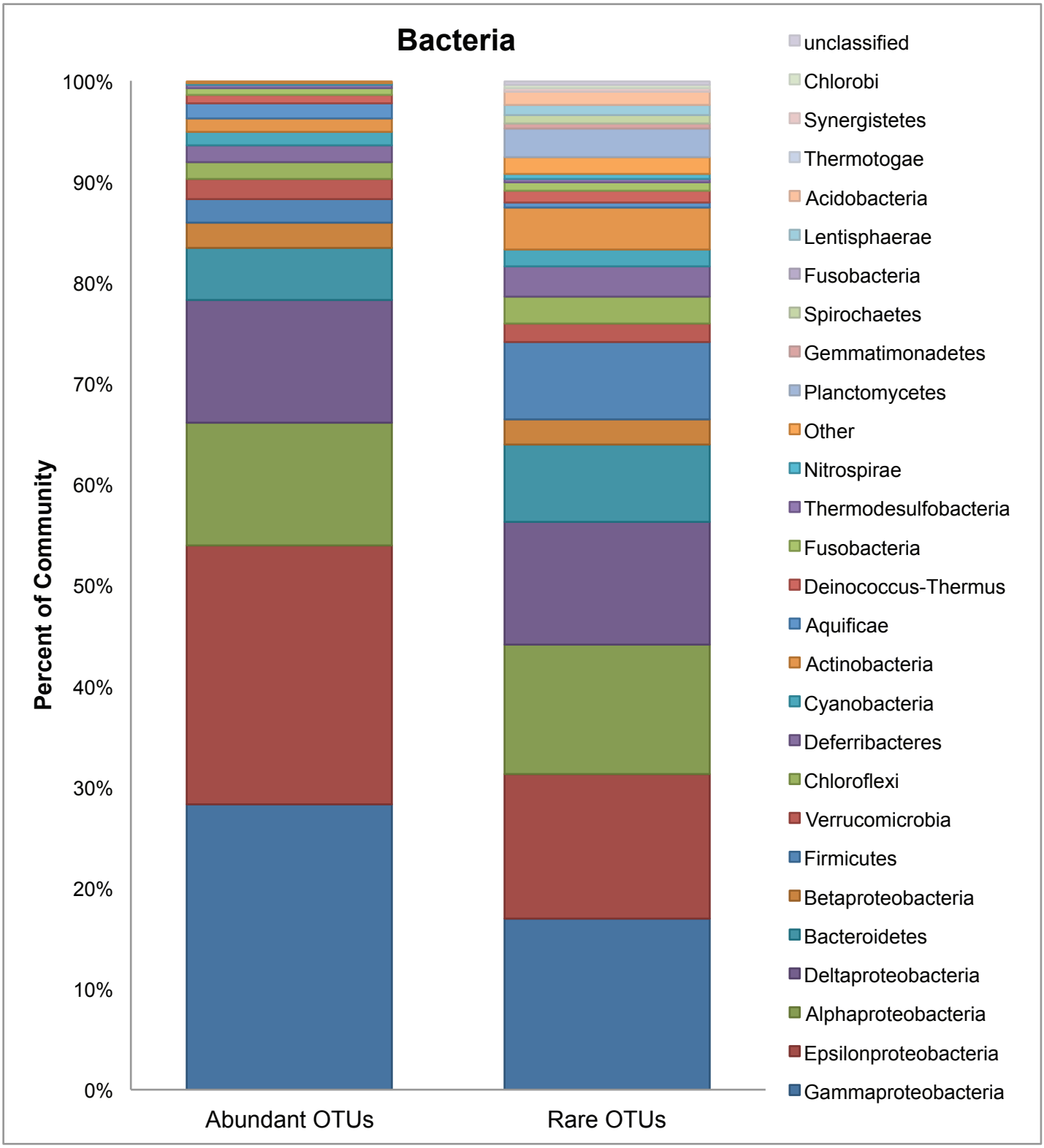


Figure S5. Bar charts of bacterial taxonomy for abundant and rare OTUs. Each category includes OTUs that were abundant or rare in at least one sample. There were 434 OTUs that were abundant in at least one sample, and 21733 OTUs that were rare in at least one sample. Taxonomy was assigned in mothur (Schloss *et al.*, 2009) according to alignment to the SILVA database (Quast *et al.*, 2013).

