Relating prokaryotic and microeukaryotic diversity to community function and ecosystem variability at deep-sea hydrothermal vents

by

Sheryl Murdock
B.Sc., University of Washington, 1998

A Dissertation Submitted in Partial Fulfillment of the Requirements
for the Degree of

DOCTOR OF PHILOSOPHY

in the School of Earth and Ocean Sciences

Relating prokaryotic and microeukaryotic diversity to community function and ecosystem variability at deep-sea hydrothermal vents

by

Sheryl Murdock
B.Sc., University of Washington, 1998

Supervisory Committee

Dr. S. Kim Juniper, Supervisor
School of Earth and Ocean Sciences

Dr. John Dower, Departmental Member
School of Earth and Ocean Sciences

Dr. Verena Tunnicliffe, Outside Member
Department of Biology

Dr. Connie Lovejoy, Additional Member
Université Laval, Department of Biology

Abstract

Despite over four decades of research on deep-sea hydrothermal vent ecosystems, major gaps remain in our understanding of these systems. Knowledge of microeukaryote diversity, abundance, and involvement in ecosystem function lags far behind that of prokaryotes, and contributions of the non-endosymbiotic microbiome in faunal assemblages to ecosystem processes and overall hydrothermal vent microbial diversity are not known. This research addresses these gaps using high-throughput sequencing of 16S/18S rRNA genes and metagenomes from vent and surrounding non-vent habitats encompassing diffuse hydrothermal fluids, plumes, deep seawater, and microbes in assemblages of the foundation tubeworm species *Ridgeia piscesae*. Co-occurrence/covariance is a central method used, first, between prokaryotes with known extreme habitat preferences and microeukaryotes to infer potential endemism in the latter, and then between microbes and fauna in *R. piscesae* assemblages to infer interspecies interactions. Microeukaryote distribution and abundance suggest potential vent endemic microeukarya are infrequently encountered, potentially in low abundance, and belong to novel lineages of Rhizaria and Stramenopila. Potential endemism is inferred for relatives of known apusomonads, excavates, and some clades of Syndiniales. *R. piscesae* assemblages are shown to be hotspots of microbial taxonomic richness and exhibit a robust temperature-driven distinction in assemblage composition above and below ~25°C spanning micro, meio and macro size classes and microbial domains (Bacteria, Archaea, and micro-Eukarya). Likely interacting faunal and microbial taxa among *R. piscesae* assemblages are identified as 'core communities', which included eight macro-

and meiofaunal taxa and members of the Bacteroidetes and Epsilonbacteraeota in highT communities (>25°C) and more meiofaunal species in addition to Alpha- and Gammaproteobacteria, and Actinobacteria in lowT communities (<25°C). Core communities were used to guide metagenomic investigations of microbial functional potential. Exploratory metagenomic analysis required development of new methods to deal with compositional data. 'Enrichment leanings' were developed to prioritize in-depth functional comparisons between sample types, which revealed clades within core community microbes with differing functional potential between highT and lowT assemblages and between assemblages and fluids. The balance of autotrophy-heterotrophy genes and patterns of genes for different carbon, nitrogen, and sulfur-cycling processes were tested as potential metrics of community-level function but did not distinguish assemblages by highT/lowT designations. This research brings us closer to understanding hydrothermal vent ecosystem function and suggests sizeable continued discovery potential.

Table of Contents

List of figures

List of tables

Acknowledgements

I am grateful to so many people for their encouragement and patience throughout this process. First of all, thanks to Kim Juniper for supporting me through every crazy idea I decided to pursue and for not saying "I told you so" when it didn't work out as planned. Thank you to Verena Tunnicliffe for answering my incessant questions about vent fauna, sharing with me your wisdom and the occasional glass of wine. Thanks to all the members of my committee for their sage advice and guidance.

I would like to thank the many friends and colleagues that have encouraged me, both professionally and personally. Julie Huber and Jon Kaye for keeping me grounded but always with support and encouragement, Cat Stevens for good chats and teaching me about lipids (I still don't get them), Dave Butterfield for so many exhausting yet awesome sampling experiences at sea, Cherisse DuPreez for just being awesome (we should definitely co-present more often), Nathalie Forget for many entertaining hours in the lab together and introducing me to good poutine and raclette.

I am grateful to the many funders and private donors from whom I received awards and scholarships—The Natural Sciences and Engineering Research Council, the Canadian Federation of University Women, and the Bob Wright and Montalbano scholarship funds.

Most importantly, I would like to thank my family—my mom, who has never wavered in her view that I can accomplish anything, and my husband for encouraging me into this journey, taking on so much responsibility so that I could focus on research, and keeping me balanced. And of course, to Kira and Simone, I am so thankful for their patience as I worked long hours and made trips away and for their encouragement to make it to the end.

Dedication

This work is dedicated to my two amazing daughters, Kira and Simone, who have brought me joy throughout this journey and for whom I have immense respect. May they have long and fulfilling futures filled with curiosity and wonder.

It is also dedicated to every woman or girl who has ever been told she wasn't enough. There is nothing we cannot accomplish.

Veni. Vidi. Vici.

Chapter 1. General introduction

Life at deep-sea hydrothermal vents exists among steep physico-chemical gradients that form between hot, reducing hydrothermal fluids and cold oxygenated seawater. The reduction-oxidation potential within these gradients fuels abundant microbial chemosynthetic primary production that supports oases of life in an otherwise energy-starved deep sea (Reysenbach et al., 2000; Govenar, 2012; Sievert and Vetriani, 2012). Current understanding of microbial roles in hydrothermal vent ecosystems stems from over four decades of research that has emphasized identification and description of extremophiles and chemoautotrophic primary producers, primarily from hydrothermal fluids and chimney surfaces and from microbes in endosymbiotic relationships with vent animals (Dick, 2019). This somewhat limited focus has left critical gaps in our understanding of microbial involvement in hydrothermal vent ecosystems. Specifically, we lack basic information about the abundance, diversity, and functional roles of microbial eukaryotes in hydrothermal vent ecosystems and have generally overlooked the functional contributions of non-endosymbiotic microbes in faunal assemblages. This research aims to address these knowledge gaps to help create a more comprehensive characterization of deep-sea hydrothermal vent ecosystems.

Improving the characterization of these ecosystems has potential future benefits for monitoring their health in the face of changing ocean conditions and potential threats as well as for identifying novel genetic resources. Comprehensive characterization to define the normal range of biological communities inhabiting a range of environmental conditions can

lead to improved recognition of excursions from normal that may be indicative of environmental stress (Carignan and Villard, 2002). This knowledge may prove useful with growing climate change influence in the deep ocean (Levin and Le Bris, 2015) or with initiation of planned seabed mining in hydrothermal systems (Thompson et al., 2018). Furthermore, the genetic novelty in all deep-sea organisms from extreme habitats holds potential promise for biomedical and biotechnological research that could result in products with widespread benefits to humankind (Thornburg et al., 2010; Navarri et al., 2016; Shi et al., 2017; Xu et al., 2017). Comprehensive characterization of these captivating ecosystems should therefore include diversity and functional contributions from all levels of community membership, from viruses to megafauna, across all major habitat types.

One subset of the microbial community that has been largely overlooked in terms of their abundance, distribution, and functional contributions in hydrothermal vent ecosystems is the microeukaryotes (protists and fungi). However, this oversight is not unique to hydrothermal vents as microeukaryotes below the photic zone have, until recently, received limited attention (Herndl et al., 2008; Arístegui et al., 2009). Although morphological studies of deep-sea microeukarya go back nearly two centuries (Gooday et al., 2020) and have provided information on the general forms (e.g., flagellate, ciliate, amoeba, etc.) and identities to the closest known relatives, knowledge surrounding their functional roles and diversity lags far behind. This disparity is beginning to change as microeukaryotes are increasingly incorporated into large research programs to understand how their distribution and abundance vary with depth and geographic location and how they contribute to marine biodiversity on a global scale (Duarte, 2015; Pernice et al., 2015; Pernice et al., 2016; Giner et al., 2020).

The 2010 Malaspina expedition, a circumnavigation cruise aimed at exploring the biodiversity of deep regions of the Pacific, Atlantic, and Indian Oceans (Duarte, 2015), has contributed greatly to knowledge of the abundance and distribution of deep-ocean microeukaryotes. Ubiquitous heterotrophic protists inhabiting global bathypelagic zones (1000-4000m) decrease in abundance, biomass, and diversity with increasing depth, correlating with a decrease in prokaryote abundance and suggesting control of prokaryote populations by grazing pressure of protists (Pernice et al., 2015; Giner et al., 2020). Varying dominance among groups of bathypelagic microeukaryotes in high-throughput DNA marker gene sequences can be partially explained by water mass and ocean basin (Pernice et al., 2016). For example, Circumpolar Deep Waters are dominated by Collodaria, an order within the class Polycystinea (Rhizaria), and North Atlantic Deep Waters are dominated by members of the Marine Alveolate Group II (MALV-II), belonging to the order Syndiniales (Alveolata). Furthermore, fungal sequences, abundant in the Pacific and Indian Oceans, are barely detected in the Atlantic. The DNA-based composition of picoeukaryotes (0.2-3μm) was also distinct between photic and aphotic zones for all three oceans sampled during the Malaspina expedition (Giner et al., 2020)

Major players among deep-sea microeukaryotes belong to Rhizaria, Stramenopila, Alveolata, and Fungi (Pernice et al., 2016; Giner et al., 2020). The Rhizaria encompass numerous species of free-living flagellates, amoeboflagellates, parasitic protists, and abundant marine amoeboid forms such as radiolaria and foraminifera (Pawlowski and Burki, 2009), as well as enormous untapped discovery potential of novel species (Burki and Keeling, 2014). Stramenopiles are a diverse group of organisms with both basal heterotrophic groups, including several parasitic lineages, and a phototrophic crown

radiation that includes common euphotic zone organisms like diatoms (Patterson, 1989; Leipe et al., 1996). The alveolates include familiar and well-characterized groups like dinoflagellates, ciliates, and parasitic apicomplexans, as well as the highly diverse, yet mostly uncharacterized, Marine Alveolate Groups (MALV). Marine fungi, understudied relative to their terrestrial relatives, reside down to the deepest depths of the ocean (Nagahama et al., 2001) and may live in commensal, mutualistic, or parasitic symbioses or as saprotrophs decomposing dead organic matter (Raghukumar, 2017). The majority of described marine fungal species belong to the phylum Ascomycota, although this may reflect a historical focus on euphotic and mesopelagic zones, as the phyla Basidiomycota, Blastocladiomycota, Chytridiomycota, and Mucoromycota also contain marine representatives (Vargas-Gastelum and Riquelme, 2020). Culture-independent studies suggest the presence of diverse novel fungal taxa in the deep sea (Nagano and Nagahama, 2012; Richards et al., 2012; Grossart et al., 2016), which are increasingly of interest as a potential source of bioactive compounds for use in medicines and pharmaceuticals (Zain Ul Arifeen et al., 2019).

Studies from the Malaspina expedition have identified various taxonomic groups within these four eukaryotic supergroups that exhibit relative dominance of DNA sequence data from different depth ranges. Groups with increasing relative abundance from the photic zone through the bathypelagic include Chrysophyceae and Bicosoecida (Stramenopila) and Group B radiolarians (RAD-B, Rhizaria) (Giner et al., 2020). Groups with high relative abundance in the mesopelagic (200-1000m) include the Labyrinthulomycetes (Stramenopila), Group C radiolarians (RAD-C, Rhizaria), and Marine Alveolate group IV (MALV-IV, Alveolata) (Giner et al., 2020). Groups with the highest relative abundance between 3000-4000m were Collodaria and MALV-II, followed by Basidiomycota (Fungi) (Pernice et al., 2016).

These dominant groups are a mix of moderately well characterized and relatively uncharacterized, yet ubiquitous, taxonomic groups. The order Collodaria and the RAD-A,B,C groups fall within the Radiolaria. Collodaria are heterotrophs that can exist either as solitary cells or as large colonies with thousands of cells embedded in a gelatinous matrix (Biard et al., 2017). They frequently host photosynthetic symbionts in surface waters (Stoecker et al., 2009), but little is known about their lifestyles in the deep ocean, particularly in relation to alternative symbioses that may form in the absence of sunlight. The RAD clusters contain environmental sequences about which little is known other than that they appear frequently in public databases as a result of culture-independent surveys, many of which remain unpublished. RAD cluster radiolarians have been reported from globally distributed deep and shallow marine settings including some chemosynthesis-based ecosystems such as hydrothermal vents and hydrocarbon seeps. The order Syndiniales is phylogenetically diverse, with two clusters of ubiquitous taxa (MALV I and II) and three clusters that are encountered less frequently (MALV III-V). Syndiniales have been found in a wide variety of marine and freshwater habitats, have few characterized representatives, and the few characterized species exist exclusively as parasites on a variety of host species, from other protists to fish (Groisillier et al., 2006; Guillou et al., 2008). Chrysophyceae belong to the photosynthetic crown within the Stramenopila, their presence in deep waters likely the result of post-bloom settling, while Bicosoecida and Labyrinthulomycetes fall into the basal heterotrophic stramenopile branches. Labyrinthulomycetes are exclusively marine and may act as decomposers of sinking aggregates formed after diatom blooms (Raghukumar, 2002), which would fit with their prominence in mesopelagic zones, or live as commensals or mutualists in tissues of marine invertebrates (Raghukumar, 2002). Bicosoecids contain

marine and freshwater species and include the highly abundant genus *Cafeteria*, a bacterial grazer found globally in surface and deep habitats (Schoenle et al., 2020). Finally, the Basidiomycota, which hosts a wide variety of single-celled yeasts and multi-cellular fungi, have few characterized marine species with most coming from shallow coastal systems (Jones and Fell, 2012). However, the findings from the Malaspina expedition suggest increased efforts to study basidiomycetes in the deep ocean may be warranted.

There are also growing efforts to define the functional roles and interactions of microeukaryotes beyond involvement in the classical concept of the microbial loop (Pomeroy, 1974; Azam et al., 1983), which focuses on upper ocean cycling of organic matter and generalizes involvement by classes of organisms (e.g., ciliates, flagellates, etc.). More recent studies are attempting to unravel the complex of interactions involving microeukaryotes at finer scales (i.e., specific taxonomic clusters) to develop models of global ocean interactomes (Lima-Mendez et al., 2015; Bjorbaekmo et al., 2020). Given the enormous size of the global deep ocean biome, these improved descriptions of the diversity and functions of microeukaryotes will certainly benefit our understanding of the global cycling of organic matter and other biogeochemical cycles (Gooday et al., 2020).

The potential involvement of microeukaryotes in deep-sea hydrothermal vent ecosystems has garnered limited attention, but similar to growing interest in characterizing deep-sea microeukaryotes in general, research addressing the unknowns surrounding vent microeukaryotes is beginning to gain momentum. Early studies of microeukaryotes in vent habitats identified various amoebae, flagellates, and ciliates from faunal assemblages and other unspecified vent habitats (Small and Gross, 1985; Atkins et al., 2000) using morphology and culturing techniques. Many were related to ubiquitous marine species

(Atkins et al., 2000), but some cultured organisms were able to adapt to high sulfide and metal concentrations (Atkins et al., 2002) suggesting potential roles in hydrothermal ecosystems. Application of culture-independent methods expanded the known diversity of microeukaryotes in hydrothermal systems to include representatives from most, if not all, major non-photosynthetic clades (Edgcomb et al., 2002; López-García et al., 2003; López-García et al., 2007; Sauvadet et al., 2010; Coyne et al., 2013). However, these studies involved small numbers of samples and low throughput sequencing methods, thus providing little information about ranges or habitat preferences of these organisms. Furthermore, the aforementioned limited information about the broader deep-sea microeukaryote community available at the time of these studies complicated interpretations of potential endemism among species sampled in vent habitats. More recent studies, including those presented here, have begun to delve deeper into hydrothermal vent microeukaryote diversity and habitat preferences using high-throughput marker gene sequencing, and to address the contributions of protists to carbon cycling in vent environments using grazing experiments (Murdock and Juniper, 2019; Pasulka et al., 2019; Hu et al., 2021; Murdock et al., 2021).

Details of the distribution and abundance of microeukaryotes among vent and non-vent habitats can provide information on habitat preferences and aid in developing hypotheses about potential endemism and contributions to hydrothermal vent ecosystem function. Microeukaryotes in hydrothermal habitats may be vent specialists or deep-sea opportunists taking advantage of abundant prokaryotic food items. Given the variety of known microeukaryote lifestyles, from heterotrophy to parasitism (Worden et al., 2015), the distinction between endemic and opportunist may have considerable effects on the flow of

energy among vent surrounding deep-sea habitats. Vent endemic microeukaryotes may play important roles as primary consumers or cyclers of organic matter within hydrothermal systems, while opportunists may act as conduits for moving chemosynthetically-produced organic carbon into deep-sea food webs. Therefore, the ranges and habitat preferences of various microeukaryotes warrant description, as they can affect the reach of hydrothermal vent influence (Levin et al., 2016).

Since the first report of hydrothermal vent microeukaryotes in 1985, no studies have revisited faunal assemblages as potential free-living microeukaryote habitat. One study reported evidence of parasitic protists from above vent bivalve communities and within their pallial cavities (Sauvadet et al., 2010) but did not sample among the bivalves or from any associated hydrothermal habitats. The faunal density of many vent assemblages would suggest opportunities for protist grazing of microbial biofilms on animal and rock surfaces in addition to being a haven for parasitic protists (Moreira and López-García, 2003). For example, assemblages of the tubeworm *Ridgeia piscesae* and their associated fauna (Sarrazin and Juniper, 1999) are known to host a diverse community of non-endosymbiotic microbes (Forget and Juniper, 2013), suggesting an abundant food source for heterotrophic protists. However, the potential interactions between prokaryotes, microeukaryotes and fauna in *R. piscesae* assemblages have not been explored, nor have the overall microbial contributions of faunal assemblages been placed in the broader context of the taxonomic and functional diversity of hydrothermal vent ecosystems.

The view of faunal assemblages primarily as symbiont-dominated microbiomes (Dick, 2019) has overshadowed their potential as habitat for non-endosymbiotic microbes. Faunal assemblages are exposed to a combination of diffuse hydrothermal fluids and deep seawater,

each with its own microbial signature that likely influences microbial diversity in the assemblage. Mixing of these two fluid sources—reducing hydrothermal fluids and oxygenated seawater—also creates ideal reduction-oxidation gradients that fuel chemosynthetic primary production (Reysenbach et al., 2000; Sievert and Vetriani, 2012). Although studies have investigated microbes on external surfaces of individual species of shrimp (Polz and Cavanaugh, 1995; Zbinden et al., 2008) and siboglinid (Lopez-Garcia et al., 2002) and alvinellid worms (Prieur et al., 1990; Campbell et al., 2001; Alain et al., 2002; Pagé et al., 2004) using culturing and low-throughput sequencing, only one study to date has characterized the full non-endosymbiont microbial community within faunal assemblages using high-throughput sequencing (Forget and Juniper, 2013).

The siboglinid tubeworm *R. piscesae* provides structure and habitat for other invertebrates (Tunnicliffe and Cordes, 2020) and diverse autotrophic and heterotrophic microbes (Forget and Juniper, 2013) across a wide range of environmental conditions (Tunnicliffe et al., 2014). Differences in diffuse fluid flux and sulfide concentrations are reflected by distinct assemblages of associated faunal species (Sarrazin and Juniper, 1999; Sarrazin et al., 1999; Govenar et al., 2002). It is these variations in habitat conditions and *R. piscesae*-associated fauna that provide a unique opportunity to explore variation in microbial diversity in the assemblages relative to the associated diffuse fluids and deep seawater, while also evaluating variation in functional contributions and potential interactions between faunal and microbial species.

1.1 Research objectives

The objectives of this research were threefold. The first objective used phylogenetic markers

to discern the habitat preferences of hydrothermal vent microeukaryotes by assessing their distributions and abundances across a range of hydrothermally influenced and adjacent deep-sea habitats, thereby identifying potential vent endemics. The second objective, also employing phylogenetic markers, explored microbial diversity in all three domains (Bacteria, Archaea, and micro-Eukarya), in the stable habitat provided by foundation species assemblages. This part of the research evolved into a larger project identifying 'core communities' of likely interacting microbial and faunal taxa as a way to simplify characterization of diversity-function relationships in ecosystems. And finally, objective three applied metagenomics to evaluate the potential functional contributions of microbes in faunal assemblages beyond the typical view of faunal assemblages solely as endosymbiotic microbiomes.

1.2 Methodological approach

Methodological developments, particularly in the production and comparison of gene sequence data, constitute an important and original component of the research presented here. The subsections below and the graphical representation in Figure 1.1, highlight some of the key methodologies used to achieve the stated objectives. Detailed descriptions are in Chapters 2 to 4.

1.2.1 Sample collection

All samples were collected using remotely operated vehicles (ROVs) aboard oceanographic research vessels. Microorganisms from hydrothermal and background fluids were collected on filters either in-situ, using the Hot Fluid and Particle Sampler (Butterfield

et al., 2004), or onboard following recovery of bulk water samples. Non-endosymbiotic microorganisms associated with faunal assemblages were collected from tubeworm grabs that were brought to the surface in sealed boxes on the ROV. Two methods of microbial assemblage removal from faunal grab samples are described in Chapter 3, and microorganisms were then collected by serial filtration onto 20 and 0.2 μm pore size filters. Serial filtration aimed to separate microeukaryotes (>20 μm) from the pico/nano-sized (0.2-20 μm) eukaryotes and prokaryotes.

1.2.2 DNA extraction and sequencing

DNA was extracted from cells collected on filters and sent to core facilities for high-throughput sequencing. Taxonomic composition of samples was characterized using marker gene sequencing of 16S (Bacteria and Archaea) and 18S (Eukarya) rRNA genes on Illumina MiSeq targeting variable regions V1-V3 for Bacteria, V6-V8 for Archaea, and V4 for Eukarya. Functional capacity of the microbial communities was characterized using shotgun metagenomic sequencing of twenty DNA extracts spread over three separate Illumina NextSeq runs to obtain reasonably good sequencing coverage of highly diverse microbial assemblages.

1.2.3 Microbial assemblage comparisons

After data quality control, marker genes were binned into operational taxonomic units (OTU) and taxonomically identified by comparison with the Silva (Bacteria and Archaea) and PR2 (Eukarya) reference databases. Metagenomes were assembled and annotated with genes identified using the Kyoto Encyclopedia of Genes and Genomes (KEGG) Orthology terms. Samples were compared based on their OTU (Chapters 2 and 3) or gene (Chapter 4)

compositions.

Between Chapters 2 and 3, there is a methodological shift that moves to recognize the compositional nature of high-throughput sequence data (Gloor et al., 2017). This represents a major realignment of thinking and methodologies in microbial ecology and although necessary, it complicates comparisons with previous published studies that did not treat sequence data as compositional.

1.2.4 Enrichment of genes and taxa

Major contributors to significant differences between sample types in Chapters 3 and 4 were identified using differential relative abundance (i.e., enrichment) of OTUs (taxa) or genes in one sample type over another. Differences between within-group and between-group variances for individual taxa or genes are given as "effect sizes"— the magnitude of the effect indicates the strength of enrichment in that group. These are referred to as enrichment scores in Chapters 3 and 4.

1.2.5 Quantitative PCR and congruence testing among size classes

In Chapter 3, evaluating the level of congruent variation in taxonomic composition among macro, meio, and micro size classes across samples from a range of tubeworm habitats required a balanced microbial assemblage created from bacterial, archaeal, and microeukaryal 16S and 18S rDNA sequences. The relative contributions of the three microbial domains were determined by quantitative PCR (qPCR) and used to scale the OTUs from each individually sequenced domain.

1.2.6 Co-occurrence and covariance network analyses

Patterns of species co-occurrence and covariance were used to infer niche preferences of microeukaryotes and potential species interactions. The analysis of co-occurrences between prokaryotes with known extreme habitat niche preferences and relatively uncharacterized microeukaryotes represents a novel approach developed to infer adaptations among the latter to the extreme conditions (e.g., high temperature, low oxygen, acidity) associated with hydrothermal vent ecosystems. In Chapter 3, potential interspecies interactions between microbes and fauna were inferred using proportionality (covariance), a compositional data alternative to co-occurrence (Lovell et al., 2015). Network mapping was used to visualize the complex of pairwise associations and identify structure within sampled assemblages.

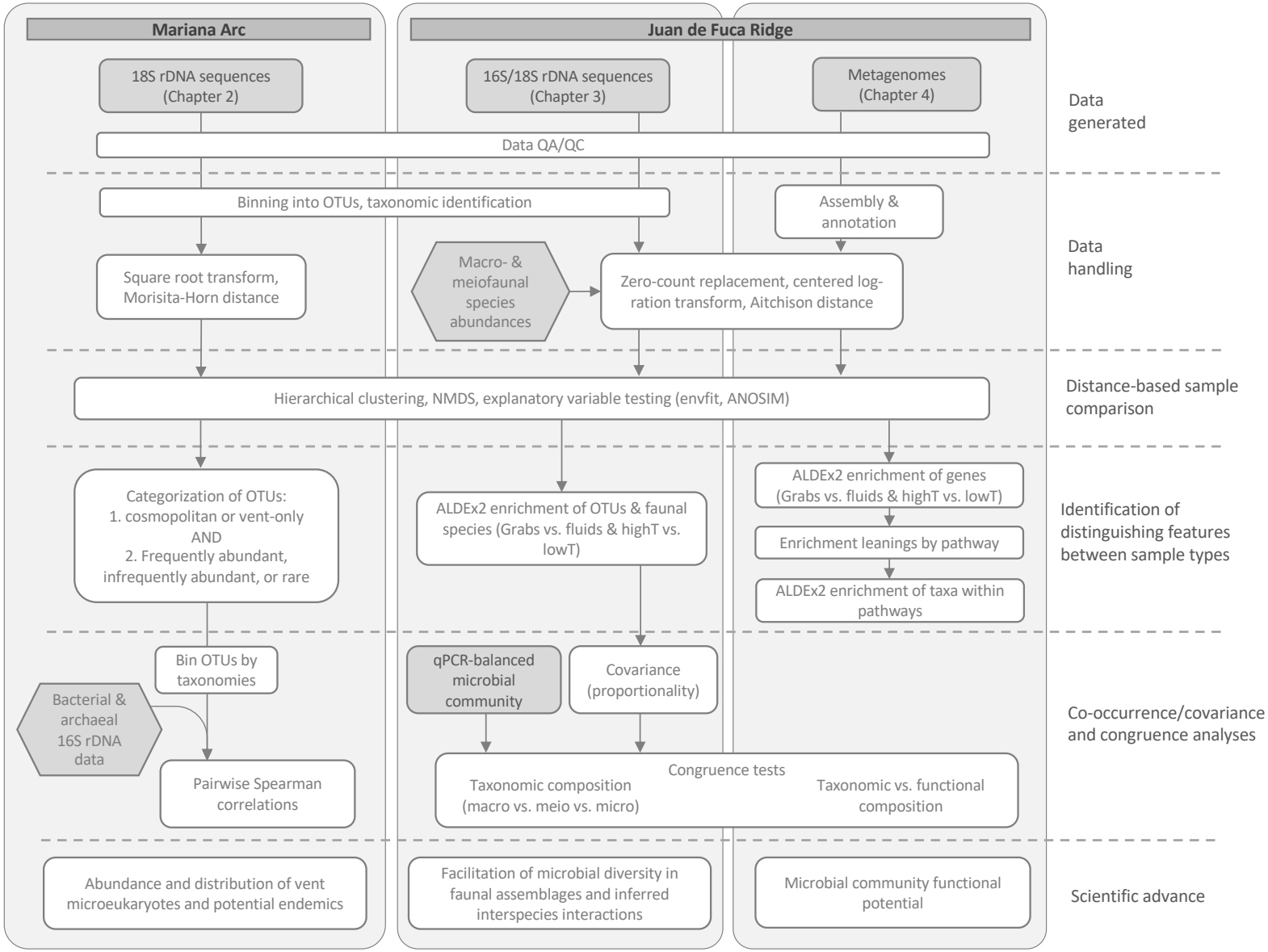
1.2.7 Compositional methods development for metagenomic data

Comparative metagenomic analyses have been slow to adopt compositional data approaches. Therefore, new methods were developed to identify distinguishing functional features between habitat and community types. A new metric of "enrichment leanings" was developed and tested for prioritization of functional pathways for in-depth analysis.

Variations in structure within the following chapters reflect section ordering preferences of the journal to which it was submitted for publication.

Figure 1.1 Graphical overview of methods

Data in grey polygons were publicly available or provided by collaborators. Data in grey rounded rectangles are produced by the author.



Chapter 2. Hydrothermal vent protistan distribution along the Mariana arc suggests vent endemics may be rare and novel

Chapter 2 has been published in the peer-reviewed journal *Environmental Microbiology*: Murdock S and SK Juniper (2019). Hydrothermal vent protistan distribution along the Mariana Arc suggests vent endemics may be rare and novel. *Environmental Microbiology* 21(10):3796-3815.

All reported data, analyses, figures and interpretations were created or performed by S Murdock, with editorial contributions on the manuscript from SK Juniper.