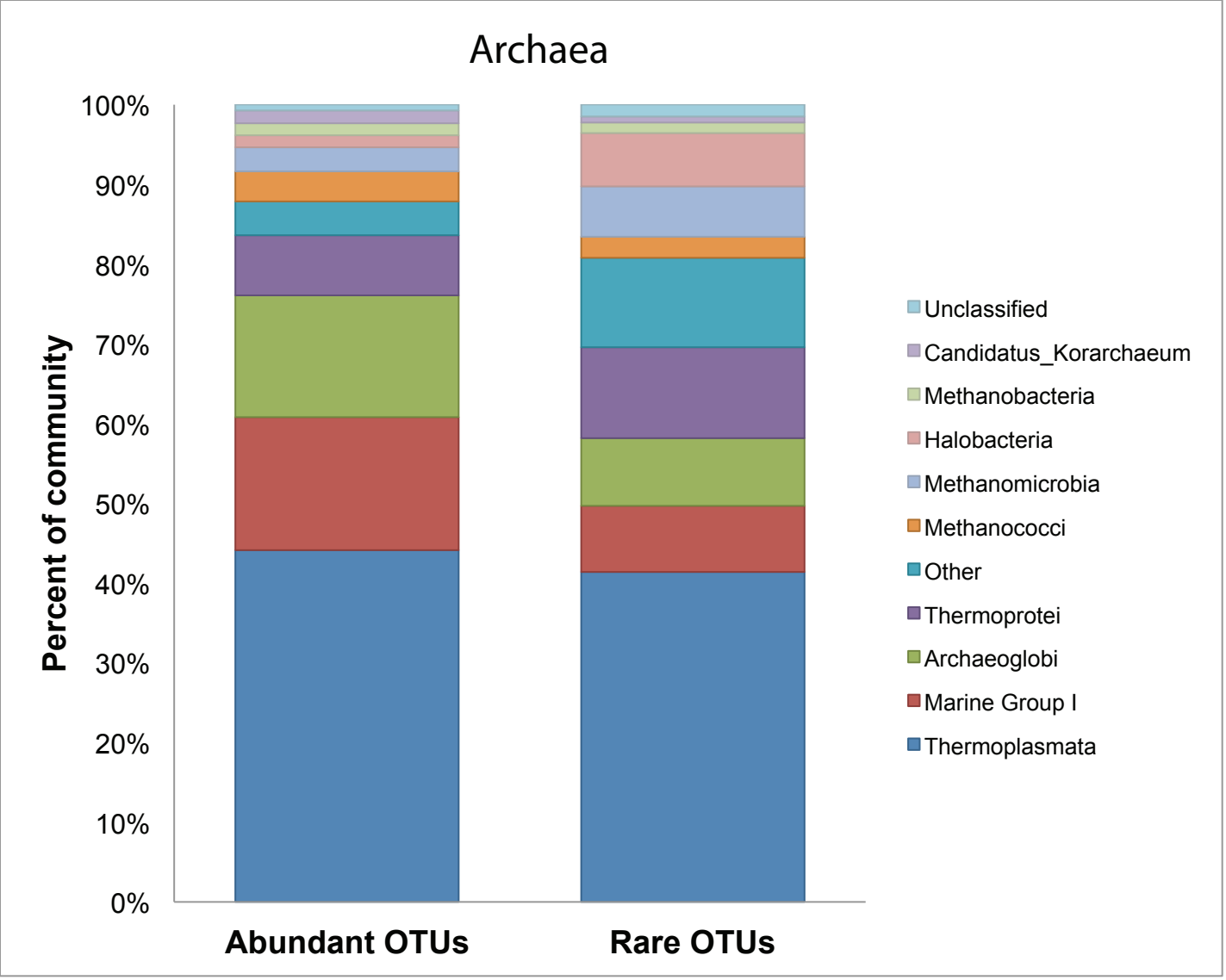


Figure S6. Bar charts of archaeal taxonomy for abundant and rare OTUs. Each category includes OTUs that were abundant or rare in at least one sample. There were 263 OTUs that were abundant in at least one sample, and 3435 OTUs that were rare in at least one sample. Taxonomy was assigned in mothur (Schloss *et al.*, 2009) according to alignment to the SILVA database (Quast *et al.*, 2013)