Abstract

Elucidation of the potential roles of single-celled Eukaryotes (protists) in ecosystem function and trophodynamics in hydrothermal vent ecosystems is reliant on information regarding their abundance, distribution, and preference for vent habitats. Using high-throughput 18S rRNA gene sequencing on a diverse suite of hydrothermally-influenced and background water samples, we assess the diversity and distribution of protists and identify potential vent-endemics. We found that 95% of the recovered sequences belong to operational taxonomic units (OTUs) with a cosmopolitan distribution across vent and non-vent habitats. Analysis of 'vent only' OTUs found in more than one vent sample and co-occurrence network analysis comparing protist groups to extremophilic reference organisms suggest that the most likely vent endemics are infrequently encountered, potentially in low abundance, and

belong to novel lineages, both at the phylum level and within defined clades of Rhizaria and Stramenopila. Potential endemism is inferred for relatives of known apusomonads, excavates, and some clades of Syndiniales. Similarity in community composition among samples was low, indicating a strong stochastic component to protist community assembly and suggesting that rare endemics may serve as a reservoir poised to respond to changing environmental conditions in these dynamic systems.

2.1 Introduction

There is a disproportionate focus in microbial ecology that favors prokaryotes and largely ignores the contribution of microbial (unicellular) eukaryotes to biodiversity and ecosystem functioning (Caron, 2009; Caron et al., 2009a). In marine systems, there is further bias towards studies of upper ocean photosynthetic micro-eukaryotes, while little is known about heterotrophic species that inhabit the deep, dark ocean interior (Herndl et al., 2008; Arístegui et al., 2009). This latter discrepancy is being remedied gradually through large initiatives such as *Tara* Oceans (Karsenti et al., 2011) and the Malaspina Expedition (Pernice et al., 2015; Pernice et al., 2016), both global sampling efforts characterizing microbial populations from sunlit to bathypelagic depths, inclusive of prokaryotic, eukaryotic, and viral members. While prokaryotes are acknowledged key players in global biogeochemical cycles (Cho and Azam, 1988; Arrigo, 2005; Falkowski et al., 2008) and chemosynthetic producers of organic carbon below the euphotic zone (Herndl et al., 2005; Arístegui et al., 2009), much less is known about the role microbial eukaryotes (protists) play in the trophic transfer of organic carbon, particularly in the deep ocean.

Chemolithoautotrophic (chemosynthetic) primary production in the deep ocean is

most intense in areas around hydrothermal vents, where it is supported by reducing substances formed during hydrothermal circulation through crustal rocks beneath the seafloor (Jannasch and Wirsen, 1979; Karl et al., 1980; Reysenbach and Shock, 2002; Amend et al., 2011; Olins et al., 2017). Global primary production by prokaryotes in subseafloor hydrothermal systems is estimated to be roughly 1.4 Tg C y^{-1} (McNichol et al., 2018), resulting in hot spots of labile carbon availability around areas of active hydrothermal discharge from the subseafloor. There is additional chemosynthetic productivity by free-living and symbiotic prokaryotes around vent openings. This intense productivity supports a diverse community of microbial heterotrophs that includes protists, about which little is known.

Early studies of protist diversity at deep-sea hydrothermal vents using microscopy and culturing methods reported species of ciliates, flagellates, and amoeboid forms that were a mix of known ubiquitous marine organisms and unknown relatives of described species (Small and Gross, 1985; Atkins et al., 2000). A few isolates of ubiquitous marine flagellates were found to survive exposure to high levels of sulfide and metals (Atkins et al., 2002) suggesting tolerance by some species of the extreme conditions commonly encountered in vent environments, and the potential for contribution to vent food webs. More recent culture-independent studies using the 18S rRNA gene have revealed a complex protist community, with unexpectedly high levels of diversity and taxonomic novelty, at hydrothermal vents at Guaymas Basin (Edgcomb et al., 2002; Coyne et al., 2013), the Mid-Atlantic Ridge (López-García et al., 2003; López-García et al., 2007), and on the southern East Pacific Rise (Sauvadet et al., 2010). While the potential functional contributions of protists to these ecosystems remain unknown, so too does fundamental information on diversity and

distribution of species in various vent habitats.

Before delving into complex ecosystem-level questions about the functional roles of protists in hydrothermal systems it is necessary to understand their degree of tolerance or specialization for the extreme conditions found at vents. Are they endemics, adapted to thrive and compete in this extreme environment? Or are they transient opportunists taking advantage of the abundant biomass in an otherwise energy-starved deep ocean? Protist populations in general may be composed of a mixture of those with near-global distribution and those with a more endemic nature (Caron et al., 2009a). The balance of these two distribution patterns among heterotrophic protists associated with hydrothermal vents could have potential impacts both within and beyond the immediate area. For example, endemic consumers may limit the export of chemosynthetically based organic carbon to the wider deep ocean, whereas opportunists with a cosmopolitan distribution would act as conduits for carbon export to the deep sea.

To date, only one vent-endemic protist has been described, a sessile, colonial symbiont-bearing folliculinid ciliate that forms dense bright blue carpets at vent sites on the Juan de Fuca Ridge (Tunnicliffe et al., 1985; Kouris et al., 2007; Kouris et al., 2010). Additional hydrothermal vent endemism has been proposed based on previous culture-independent studies of protists but most of these studies made no comparisons with protist communities in nearby non-vent-influenced habitats. Sauvadet et al. (2010) conducted the most thorough sampling of protists in vent habitats to date but the single non-vent sample was collected over 1,000 kilometers away from their vent sites making it difficult to distinguish vent specialist protists from widely distributed deep-sea opportunists that may have been entrained from surrounding seawater during sampling.

In addition to comparing protists from vent and nearby non-vent habitats, associations with known vent indicator microorganisms can be used to help identify vent-specific protists. For example, among the bacteria, Epsilonbacteraeota (formerly Epsilonproteobacteria) are considered true endemic indicator organisms of hydrothermal habitats because of their frequent incidence and high abundance in almost every surveyed vent habitat and near absence from the surrounding deep seawater (Huber et al., 2003; Nakagawa et al., 2005; Campbell et al., 2006; Huber et al., 2007; Goffredi, 2010; Huber et al., 2010; Djurhuus et al., 2017). Protists with a similar distribution to the Epsilonbacteraeota across a wide range of vent and non-vent habitats could be inferred to have similar niche specificity.

Co-occurrence network analyses using pairwise correlations can serve as a tool for inferring similarities in niche preference (Steele et al., 2011) and are effective predictors of physical associations such as parasitism or commensalism (Lima-Mendez et al., 2015). Pairwise correlations between known and unknown taxa can help constrain the preferred habitat of the latter. Additionally, previous network analyses that also included environmental parameters found that some deterministic effects on microbial community members were identified, although links between microbial species were stronger than those with environmental variables (Steele et al., 2011) indicating that interspecies interactions were stronger deterministic factors on community composition than the environment.

In this study we evaluate the diversity of protists from a range of hydrothermal vent-influenced and nearby background seawater environments at nine locations along the 1100 km length of the Mariana arc, sampled during several research expeditions, using high-

throughput sequencing of the 18S rRNA gene. We use this broad dataset in combination with deep community sequencing to evaluate the potential specialization of various protist groups for life in vent habitats. We utilize the incidence and relative abundance of sequences from different protist taxa and co-occurrences of protists with prokaryotes and environmental data to 1) identify potential vent endemic and/or vent adapted protists, 2) explore biogeographic patterns in the composition of vent protist assemblages, and 3) identify deterministic factors that affect the incidence and relative abundance of different groups of protists. This information will aid in the advancement of hypotheses regarding the role of protists in vent ecosystem trophodynamics.

2.2 Results

2.2.1 Site descriptions

The geologic and volcanic settings of Nikko, NW Eifuku, Daikoku, East Diamante, and NW Rota-1 have been described in detail by Embley et al. (2006); Embley et al. (2007); Chadwick et al. (2008), Lupton et al. (2006) and references therein. Kasuga-2 has a shallow summit (200 m depth) and multiple hydrothermally active sites venting through sulfur-encrusted volcaniclastics (McMurtry et al., 1993; Fryer et al., 1997). In 2004, iron-oxyhydroxide crusts were observed over extensive areas of both Kasuga-2 and NW Eifuku suggesting availability of both reduced sulfur and iron as energy sources for chemosynthetic microbes at both sites. The Maug islands are remnants of a volcanic caldera wall with a shallow (30 m depth) central cone. Only low-level venting with small areas of fluffy yellow or white filamentous microbial mats were observed on the flanks of the cone during the 2004 expedition with fluid temperatures 4°C above ambient.

Forecast and Urashima lie in the southern part of the back-arc basin where it comes to within 20 km of the volcanic arc (Ishibashi and Urabe, 1995; Martinez et al., 2000; Nakamura et al., 2013). At Forecast we found chimneys venting moderately high temperature clear fluids (up to 195°C) as well as low temperature diffuse vents. Urashima, the deepest site we sampled (Table S1), hosts active chimneys that are 5-10 m in height, venting clear diffuse fluids through iron-oxyhydroxide rich structures as well as high temperature black smoke with inferred end-member fluid temperatures of 326°C (Nakamura et al., 2013).

2.2.2 DNA sequencing

DNA was extracted from 68 fluid samples covering a range of background and vent-influenced sample types (9 background, 46 diffuse vent fluid, 3 high-temperature diffuse, 4 diffuse/plume, 2 near-field, and 4 buoyant plume). Nine samples were removed from further analysis after unsuccessful attempts at amplification of 18S rRNA genes or insufficient sequencing returns (Table S2). After quality filtering, removal of chimeric sequences, and those identified as Metazoa, a total of 4,705,713 high-quality sequences of the 18S rRNA V4 region remained from 59 samples (Table S3). Eight samples were retained for analysis despite lesser numbers of sequences (Table S3, grey shading). Four of these samples produced very low numbers (<10,000) of usable protist sequences, well below the average return of around 80,000 sequences;— one was from a high temperature vent (155°C, FS430), one was collected immediately adjacent to the actively erupting site on NW Rota-1 (FS660) and the other two were from the Iceberg complex on NW Rota-1 (FS446 & J800-10), an area from which one of the original 68 samples was discarded because of failed attempts at amplification of the 18S rRNA gene (Table S2). The other four samples had few sequences

that were not classified as Metazoa (FS349, FS357, J798-02, & J799-33; Table S3).

As operational taxonomic units (OTUs) that are represented by a single sequence (singletons) are often experimental artifacts on some high-throughput sequencing platforms (Behnke et al., 2011), we adopted a conservative approach by removing sequences that appeared only once in the entire dataset (global singletons), leaving 13,433 non-singleton protist OTUs. Singleton OTUs in our study typically accounted for <1.5% of the sequences in any given sample, with a few exceptions where they made up as much as 5.5%. Overall removal of global singleton OTUs accounted for 0.7% of the total dataset.

2.2.3 Richness and diversity

OTU richness varied widely among samples (Table S3) and ranged between 2,428 OTUs (sample FS473) and 62 OTUs (sample FS480). Sample FS473 was collected in the NE pit crater of Daikoku where intense venting was evident from a dense smoky plume that filled the crater and precluded sampling at the source of venting. This sample is therefore termed "diffuse/plume" indicating the likely influences of both intense hydrothermal venting and dilution with the surrounding seawater. Sample FS480 was collected in an area characterized by abundant tubeworm communities and microbial mats on Nikko volcano. For the five sites from which we collected both vent and background samples (Nikko, NW Eifuku, Daikoku, NW Rota-1, and Forecast), the combined OTU richness in vent samples was higher than in background samples, with the exception of Nikko where the background sample had higher richness. The combined richness from a site was strongly correlated with the number of samples analyzed from it (r = 0.93, R^2 = 0.87, p<0.001).

Diversity estimates using the Shannon and Simpson (reported as 1/Simpson) indices

before and after removal of singleton OTUs were widely ranging across the 59 samples (Table S3). Values with singletons removed ranged from 2.2 to 5.3 for Shannon diversity and 4.8 to 57.6 for the inverse Simpson measure. Values with singletons included were nearly identical with slightly higher maximum values and are not discussed further. Shannon diversity values with samples normalized to n=2000 sequences (which excluded three samples) varied by ±0.2 from those using the full dataset of nonsingleton OTUs and caused only minor changes to the ordering of samples by diversity, mostly within the intermediate diversity samples. Therefore, for consistency with the rest of the study, diversity values reported in table S3 are calculated based on the 13,433 nonsingleton OTUs.

Both indices found the highest diversity in the lowest temperature diffuse fluid sample from the Iron599 site on NW Rota-1 (FS653) followed closely by the background sample taken near Nikko (FS482), and the lowest diversity was from the Tubeworm Hangover site at Nikko (FS480) that also had the lowest OTU richness. The ordering of samples by diversity, from highest to lowest, was otherwise variable between the two indices. Similar to the combined OTU richness from a site, the combined diversity from a site was higher in the vent samples than the background, except at Nikko. There were no correlations between Shannon or Simpson diversity estimates and sample temperature or depth, and analysis of variance did not indicate significant differences in diversity tied to location (site) or sample type.

2.2.4 Incidence and relative abundance of OTUs

Table 2.1 summarizes the incidence and relative abundance of OTUs encountered in this study. Relative abundance as we use it here refers to the frequency of encountering sequences belonging to individual OTUs within a single sample, which may or may not be

indicative of the relative abundance of the organism in nature. Caveats to the usage of relative abundance on PCR-based studies are addressed in the *Discussion section*. We used the conventions of 'ubiquitous', 'common', and 'rare' to indicate the incidence of individual OTUs across the sample set, defined respectively as presence in 30 or more samples, 6-29 samples, and 5 or fewer samples. In addition to presence in 30 or more samples, OTUs labeled as ubiquitous also had to be present in at least 3 of the 5 sample years and at both the northern and southern arc sites. Relative abundance of OTUs was then evaluated with respect to a threshold level of 1% relative abundance (within individual samples). We defined 'frequently abundant', 'infrequently abundant', and 'never abundant' as abundant (>1%) in 50% or more of the samples in which it is found, abundant in less than 50% of the samples in which it is found, and never present at 1% or greater in any sample, respectively. Because the term 'rare' can be used to mean both uncommonly encountered as well as low relative abundance, we feel it is important to explicitly define its usage in this study.

Based on the above criteria, 107 of the 13,433 OTUs (0.8%) were classified as ubiquitous (Table 2.1), accounting for 67.2% of the sequences in the study. Ubiquitous OTUs belonged primarily to the Syndiniales (marine alveolate groups I-IV, MALV) and radiolaria but also included some haptophytes, cercozoa, and an OTU belonging to the recently described phylum Picozoa (Seenivasan et al., 2013). Of the 107 ubiquitous OTUs, 73 occurred in high relative abundance within some of the samples, but increased relative abundance was not linked to any particular sample type or site. The OTU with the highest frequency of abundance (present in all 59 samples and abundant in 53) was identified as a member of the Syndiniales (MALV-I).

Table 2.1 Incidence and frequency of abundance of OTUs.

	Ubiquitous (30+ samples)		Common (6-29 samples)		Rare (<6 samples)	
	No. of OTUs (% OTUs)	No. of reads (% reads)	No. of OTUs (% OTUs)	No. of reads (% reads)	No. of OTUs (% OTUs)	No. of reads (% reads)
Frequently abundant [a]	14 (0.1%)	1737391 (36.9%)	3 (<0.1%)	103820 (2.2%)	13 (<0.1%)	43985 (1.0%)
Infrequently abundant [b]	59 (0.4%)	1327328 (28.2%)	156 (1.1%)	747016 (15.9%)	37 (0.3%)	61385 (1.3%)
Never abundant [c]	34 (0.3%)	98380 (2.1%)	1204 (8.7%)	378930 (8.0%)	11913 (89.0%)	207478 (4.4%)
Totals =	**107 (0.8%)**	**3163099 (67.2%)**	**1363 (10.1%)**	**1229766 (26.1%)**	**11963 (89.1%)**	**312848 (6.7%)**

[a] >1% relative abundance in >50% of samples in which it occurs.
[b] >1% relative abundance in <50% of samples in which it occurs.
[c] Never >1% relative abundance.

Only 282 OTUs (2% of total) reached the 1% relative abundance threshold, but these OTUs made up 85% of the sequences (Table 2.1). Each sample contained between 12 and 32 abundant OTUs, such that, on average, 8% of the OTUs accounted for 69% of the sequences in a sample. Syndiniales made up about half of the 282 OTUs that reached the abundance threshold, some representing as much as 39% of the sequences from a sample, followed by radiolaria and dinoflagellates. There were also 8 frequently abundant and 15 infrequently abundant OTUs that could not be identified beyond the domain Eukarya. The majority of OTUs (98%) could never be classified as abundant, regardless of the frequency with which they were found in our 59 samples. Additionally, most OTUs were rarely encountered (11,963 rare OTUs, Table 2.1) with only 50 of these contributing to the 282 that reached the 1% threshold of abundance. Nearly half of rare OTUs (6,448) were present in only one sample.

2.2.5 Distribution of OTUs across sample types and locations

Of the 13,433 OTUs, 10.6% (1,429) were found only in background samples (not in any hydrothermally-influenced sample), 31.1% (4,178) had a cosmopolitan distribution (i.e. present in both background and hydrothermally-influenced samples), and 58.3% (7,826)

were found only in samples with some hydrothermal influence ('vent-only', never in any background sample) (Figure 2.1a). The majority of the vent-only OTUs (7,774) fell into the 'never abundant' category discussed in the previous section, with the remaining 52 OTUs being abundant in three or fewer samples. The other 230 of 282 OTUs that reached the 1% relative abundance threshold had a cosmopolitan distribution. Incidence of vent-only OTUs was either common or rare with none found in more than 15 samples and most in 3 or fewer vent-influenced samples.

The taxonomic composition of OTUs specific to hydrothermally-influenced samples differed slightly from the combined cosmopolitan and background OTUs, hereafter simply referred to as cosmopolitan OTUs (Figure 2.1b,c). While Alveolates dominated both vent-only and cosmopolitan numbers of OTUs, they made up a smaller proportion of OTUs in the vent samples (56%, compared to 71% of cosmopolitan OTUs). This difference was primarily attributed to a large decrease in the number of OTUs belonging to the Syndiniales in vent samples, which was accompanied by a small increase in OTUs with uncertain placement within the alveolates. The second most speciose group in both vent-only and cosmopolitan categories were the Rhizaria, which had slightly greater representation in vent-only (21%) compared to cosmopolitan (18%) OTUs, primarily in the polycystine radiolaria. A final notable difference between the two groups of OTUs was the proportion of OTUs that could not be classified beyond the domain Eukarya (Figure 2.1b,c). This group made up the third largest proportion of both vent-only (18%) and cosmopolitan (7%) OTUs. Vent-only OTUs also included members of the Amoebozoa, Apusozoa and Excavata that were absent from cosmopolitan OTUs.

The taxonomic depth to which sequences could be assigned was quite shallow

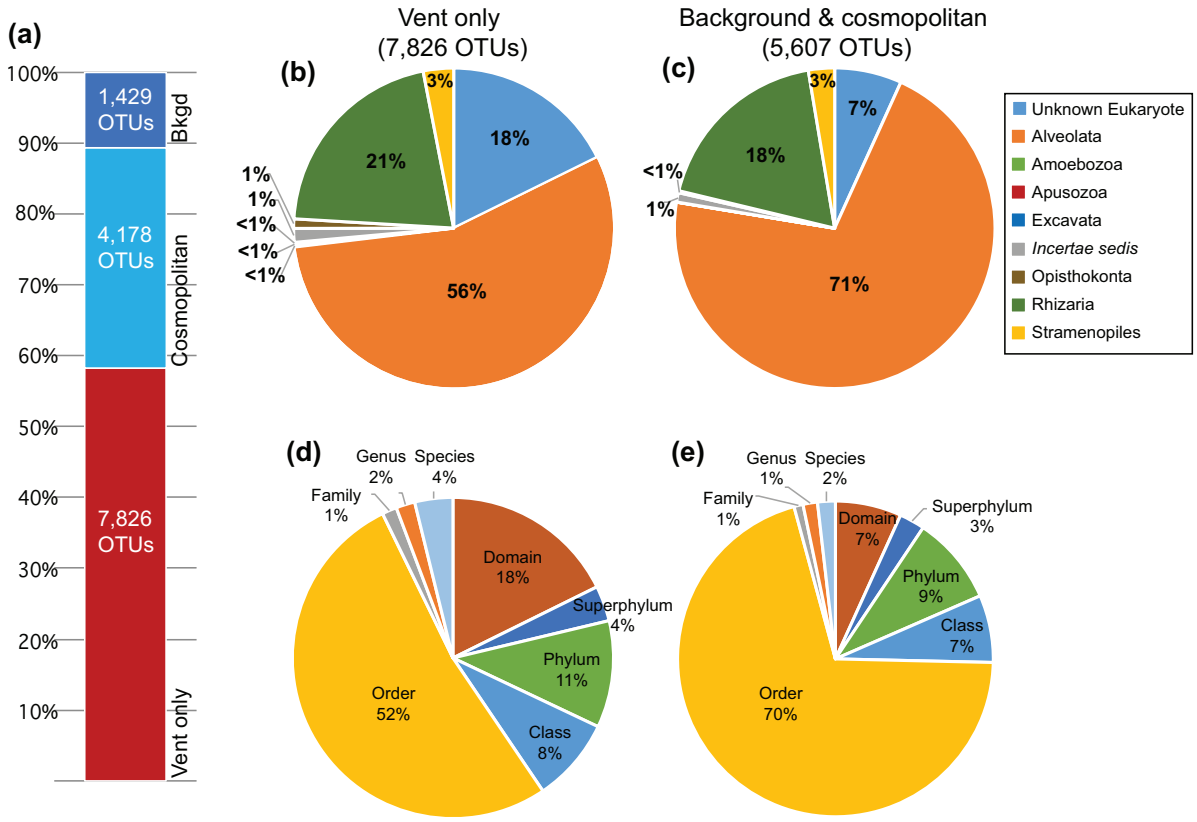


Figure 2.1 Distribution and classification of OTUs.

a) Relative proportions of OTUs found only in vent samples (red), in vent and background (cosmopolitan, light blue), and only in the background (dark blue). **b) and c) T**axonomic breakdown of vent-only OTUs (b) and background and cosmopolitan OTUs combined (c). The group *Incertae sedis* includes the Centroheliozoa, Haptophyta, Picozoa, and Telonema. **d) and e)** Depth to which vent-only (d) and background and cosmopolitan (e) OTUs could be classified using the PR^2 database.

across the entire dataset (Figure 2.1d,e) and tended to be shallower for vent-only OTUs. OTUs classified to the level of order made up the bulk of cosmopolitan and vent OTUs, at 70% and 52% respectively. Vent-only OTUs were more frequently identified to the level of class, phylum, superphylum, and domain, along with small increases in OTUs identified to genus and species level.

While the few ubiquitous and frequently abundant OTUs, all of which had a cosmopolitan distribution, provided a strong basis of similarity among samples, this was balanced by numerous rare OTUs that were unique to individual samples and sites. Cosmopolitan OTUs made up less than half of total OTU richness in this study but accounted for >96% of the sequences from each sample. The remaining sequences in a sample (<4%) belonged to vent-only OTUs. OTUs that were unique-to-site varied between 10.6% of OTUs at Forecast and 58.6% of OTUs at NW Rota-1 (Table 2.2) with an average of 23% of OTUs that are unique to a site. OTUs that were unique-to-sample made up 12 ± 7% of OTUs in a sample (not shown). Samples from East Diamante had the highest unique-to-sample average of 18% and samples from Forecast had the lowest (7%). NW Rota-1 had the highest percentage of vent-only OTUs (55%) followed closely by Daikoku and East Diamante with 45% and 44%, respectively (Table 2.2).

Table 2.2 Total numbers of OTUs identified at each site.

Site	Number of samples	Total number of OTUs	*Unique to site (% of site OTUs)	‡Vent-influenced OTUs (% of site OTUs)
Nikko	4	2243	486 (21.7%)	477 (21.3%)
Kasuga-2	1	770	88 (11.4%)	216 (28.1%)
Daikoku	5	4150	1206 (29.1%)	1888 (45.5%)
NW Eifuku	12	3226	820 (25.4%)	973 (30.2%)
Maug	1	188	29 (15.4%)	73 (38.8%)
E Diamante	2	1272	226 (17.8%)	555 (43.6%)
NW Rota-1	26	9667	5667 (58.6%)	5324 (55.1%)
Forecast	5	615	65 (10.6%)	167 (27.2%)
Urashima	3	430	59 (13.7%)	152 (35.3%)

*Includes OTUs of all sample types not found in samples at/near other sites.

‡Includes OTUs tagged as 'vent only' (i.e. absent from all background samples) that may be found at any number of the sampled sites.

Comparison of community composition by Nonmetric Multidimensional Scaling (NMDS) and hierarchical cluster analysis showed some grouping of samples from NW Rota-1 and NW Eifuku but with considerable overlap between them, as well as outliers from both sites that did not group with the rest (Figure 2.2). Samples from Daikoku grouped close together, but samples from most other sites did not form cohesive clusters. The stress of the 2-dimensional NMDS was high (0.20) but decreased to 0.14 when represented in three dimensions (not shown). Neither visualization method grouped samples in distinct clusters based on sample type or site. However, one-way ANOSIM results showed significant differences in sample composition between sites (global R value=0.40, $p<0.001$) with the highest degree of dissimilarity between samples from NW Rota-1 and Urashima (R=0.85) followed by Daikoku and Forecast (R=0.77) and Urashima and NW Eifuku (R=0.76). NW Rota-1 also differed significantly but to a lesser degree with Forecast (R=0.51) and NW Eifuku (R=0.26). ANOSIM results based on sample type did not identify any significant differences. NMDS was also performed using only abundant OTUs (>1% relative abundance) as well as with vent-only OTUs. The 2-dimensional stress increased slightly when using only the abundant OTUs, and no solution could be reached in 1000 iterations using vent-only OTUs (not shown).

2.2.6 Comparison of close proximity sample pairs

Small-scale protist community heterogeneity was assessed by comparison of the protist community compositions in two pairs of diffuse fluid samples that were each collected 11 meters apart on NW Rota-1 volcano, one pair in 2009 and one in 2010. These close proximity sample pairs had highly dissimilar populations despite similar habitat appearance and

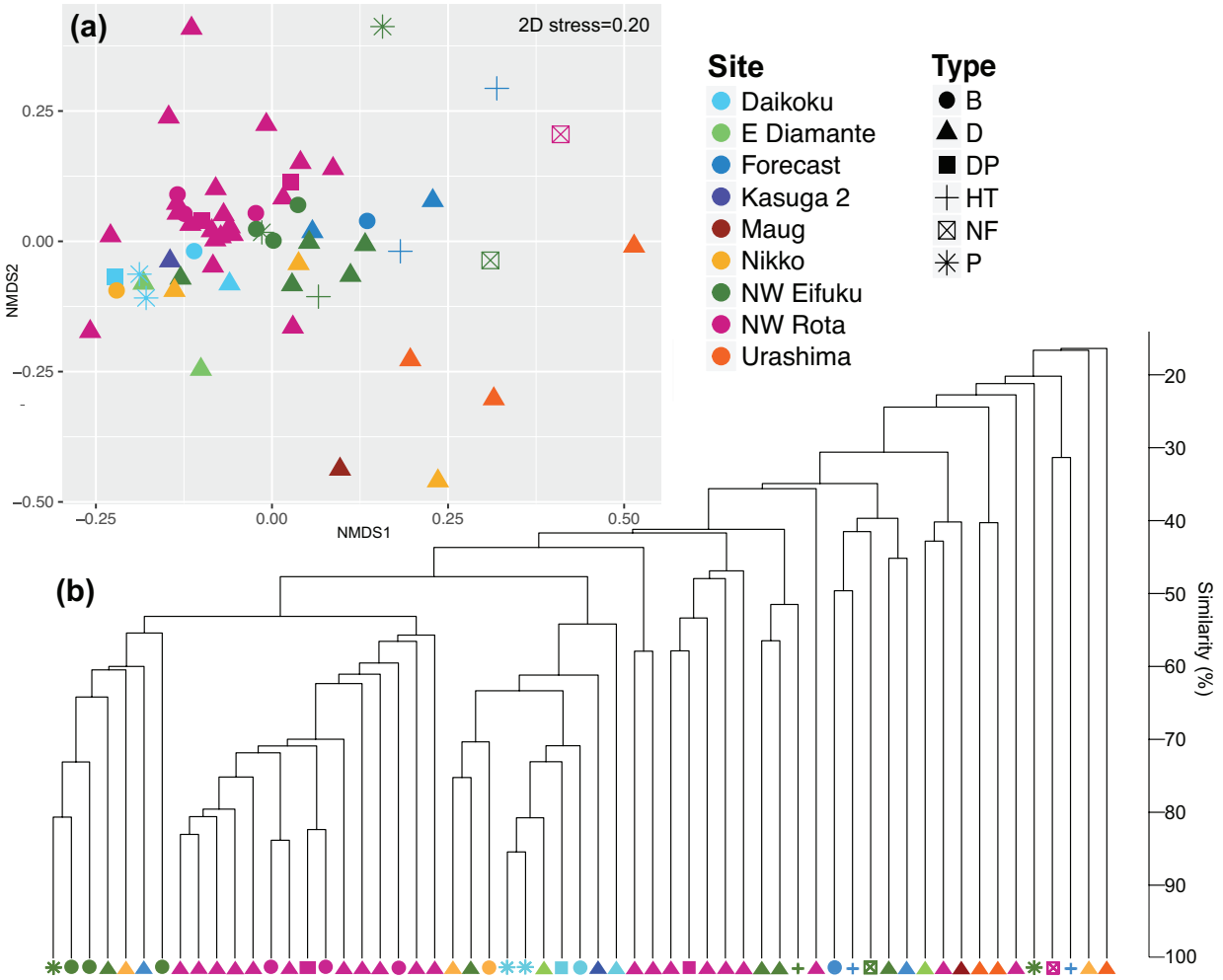


Figure 2.2 Relatedness patterns of the Mariana Arc samples.