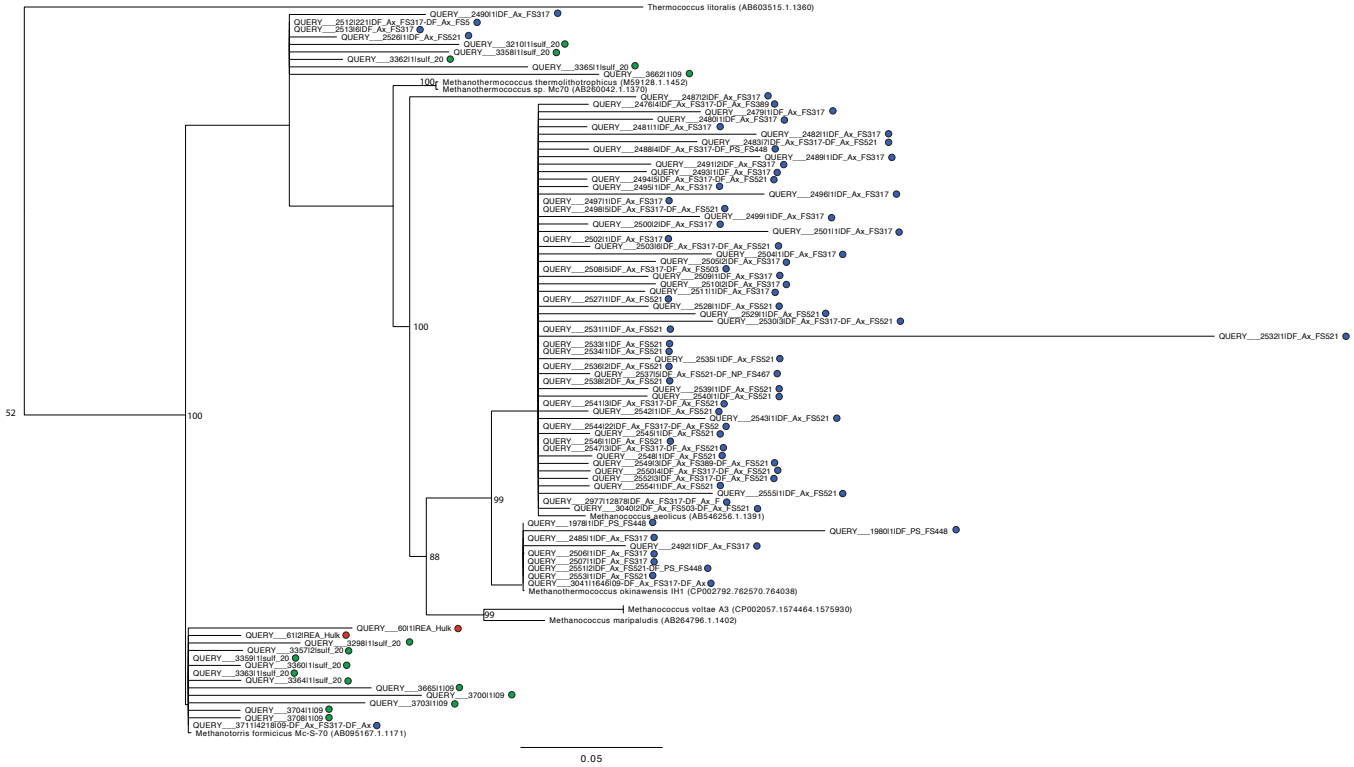


Figure S7. Phylogenetic tree of *Methanococcales* based on 16S rRNA gene sequences, with pyrotag sequences added into reference tree. Red dots indicate a sequence found in the high-temperature Hulk sample, blue dots indicate sequences found in diffuse flow samples, and green dots indicate sequences found in sulfide samples. Collapsed wedges are annotated with the number of sequences in each cluster that was found in each respective environment. Evolutionary history was inferred using a rapid bootstrap, maximum likelihood method with 100 alternative runs on distinct starting trees, using the GTR+ optimization of substitution rates and the GAMMA model of heterogeneity in RAxML (Stamatakis 2006). The Evolutionary Placement Algorithm (EPA) (Berger et al, 2011) was used to insert short reads into the reference tree. Bootstrap values for the reference tree are labeled.