Relatedness based on the Morisita-Horn distances of the square root transformed abundance matrix for all nonsingleton protist OTUs shown as **a)** a Nonmetric Multidimensional Scaling plot and **b)** a hierarchical cluster diagram**.** Sample types: B=background, D=diffuse, DP=diffuse/plume, HT=high-temperature, NF=near-field, P=plume.

comparable temperatures. Samples FS741 and FS753, both from the Iceberg complex, had measured temperatures of 33.3 and 20.7°C, respectively, and samples FS646 and FS664 had nearly identical temperatures around 46°C. All four samples were collected from diffuse fluids venting through cracks in the seafloor in areas characterized by thick microbial biofilms that coated rock surfaces and appeared as flocculent material in the surrounding water. FS741 and FS753 shared 276 OTUs (41% of FS741 OTUs, 26% of FS753 OTUs) and FS646 and FS664 shared 87 OTUs (9% of FS646 OTUs, 25% of FS664 OTUs). Most of the shared OTUs were categorized as ubiquitous or common and had a cosmopolitan distribution. Only 11 out of 276 and 7 out of 87 shared OTUs were classified as 'vent only'.

In addition to having few OTUs in common, those that were shared had differential abundance of sequences from the two samples (Figure S3). OTUs belonging to the haptophyte order Prymnesiales, MALV-I (Syndiniales), and the radiolarian order Spumellaria made up a greater proportion of sequences from FS646, while OTUs identified as unknown alveolates, RAD-B cluster radiolarians, and MALV-II (Syndiniales) were more abundant in FS664. Similarly, shared OTUs belonging to MALV-II, the radiolarian order Nassellaria, and unknown dinoflagellates have greater representation in sample FS741, with MALV-I dominating shared OTUs in sample FS753.

2.2.7 Drivers of protistan diversity

Multivariate statistical analyses applied to the 2006 subset of protist diversity and physiochemical data used by Huber et al. (2010) revealed no significant correlations by any of the three methods applied. No single variable was able to explain more than 19% of the variability in community composition of the samples in either the NMDS or CCA plots (data

not shown) and no significant parameters were identified by PERMANOVA. Pairwise co-occurrences between OTUs (grouped by similar taxonomic identities) and environmental parameters revealed mostly negative relationships (Table S4) with some exceptions: positive correlations occurred between concentrations of hydrogen gas and polycystine radiolaria of the order Nassellaria and the stramenopile genus *Pirsonia*, a few MALV clades with sulfate and pH, the fungal genus *Trichosporon* with depth, and unknown Eukaryotes were positively correlated with concentrations of iron.

Co-occurrence network visualizations were used to evaluate the overall structure of the microbial community and pairwise Spearman correlations between protist and prokaryote OTUs from Huber et al. (2010), with known preference for extremes, were used to infer groups of protists with potential adaptation to the extreme conditions found at vents. Structurally, several protist nodes (taxa) aggregated together to form one visually cohesive cluster (Figure 2.3a) that included many of the archaeal taxa, with scattered distribution of the remaining protist nodes. Although Markov clustering identified two overlapping clusters within this aggregation, the protistan taxa included had consistent positive correlations with reference taxa of known thermophiles and hyperthermophiles, anaerobic methanogens, and novel archaeal groups belonging to the candidate phyla Lokiarchaeota, Bathyarchaeota, and Geothermarchaeota, so far associated with deep sediment and crustal subsurface habitats (Meng et al., 2014; Spang et al., 2015; Jungbluth et al., 2017) (Figure 2.3b). Each of the protist taxa within these clusters had low relative abundance (indicated by the size of the node in Figure 2.3), while the five protist groups with high relative abundance, which collectively accounted for 60% of protist sequences, were all located on the periphery of the network. Protist taxa within the aggregation belonged to every eukaryotic supergroup included in the

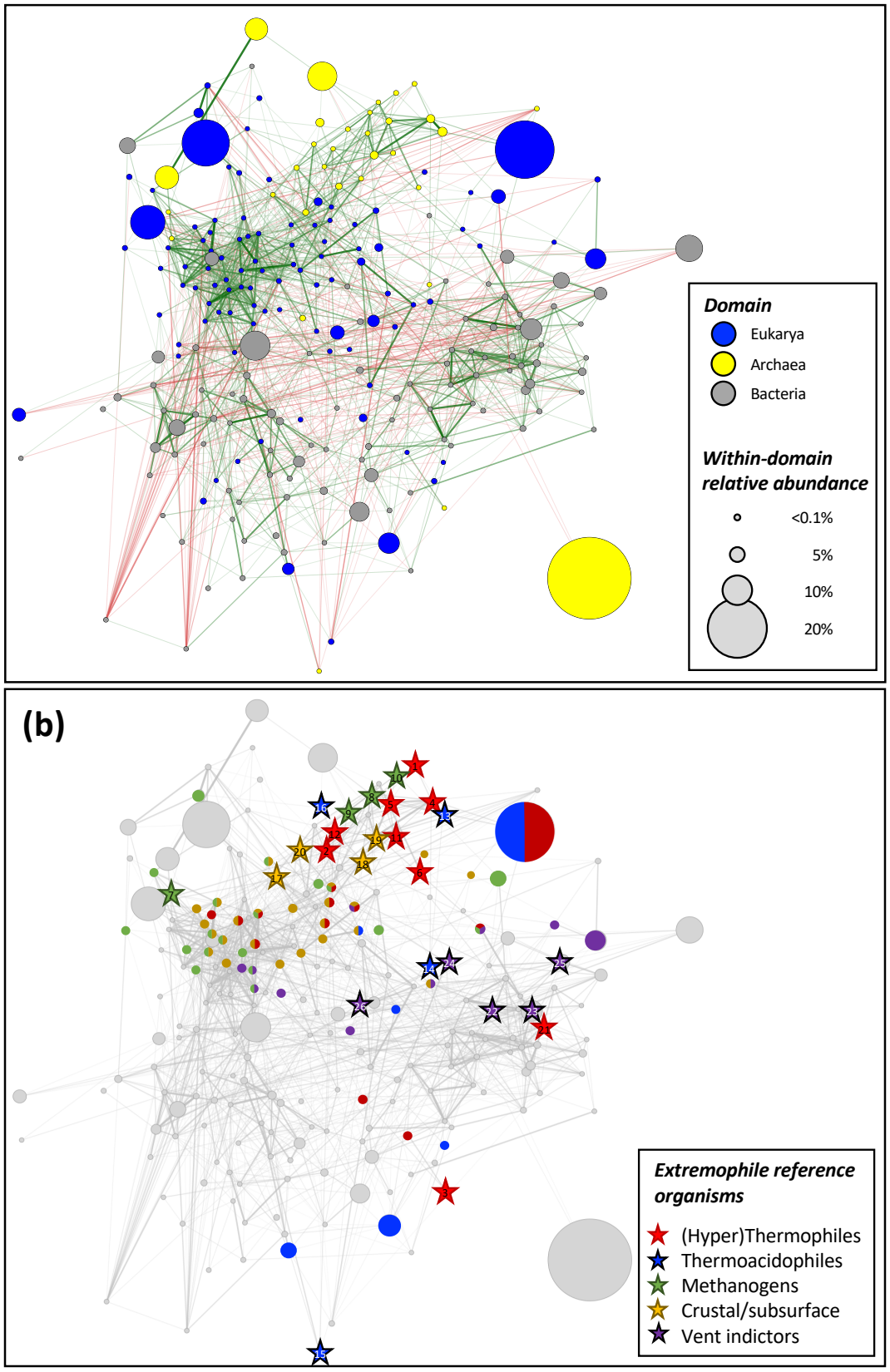

Domain
Eukarya
Archaea
Bacteria
Within-domain relative abundance
<0.1%
5%
10%
20%
(b)
Extremophile reference organisms
(Hyper)Thermophiles
Thermoacidophiles
Methanogens
Crustal/subsurface
Vent indictors

Figure 2.3 Co-occurrence network

Network diagram showing a) positive relationships (green lines) and negative relationships (red lines) between protists, Bacteria, and Archaea in the Mariana Arc region. Line (edge) thickness indicates the strength of the relationship. Size of the circles (nodes) indicates the relative contribution of the taxa to the encompassing domain. In b), protist nodes with positive relationships to known extremophilic reference organisms (numbered stars) are indicated by pies colored according to the relevant associations. Numbers in extremophile reference nodes link to identities listed in Table S5. All other nodes are in grey. Analysis performed on the 2006 subset of samples, consisting of 17 diffuse vent and 3 background samples.

co-occurrence analyses. The most connected protistan taxa within these clusters belonged to MALV-II clade 12, group RAD-A radiolarians, and the apusozoan genus *Amastigomonas*.

A few of the more sparsely distributed protist nodes had significant links to vent indicator organisms belonging to the Epsilonbacteraeota. These included MALV-I clades 1 and 8, MALV-II clade 26, MALV-IV, and the fungal genus *Rhodotorula*. The stramenopile genus *Pirsonia* had strong links to anaerobic methanogens and thermophiles as well as Epsilonbacteraeota (Figure 2.5). Protist nodes that exhibited a distinctively non-extreme profile, with only significant negative correlations ($R<-0.5$, $p<0.01$) with extreme reference organisms, included MALV-I clade 6, MALV-II clades 3 and 20, MALV-III, and radiolaria belonging to the Group I Spumellaria and the genus *Pterocorys* (Figure 2.5).

2.3 Discussion

Categorization of the 13,433 OTUs in this study by distribution across different sample types (background only, cosmopolitan, or vent only), by their incidence among the 59 samples (ubiquitous, common, or rare), and by their tendency to reach high relative abundance within individual samples (frequently, infrequently, or never) has provided insight into the

potential composition and distribution of hydrothermal vent protists. We cautiously interpret relative abundances of OTUs to mitigate potential amplification bias in PCR-based studies (v. Wintzingerode et al., 1997) and variable copy numbers of the rRNA gene in different species. While information on rRNA gene copy numbers in prokaryotes is increasingly available (Stoddard et al., 2015), it is far less known for eukaryotes. Copy numbers have been linked to genome size in plants and animals (Prokopowich et al., 2003), to cell size in algae (Zhu et al., 2005), and can range from a single copy up into the hundreds of thousands for some ciliates (Gong et al., 2013). These copy number dynamics would tend to overestimate the natural abundance of organisms with high rRNA gene copy numbers, as they would produce OTUs with high relative abundance within a sample. Therefore, caution is required when inferring natural species abundance, particularly from high abundance OTUs.

The distribution patterns of protist OTUs identified in this study are congruent with early protist culturing attempts from hydrothermal vents which primarily yielded ubiquitous members of marine and freshwater ecosystems (Atkins et al., 2000). Similarly, 95% of the sequences in our study belonged to OTUs with a cosmopolitan distribution across a broad range of vent and non-vent habitats. Most had an incidence classified as ubiquitous or common and occurred with high relative abundance in at least one sample. These cosmopolitan OTUs may represent deep-sea heterotrophic opportunists drawn to the intense chemosynthetic primary productivity associated with the vents, or prolific marine parasites. In general, OTUs with a cosmopolitan distribution made up less than half of total OTU richness in this study, but >96% of the sequences from each sample belonged to this group. The remaining sequences in a sample (<4%) belonged to vent-only OTUs, which

represented as much as 55% of OTUs from the sample, indicating that vent-only OTUs are highly diverse and may include many species with low abundance in nature or low 18S gene copy numbers relative to abundant OTUs. If the latter is the case, then the natural relative abundance of potential vent organisms may be higher than suggested by these results.

Among our vent-only OTUs, there is little evidence of generalized vent endemism similar to that seen in the prokaryotic phylum Epsilonbacteraeota. Instead, endemism and the potential for protists to thrive seems to exist in narrow ranges of vent conditions and is not tied to any high-level protist group. This is evidenced by a lack of vent-only OTUs classified as ubiquitous and the finding that the few vent-only OTUs that occurred in abundance within a sample (52 of 7,826 OTUs) did so infrequently. The most frequently encountered vent-only OTU with occasional abundance occurred in 13 samples (from 5 of the 9 sites) but only in abundance in three samples with very different physical characteristics— one with fluids venting through sandy volcaniclastics near the eruptive pit on NW Rota-1 (FS447) and two from venting chimneys covered in iron-rich microbial mats from Urashima (J797-41 and J801-37). This OTU could not be identified beyond the domain Eukarya. A BLAST search (www.blast.ncbi.nlm.nih.gov) revealed an identical match to a clone from seawater collected at 2500 meters depth in the vicinity of active hydrothermal vents on the East Pacific Rise (Lie et al., 2014), suggesting that widespread vent endemism may occur within completely unknown branches of Eukarya, in species that lie poised to respond when environmental conditions become favorable for their growth.

A common finding of this and previous studies of vent-associated protists is the prevalence of alveolates, specifically the group known as the Syndiniales, previously referred to as the marine alveolate (MALV) groups. The Syndiniales, a sister group to the

dinoflagellates, have been found in diverse marine ecosystems (Groisillier et al., 2006) and are closely related to several known parasitic lineages (Guillou et al., 2008). They contain many distinct clades with few described species. The Syndiniales made up the majority of ubiquitous OTUs (65 out of 107) and all but one of the OTUs that were both ubiquitous and frequently abundant. The ubiquity of the Syndiniales and their close relationship to parasitic lineages has led to the suggestion that parasitism is a common mode of life at the hydrothermal vents and may explain sudden die offs of vent animal populations (Moreira and López-García, 2003). However, the cosmopolitan distribution of the MALV OTUs may indicate that parasitism is prevalent in much of the deep ocean, not just at vents. Indeed, the second most speciose group in our study were radiolarians, which often host parasitic or symbiotic members of the Syndiniales (Bråte et al., 2012).

The large number of unique-to-sample OTUs in this study (6,448; 48% of OTUs) contributed to high beta diversity among individual samples, which is evidenced by the relatively high stress of the NMDS and low similarity percentages between samples in Figure 2. These unique-to-sample OTUs were also the likely contributors to the strong relationship between the number of samples per site and its OTU richness. Patterns in community composition, as determined by ordination (NMDS) and hierarchical cluster analyses, were not explained by location, sample type, or any of the 15 physiochemical parameters that we evaluated against the 2006 samples, but ANOSIM results did identify significant site-based distinctions in community composition. High beta diversity that was not explained by physiochemical parameters has been reported previously among prokaryotic communities at Axial volcano and among the Epsilonbacteraeota community in the same 2006 samples analyzed in this study (Opatkiewicz et al., 2009; Huber et al., 2010), but this is the first report

of a similar pattern in protist community composition.

Protist community heterogeneity over small-scales is demonstrated by comparison of community compositions between the two pairs of diffuse fluid samples that were each collected 11 meters apart on NW Rota-1. The differential relative abundances of ubiquitous OTUs between these close proximity samples indicates that environmental selection of cosmopolitan species can happen over relatively small scales. It is important to note that, although these sites were geographically close together and had similar fluid temperatures and appearance, similarity in temperature of venting fluids is not necessarily indicative of similarity in the subsurface conditions. There are many unknown aspects of subseafloor plumbing and fluid circulation associated with hydrothermal vents which may affect the chemical and microbiological characteristics of fluids manifesting at the seafloor.

Temporal variability in protist community composition was not evaluated in this study despite collection of samples from NW Rota-1 during all five expeditions. The dynamic nature of this erupting volcano was continuously reshaping the landscape and made repeat sampling of individual vents nearly impossible. During the 2014 expedition, when eruptive activity had decreased significantly, it was difficult to locate any diffuse venting to sample. Therefore, the sample set was deemed inappropriate for temporal comparisons.

The distinctiveness of microbial communities shown previously in the bacterial domain (Huber et al., 2010) and now in protists in the Mariana arc region indicates a strong stochastic component to microbial community assembly. The inability to identify deterministic physiochemical factors influencing microbial community composition is not new, nor is it surprising given the large number of OTUs in this study with rare incidence and frequent low relative abundance. In a study of coastal surface waters most OTUs were

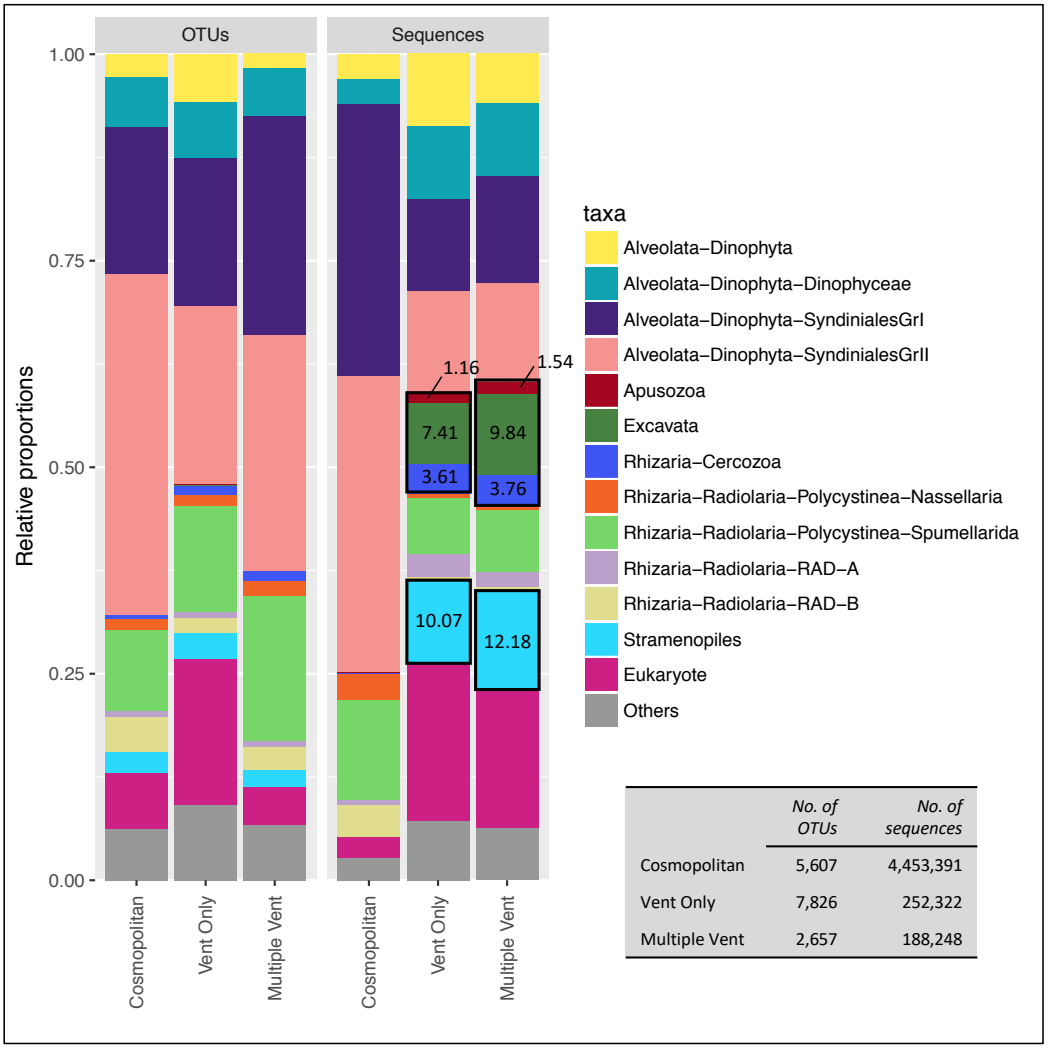
classified as rare but were shown to be metabolically active (Logares et al., 2014) supporting the idea that the protistan rare biosphere may act as a reservoir of functional potential, ready to respond to changing conditions (Caron and Countway, 2009). Therefore, rather than trying to explain whole communities by particular environmental conditions, we must consider individual species-species and species-environment interactions that may respond in a cascade to changes in environment. While we found few significant positive correlations between protists and environmental parameters, the species-species interactions proved more useful and are discussed below.

To apply this large dataset of protist sequences to the identification of likely vent endemic species we utilized two approaches—identification of OTUs found only in vent-influenced samples (vent-only OTUs) and co-occurrence patterns with extremophilic reference organisms. Both methods had strengths and weaknesses. For co-occurrence methods, grouping of OTUs by taxonomic identity allowed OTUs with low incidence and abundance to be included in the analyses. However, the taxonomic level at which OTUs were grouped may be problematic as many OTUs had shallow classification depth, particularly among vent-only OTUs. The taxonomic depth at which OTUs are aggregated in co-occurrence analyses can affect the reliability of results. Williams et al. (2014) found that the identification of keystone taxa was more consistent across varying correlation thresholds when aggregating at the family level versus the order level, which suggests that clusters of high-level novel OTUs will be more difficult to interpret. Additionally, because DNA does not distinguish between active and dead or dormant forms, we cannot say with certainty whether protists exiting the seafloor in diffuse hydrothermal fluids, along with known extremophilic prokaryotes, are cycling through, or resident in, warm subsurface

hydrothermal habitats. However, co-occurrences can provide testable hypotheses and narrow the focus of taxa and habitats for future research.

Taking a conservative approach to analysis of vent-only OTUs, we focused on those that were present in more than one sample to allow stronger inference of vent endemism within these OTUs. To that end, we identified 2,657 vent-only OTUs that were present in multiple samples and identified taxonomic groups with decreased OTU richness but an increased proportion of the sequences, relative to the entire set of 7,826 vent-only OTUs (Figure 2.4). Taxonomic groups that fit these criteria belonged to apusomonads, excavates, cercozoan Rhizaria, and stramenopiles (Figure 2.4). Decreased richness with increased relative abundance may indicate selective processes and specialization for general vent conditions within these taxonomic groups. However, we use this approach with caution as this pattern could also be attributed to increased copy numbers of the 18S rRNA gene, the variation in which is not well described among protist species.

Apusomonads are one of the broad taxonomic groups that was absent from all nine background samples in this study. A single OTU related to *Amastigomonas debruynei* was found in one sample from East Diamante, four samples from NW Eifuku with a broad range of fluid temperatures (3-103 °C), and in abundance (>1%) in one sample of clear 35 °C fluids venting through a crack in the seafloor at Nikko. *Amastigomonas* is a widely-distributed gliding heterotrophic flagellate belonging to the order Apusomonadida, which is found in both freshwater and marine systems and soils (Cavalier-Smith and Chao, 2010). It has been reported in seawater samples surrounding sulfide chimneys and within vent bivalves (Sauvadet et al., 2010), and other apusomonads occur in hydrothermal sediments with high metal concentrations (López-García et al., 2003). *Amastigomonas* fell within the tight



	No. of OTUs	*No. of sequences*
Cosmopolitan	5,607	4,453,391
Vent Only	7,826	252,322
Multiple Vent	2,657	188,248

Figure 2.4 Taxonomic composition of OTUs and sequences

OTUs and sequences identified as cosmopolitan (in vent and background samples), vent only (never in background samples), or multiple vent (in more than one vent sample, never in background samples). Black-framed boxes highlight taxa that make up a greater proportion of the sequences belonging to multi-vent OTUs than in those belonging to the broader group of vent-only OTUs. Numbers within black boxes indicate the relative proportion of sequence reads listed for each category in the table (inset). Group 'Others' included those that made up <1% of sequences in the study, namely Ciliophora (Alveolata), Syndiniales MALV groups III-V, Perkinsea (Alveolata), Amoebozoa, Centroheliozoa, Picozoa, Opisthokonta, Telonemia, Acantharea (Rhizaria), Collodaria (Rhizaria), and other unknown Rhizaria.

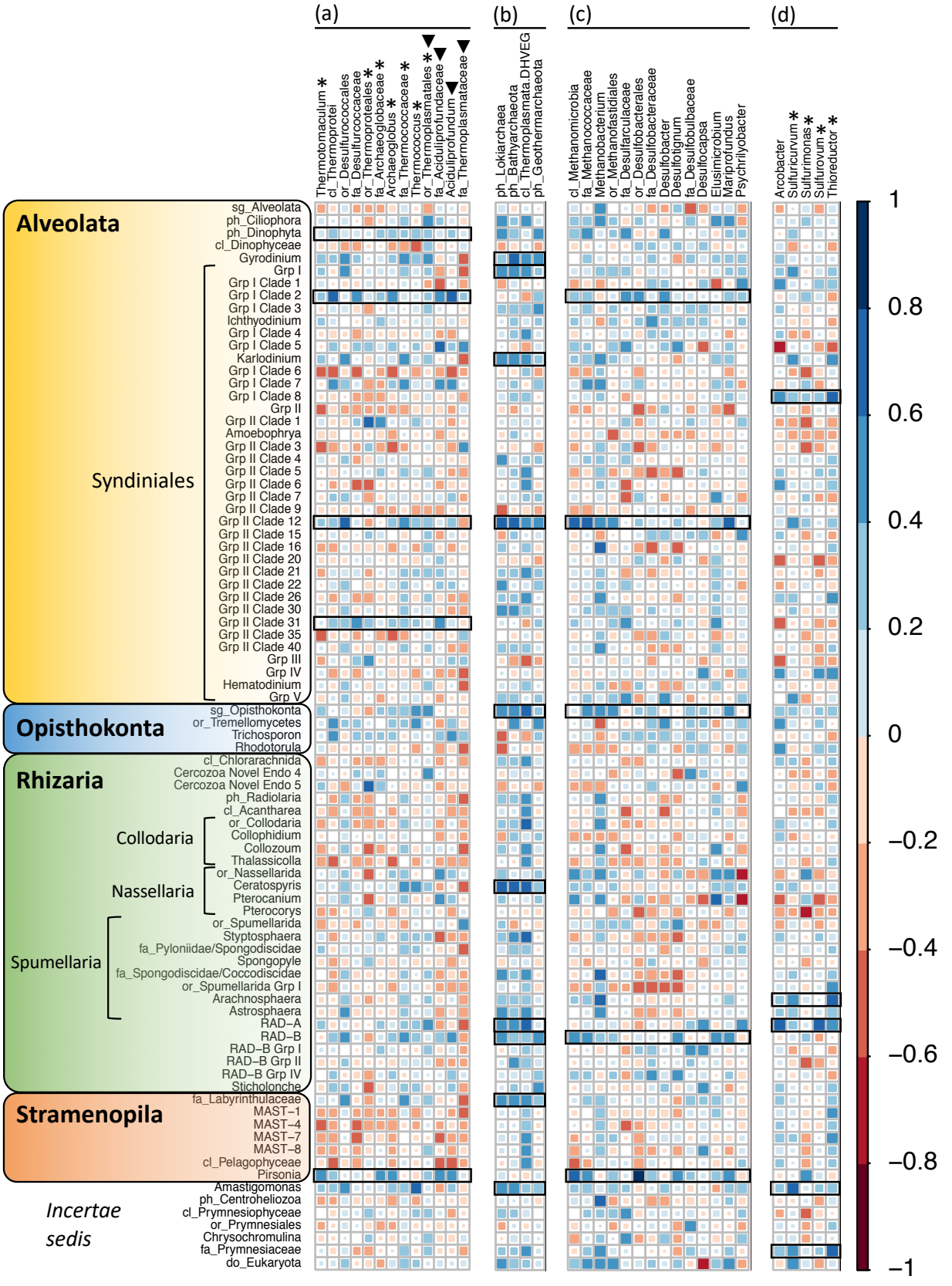

(a)
(b)
(c)
(d)
Thermotomaculum *
cl_Thermoprotei
or_Desulfurococcales
fa_Desulfurococcaceae
or_Thermoproteales *
fa_Archaeoglobaceae *
Archaeoglobus *
fa_Thermococcaceae *
Thermococcus *
or_Thermoplasmatales *
fa_Aciduliprofundaceae
Aciduliprofundum *
fa_Thermoplasmataceae
ph_Lokiarchaea
ph_Bathyarchaeota
cl_Thermoplasmata.DHVEG
ph_Geothermarchaeota
cl_Methanomicrobia
fa_Methanococcaceae
Methanobacterium
or_Methanofastidiales
fa_Desulfarculaceae
or_Desulfobacterales
fa_Desulfobacteraceae
Desulfobacter
Desulfotignum
fa_Desulfobulbaceae
Desulfocapsa
Elusimicrobium
Mariprofundus
Psychrilyobacter
Arcobacter
Sulfuricurvum *
Sulfurimonas *
Sulfurovum * *
Thioreductor * *
Alveolata
sg_Alveolata
ph_Ciliophora
ph_Dinophyta
cl_Dinophyceae
Gyrodinium
Syndiniales
Grp I
Grp I Clade 1
Grp I Clade 2
Grp I Clade 3
Ichthyodinium
Grp I Clade 4
Grp I Clade 5
Karlodinium
Grp I Clade 6
Grp I Clade 7
Grp I Clade 8
Grp II
Grp II Clade 1
Amoebophrya
Grp II Clade 3
Grp II Clade 4
Grp II Clade 5
Grp II Clade 6
Grp II Clade 7
Grp II Clade 9
Grp II Clade 12
Grp II Clade 15
Grp II Clade 16
Grp II Clade 20
Grp II Clade 21
Grp II Clade 22
Grp II Clade 26
Grp II Clade 30
Grp II Clade 31
Grp II Clade 35
Grp II Clade 40
Grp III
Grp IV
Hematodinium
Grp V
Opisthokonta
sg_Opisthokonta
or_Tremellomycetes
Trichosporon
Rhodotorula
Rhizaria
cl_Chlorarachnida
Cercozoa Novel Endo 4
Cercozoa Novel Endo 5
ph_Radiolaria
cl_Acantharea
Collodaria
or_Collodaria
Collophidium
Collozoum
Thalassicolla
Nassellaria
or_Nassellarida
Ceratospyris
Pterocanium
Pterocorys
Spumellaria
or_Spumellarida
Styptosphaera
fa_Pyloniidae/Spongodiscidae
Spongopyle
fa_Spongodiscidae/Coccodiscidae
or_Spumellarida Grp I
Arachnosphaera
Astrosphaera
RAD–A
RAD–B
RAD–B Grp I
RAD–B Grp II
RAD–B Grp IV
Sticholonche
Stramenopila
fa_Labyrinthulaceae
MAST–1
MAST–4
MAST–7
MAST–8
cl_Pelagophyceae
Pirsonia
Incertae sedis
Amastigomonas
ph_Centroheliozoa
cl_Prymnesiophyceae
or_Prymnesiales
Chrysochromulina
fa_Prymnesiaceae
do_Eukaryota
1
0.8
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
−0.8
−1

Figure 2.5 Spearman correlation matrix

Matrix between Eukarya (vertical axis) and prokaryotes that are known thermophiles and hyperthermophiles (panel a), novel lineages identified from crustal/subsurface habitats (panel b), known anaerobes (panel c), and vent indicator organisms belonging to the Epsilonbacteraeota (panel d). Strength and sign of correlation are indicated by both color and size of the squares, with size increasing away from zero. Figure includes correlations at all levels, not just those with values stronger than positive or negative 0.5. Black-framed boxes highlight the eukaryotic taxa with the top 5 (panels a, c, d) and top 10 (panel b) average R-values. Symbols in panels a and d: * also anaerobic; ▼also acidophilic. OTUs are identified to the genus level unless preceded by a 2-letter code indicating the depth of taxonomic identity: sg=supergroup, ph=phylum, cl=class, or=order, fa=family. Prokaryote taxa from Huber et al 2010.

aggregation of protists and archaea in the network diagram (Figure 2.3) and had strong pairwise correlations with Epsilonbacteraeota (Figure 2.5) and the archaeal candidate phyla Lokiarchaeota, Bathyarchaeota, and Geothermarchaeota, all of which are known to inhabit deep subsurface sediments and crustal habitats (Meng et al., 2014; Spang et al., 2015; Jungbluth et al., 2017). Together, this suggests that apusomonads may play a central role as primary grazers in subsurface food webs at hydrothermal vents.

Two excavate OTUs accounted for nearly 10% of the sequences in the multi-sample vent-only OTUs (Figure 2.4). One was related to *Jakoba libera*, a deeply-branching heterotrophic nanoflagellate from marine sediments (Patterson, 1990) that is notably absent from previous hydrothermal vent protist surveys, even those focused on hydrothermal sediments (Edgcomb et al., 2002). This OTU, present at NW Eifuku, NW Rota-1, and the iron-rich Urashima site, was found in abundance in diffuse fluids from a low temperature vent on NW Rota-1 (J800-18) and the top of a large chimney with a thick covering of orange microbial mat at Urashima (J801-37). *J. libera* has been successfully cultured from diffuse hydrothermal fluids from the Mariana region (SM, unpublished data), further suggesting adaptation to the vent environment in this organism. The second multi-sample excavate OTU

was related to *Trimastix pyriformis*, a free-living anaerobic flagellate that lacks traditional mitochondria but expresses proteins that are typical of both aerobic mitochondria and anaerobic hydrogenosomes (Hampl et al., 2008). Such versatility could promote evolutionary success in areas with steep oxygen gradients such as hydrothermal systems, (although the related OTU was only found at NW Rota-1 in low abundance). Unfortunately, co-occurrences could not be applied to either of these excavate OTUs as neither appeared in the 2006 samples used in those analyses. However, to our knowledge, ours is the first report of these two genera in hydrothermal systems; their presence in multiple vent sites, absence from background seawater, and occasional abundance makes excavates interesting targets for future culturing and genomic efforts.

Cercozoan Rhizaria, particularly related to the widely-distributed *Massisteria marina* (Larsen and Patterson, 1990), have been recovered from hydrothermal systems by both culture-based and culture-independent methods from four locations in the Eastern Pacific (Atkins et al., 2000), from carbonates in the Lost City field, Mid-Atlantic Ridge (López-García et al., 2007), and within or near vent bivalves (Sauvadet et al., 2010). Twenty-eight OTUs belonging to diverse Cercozoa, many of which could only be identified to the subphylum or class level, occurred across all sites except Kasuga-2, and with very limited numbers of sequences at Daikoku and NW Eifuku. Only three of these OTUs were ever found in abundance within a sample. The OTU with the highest incidence (11 samples from NW Rota-1 and NW Eifuku) was abundant only in one sample of diffuse fluids venting through a sulfur crust on NW Rota-1 and had 97% similarity to sequences from seasonally and permanently anoxic basins (Stoeck et al., 2006; Orsi et al., 2012). The other two OTUs that reached abundance had only 90% and 91% similarity to soil and freshwater sequences in the public

databases and likely represent new cercozoan lineages.

Co-occurrences do not indicate strong adaptation to extreme conditions in Cercozoa (Figure 2.5) with two novel clades (Endo-4 and Endo-5) having only limited positive correlations with two archaeal thermophilic orders amidst a mostly negative trend with extremophiles. However, co-occurrences included vent-only and cosmopolitan cercozoan OTUs, with the former exhibiting very low incidence and abundance in contrast to the four cosmopolitan OTUs, which contributed 87% of the included cercozoan sequences. The Endo-4 and -5 clades fall within the Endomyxa, which contain many parasitic forms (Bass and Cavalier-Smith, 2009). Together these findings suggest that some Cercozoa may have limited tolerance for hydrothermal conditions and may be common opportunists or parasites in vent systems rather than endemics.

Stramenopiles are a diverse phylum of Eukaryotes composed of basal heterotrophic groups, including several parasitic lineages, and a phototrophic crown radiation (Patterson, 1989; Leipe et al., 1996; Burki, 2014). The 52 stramenopile multi-sample OTUs that had increased numbers of sequences in hydrothermal vent samples included only six that reached abundance. These six were closely related to either algal groups that have been found in seasonally and permanently anoxic basins (Edgcomb et al., 2011; Orsi et al., 2012) or the diatom-parasitizing nanoflagellate genus *Pirsonia* (Kühn et al., 2004). Additionally, many of the non-abundant stramenopile OTUs also belonged to algal groups. The presence of these seemingly common photosynthetic organisms and their parasites in our vent samples is likely the result of subsurface entrainment in hydrothermal fluids as they circulate through sediments laden with settled surface ocean communities. However, the genus *Pirsonia* presents a less clear case considering the strong pairwise correlations with

anaerobes and thermophiles (Figure 2.5) and positive correlation with concentrations of hydrogen gas (Table S4). We cannot rule out the potential that the *Pirsonia* sequences belong to a new species of this known parasite that has a vent-specific host. Future application of co-occurrence analyses that include vent endemic meio- and macrofaunal may help to address this hypothesis.

Stramenopile OTUs related to heterotrophic lineages that were found across multiple vent sites included *Wobblia lunata* (Moriya et al., 2000), found only at NW Eifuku, and two OTUs related to the pseudo-fungus *Hyphochytrium catenoides* (Cooney et al., 1985), one found only on NW Rota-1 and the other only at Urashima. OTUs belonging to these heterotrophic groups were not present in the 2006 samples and were therefore not included in co-occurrences but are potential targets for future culturing efforts. Eight OTUs belonged to the novel marine stramenopile groups (MAST), that are abundant in open oceans and coastal regions (Díez et al., 2001; López-García et al., 2001; Moon-van der Staay et al., 2001; Massana et al., 2002) and include some putatively anoxic- or microoxic-adapted clades (Massana et al., 2004). However, all of the MAST clades included in co-occurrence analyses (MAST-1, -3, -4, -7, and -8) are thought to be made up of widely distributed, marine aerobic organisms (Massana et al., 2004), which is confirmed here by mostly negative relationships to extremophilic reference organisms (Figure 2.5). As with the Cercozoa, several multi-vent stramenopile OTUs could not be identified below the phylum level and may represent further novel vent-adapted lineages.

One group within the stramenopiles exhibited a strong co-occurrence relationship with crustal/subsurface archaea— the family Labyrinthulaceae (Figure 2.5). Labyrinthulids are colonial heterotrophs that are widely distributed in marine systems and can exist as

parasites and commensals of both marine plants and animals (Raghukumar, 2002). They have also been found in anoxic basins (Edgcomb et al., 2011; Orsi et al., 2012) and in hydrothermal and cold seep sediments (Edgcomb et al., 2002; Takishita et al., 2005; Takishita et al., 2007). This wide distribution across various extreme and non-extreme environments indicates that some members of the Labyrinthulaceae have adapted a tolerance for anoxic, reducing-habitat conditions and may include some vent endemics.

As previously discussed, OTUs identified only as Eukarya were more speciose in the vent-only OTUs (18%) than those with cosmopolitan distribution (7%) and made up the third most abundant component of OTUs in our study. These unknown Eukarya also made up the greatest proportion of sequences among the multi-vent OTUs at 16.6% (Figure 2.4). While few inferences can be made regarding unknown Eukarya due to the likelihood of containing several unknown species, it is important to note that these widely-distributed and potentially vent-endemic organisms belong to novel phyla with unknown genetic potential. This finding should serve as a call to action for increased culturing efforts in hydrothermal systems using a broad suite of culturing techniques in order to identify and characterize these novel organisms.

Co-occurrences revealed useful distinctions between habitat conditions preferred by various clades within the MALV group Syndiniales, which currently have few cultured representatives. Potential specialization for hydrothermal or anoxic habitats has been suggested for MALV-II clades 9 and 15 (Guillou et al., 2008), but our co-occurrence results suggest that MALV-II clade 12 is heavily linked to anoxic, thermophilic and crustal/subsurface habitats (Figure 2.5). This clade was the most connected node within the dense aggregation of protist nodes in the co-occurrence network (Figure 2.3) suggesting a

central role in the deep subsurface microbial community, potentially one of a parasitic nature due to the prevalence of parasitism among the Syndiniales. However, no OTUs in this clade were ever found in high relative abundance furthering the case that the most likely vent endemics are rare and may be in low abundance. Analysis of deep sediment and crustal protist communities are needed to confirm this hypothesis and future inclusion of vent endemic meio- and macrofaunal species in co-occurrences with protists may help to identify potential symbiotic associations such as parasitism.

MALV-I clade 2 exhibited a consistent trend of positive linkages with thermophiles and acidophiles, with some positive linkages with anaerobes (Figure 2.5). This clade was the most abundant group of OTUs included in the co-occurrences at 21% of sequences (large red/blue node in upper right quadrant of figure 2.3b), with most of the OTUs having a ubiquitous distribution. This clade has been identified in a variety of deep ocean and hydrothermal vent habitats (Groisillier et al., 2006; Guillou et al., 2008) and is therefore considered here as a deep-sea opportunist that may be able to tolerate or adapt to extreme conditions. At the other end of the spectrum, MALV-I clade 6, MALV-II clades 3 and 20, and MALV-III all exhibited distinctly non-extremophilic profiles, with mostly negative correlations to extremophiles (Figure 2.5) as well as some positive links to common seawater prokaryotes (Table S6), possibly an indication of general avoidance of extreme conditions within these clades.

2.4 Conclusion

Protists associated with hydrothermal systems are a mix of potential endemics, surface detritus entrained from local sediments, abundant cosmopolitan organisms from seawater,

and possible parasites. Elucidating the roles these protists play in ecosystem function and trophodynamics will require disentanglement of the various influences on community composition and identification of the active contributors to the system through future RNA-based studies. While molecular methods have opened our eyes to the greater breadth of diversity in these extreme environments, culturing efforts must be maintained in order to discover and describe new species and identify their preferred conditions for life at vents. As has been demonstrated by the decreased classification depth among vent-only OTUs, the reference databases lack close relatives of potential vent endemics, which further argues for increased culturing efforts of protists from hydrothermal vents, using a wide array of culturing techniques.

Co-occurrence analyses have provided a strong basis for the formation of hypotheses regarding the habitat preference and potential roles of undescribed clades of protists within the vent environment, including members of the Cercozoa and Syndiniales. Going forward, inclusion of vent endemic meio- and macrofaunal species in co-occurrence analyses with protists may help to further identify specific symbiotic associations.

The combination of high-throughput sequencing with a large and diverse suite of hydrothermally-influenced and non-vent seawater samples has allowed us, for the first time, to evaluate the potential specialization of both ubiquitous and rare protist groups for life in vent habitats. While we cannot conclusively identify any protist group to be a vent endemic there is strong evidence to focus future searches towards labyrinthulids, apusomonads, and excavates. We have demonstrated that 1) the majority of sequences recovered in samples from hydrothermal vents likely come from ubiquitous deep-sea protists, 2) potential vent endemics make up a small percentage of the sequences but a large portion of the diversity in

most samples, 3) many potential vent endemics belong to novel lineages both at the phylum level and within defined clades, and 4) co-occurrence patterns with prokaryotes have the potential to provide insight into the nature of the vent protist community, if only by providing future direction and reasonably testable hypotheses.

2.5 Experimental Procedures

2.5.1 Site descriptions and sample collection

The Mariana volcanic arc and southern back-arc basin were sampled during five research expeditions, with funding from NOAA Ocean Exploration and Research (all cruises) and the National Science Foundation (2009, 2010, 2014). During the exploratory 2004 and 2006 expeditions (Embley et al., 2007) a mix of hydrothermal and background fluid types were sampled from seven submerged sites along the Mariana arc and one site in the southern back-arc basin (Forecast; Figure S1, Figure S2, Table S1) using the Remotely Operated Vehicles (ROVs) *ROPOS* and *Jason*. The subsequent 2009 and 2010 missions focused on the actively erupting site NW Rota-1 (Embley et al., 2006) and allowed for more thorough sampling of active vent sites (Figure S2). The 2014 expedition was more limited in scope but included revisiting sites along the arc that had not been sampled since 2006, sampling one additional site on the back-arc (Urashima), and returning to NW Rota-1 (Figure S2) to find greatly decreased volcanic activity relative to all prior visits in the previous decade (Schnur et al., 2017).

Types of fluids collected for culture-independent surveys of microbial diversity included focused diffuse hydrothermal fluids classified as either low temperature (<100°C, “diffuse”) or high temperature (>100°C, “high temp”), dense hydrothermal discharge for

which no point source could be found but which exhibited a measurable temperature anomaly ("diffuse/plume"), buoyant plumes collected above vent fields with minimal detectable temperature anomalies ("plume"), deep seawater immediately adjacent to discharging hydrothermal fluids ("near-field"), and deep seawater with no detectable hydrothermal signal ("background").

Diffuse, high temp, diffuse/plume, and near-field fluids were sampled by *in situ* filtration onto Sterivex-GP filters (0.2 μm pore size) using the Hydrothermal Fluid and Particle Sampler (Butterfield et al., 2004) mounted on the ROV. Background and plume samples were collected in Niskin bottles on a sampling rosette instrumented with sensors for measuring conductivity, temperature, depth, oxygen, optical backscatter, and reduction potential (Eh). Approximately 3L of fluids were filtered through individual Sterivex filters, which were subsequently stored at -80°C until further processing.

2.5.2 DNA extraction and high-throughput sequencing

DNA from the 2004 and 2006 expeditions was extracted from filters using the method of (Huber et al., 2002) with the following modifications: Lysozyme (200 μl, 50 mg/ml) was included along with the proteinase K and SDS during the first 65°C incubation step, and DNA was resuspended in 200 μl TE buffer (10 mM Tris, 1mM EDTA, pH 8.0). For later samples the method was modified further to increase DNA yields, by following the protocol in Sogin et al. (2006). A sterile blank filter was included with each batch of samples extracted and the resultant material was carried through initial screening along with samples to assure no contaminants were introduced during extraction.

Extracts (samples and blanks) were screened for the presence of eukaryotic DNA by

Polymerase Chain Reaction (PCR) using the primers EukA (5'-AACCTGGTTGATCCTGCCAGT-3') and EukB (5'-TGATCCTTCTGCAGGTTCACCTAC-3') (Medlin et al., 1988). For samples that showed eukaryote amplification, an aliquot of the genomic DNA was sent for high-throughput sequencing on an Illumina MiSeq at the Laboratory for Advanced Genome Analysis (Vancouver Prostate Centre core facility, Vancouver, Canada) using Illumina 2x300bp chemistries. The primers TAReuk454FWD1 (5'-CCAGCA(G/C)C(C/T)GCGGTAATTCC-3') and TAReukREV3 (5'-ACTTTCGTTCTTGAT(C/T)(A/G)A-3') (Stoeck et al., 2010) were used to target and sequence the V4 region of the 18S rRNA gene.

2.5.3 Sequence and statistical analyses

All sequence and statistical analysis steps were preformed using the software package Mothur v.1.35.1 (Schloss et al., 2009), following the "MiSeq SOP" analysis example (http://mothur.org/wiki/MiSeq_SOP), unless stated otherwise. Sequences were quality filtered according to the recommendations in Huse et al. (2007), removing sequences with ambiguous bases, mismatches to the primer sequence, homopolymers longer than 7 nucleotides, and lengths shorter than 200 base pairs. Primer sequences were removed, and the remaining sequences were trimmed to encompass *Saccharomyces cerevisiae* positions 586-963. All sequences were compared using the average neighbor pairwise distance method and clustered into Operational Taxonomic Units (OTUs) using a 98% similarity threshold. The diversity of each sample was estimated using both the Shannon and Inverse Simpson (1/Simpson) indices on 1) the full component of protist sequences, 2) only sequences belonging to nonsingleton protist OTUs, and 3) a normalized value of n=2000

sequences.

An OTU-by-sample matrix was created, square root transformed and used to calculate distances between samples by the Morisita-Horn measure (Horn, 1966) using the *vegan* package for R (Oksanen et al., 2015). Using this distance matrix, we performed hierarchical cluster analysis and Nonmetric Multidimensional Scaling (NMDS) in *vegan*, and significance of clustering patterns based on location (site) and sample type (diffuse, background, etc.) was evaluated using one-way Analysis of Similarity (ANOSIM) in Mothur. Taxonomic assignments of OTUs were determined by reference to the Protist Ribosomal Reference database (PR^2, version 4.9.2) of 18S rRNA V4 sequences (Guillou et al., 2013).

As an indicator of preference for vent habitats, each OTU was assigned a code based on its incidence across the different sample types (diffuse, background, etc.). This allowed us to score habitat preference of each OTU as 'background only', 'vent only', or 'cosmopolitan' (i.e., found in both vent and background habitats). The 'vent only' category included diffuse fluids (high and low temperature and diffuse/plume samples), buoyant plumes, and near-field samples, which are all known to be mixtures of hydrothermal fluids and background seawater to varying degrees.

2.5.4 Investigating potential drivers of diversity

A subset of samples, for which complementary chemical measurements and prokaryote diversity data were available, were used to evaluate potential physiochemical and interspecies drivers of protist diversity. Chemical composition of hydrothermal fluids was measured on board ship and in the laboratory of D. Butterfield at University of Washington and NOAA/PMEL, as reported in Huber et al. (2010). Potential links between protist diversity

and physiochemical parameters were investigated on the seventeen vent-influenced samples collected in 2006 (Table S1) using the *vegan* R package (Oksanen et al., 2015) with the physiochemical dataset used by Huber et al. (2010), which included temperature, depth, pH, dissolved gases (H_2S, H_2, CH_4), SO_4, Na, Li, Fe, Mn, Si, and major nutrients (NO_3, PO_4, NH_4). We tested for significant ($p<0.005$) drivers of diversity by three methods: 1) constrained ordination (Canonical Correspondence Analysis, CCA), 2) fitting environmental data to an unconstrained ordination (NMDS) using the *envfit* function, and 3) permutational multivariate analysis of variance (PERMANOVA) using the *adonis* function.

Spearman correlations between bacterial, archaeal, and protist OTUs from 20 samples (17 diffuse vent fluid, 3 background) were used to create interspecies co-occurrence networks following R scripts developed by Williams et al. (2014). Bacterial and archaeal OTU counts from samples reported by Huber et al. (2010) were obtained from the VAMPS database (Huse et al., 2014) and combined with protist OTUs from this study. OTU counts were converted to relative abundances within each of the three domains (i.e. all archaeal OTUs in a sample sum to 100%, all bacterial OTUs to 100% and all protistan OTUs to 100%). Comparisons were made using OTUs grouped by taxonomic identity at the lowest taxonomic level to which they could be classified using the PR^2 (protists) and Silva 132 (Bacteria and Archaea) databases. Taxa that were present in less than 15% of samples (3 of 20) or with a total abundance of <50 sequences among all samples were excluded from co-occurrence analyses. After removal of these low incidence and abundance taxa, samples retained 80-95%, 83-98%, and >98% of their protistan, bacterial, and archaeal sequences, respectively, for co-occurrence analysis.

Generally, only co-occurrences with correlation values ≥ 0.5 and p-values <0.01 were

considered relevant, but all values were considered in looking at general trends with extremophilic reference taxa (below). Correlation matrices showing all positive and negative correlation values between protists and select prokaryotes were created using the *corrplot* package in R (Wei and Simko, 2017). Co-occurrence networks were visualized using a force-directed layout in Cytoscape v.3.6.0 (Shannon et al., 2003) and included only significant ($p<0.01$) positive and negative correlations within and between Bacteria, Archaea, and protists. Markov clustering was used to identify clusters using the *MCL* package in R (van Dongen, 2000).

Preference of the various protist groups for the extreme conditions often encountered in hydrothermal systems (high temperature, acidity, anoxia) was inferred from positive co-occurrences with reference organisms known to inhabit them. These included prokaryote reference taxa that are known thermophiles, hyperthermophiles, acidophiles, and anaerobes, as well as known vent indicator organisms belonging to the Epsilonbacteraeota, and are listed in Table S5. Linkages to novel archaeal phyla isolated from crustal/subsurface habitats were used to infer preference for these habitats as well (Table S5). Physiochemical data were also included in co-occurrence network analyses to determine if particular taxa were more or less influenced by environmental parameters.

2.5.5 Sequence submission

Protist sequences of the 18S rRNA gene V4 region have been deposited in the NCBI Sequence Read Archive under accession no. PRJNA509492.

2.6 Article Supplementary Material

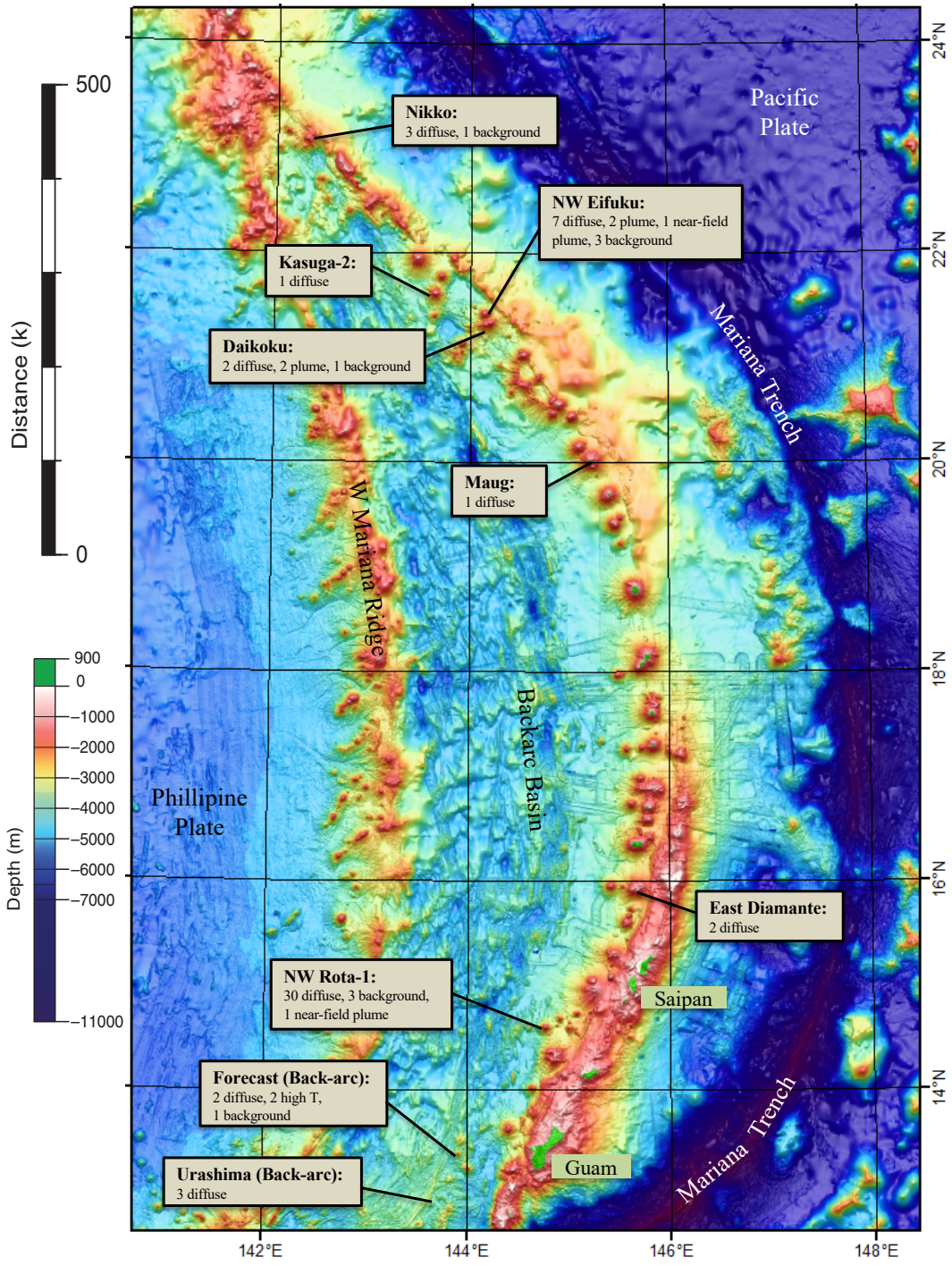


Figure 2.6 Supplementary Figure S1: Map of Mariana region

Mariana volcanic arc and back-arc regions with samples listed from each of nine sites. (Map courtesy of R. Embley and S. Merle, NOAA/PMEL.)

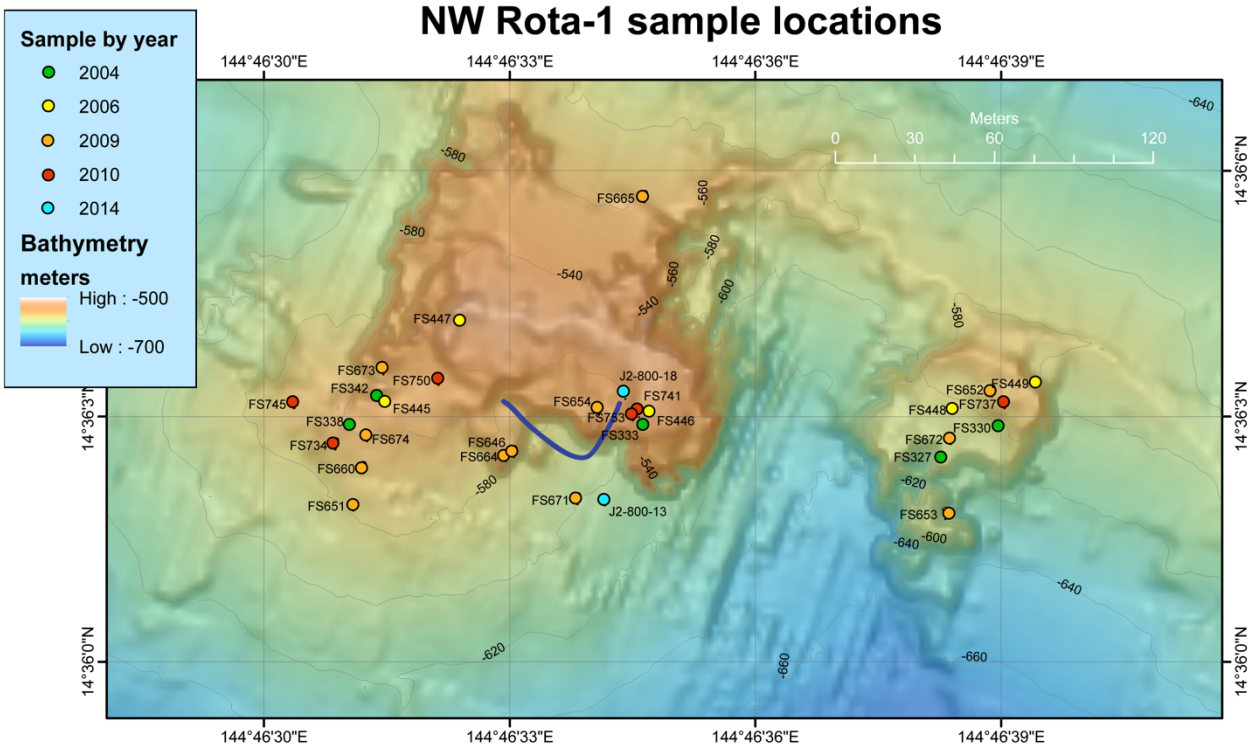


Figure 2.7 Supplementary Figure S2: Map of NW Rota-1

Map of NW Rota-1 volcano showing the locations of the 30 diffuse and diffuse/plume samples collected during five sampling years. Not shown are three background water samples that were collected away from the volcano at a similar depth to the vent samples. Near-field sample (J800-10) was collected along the path indicated by the blue line. (Map courtesy of R. Embley and S. Merle, NOAA/PMEL)

Table 2.3 Supplementary Table S1: Details of samples included in the study

Sample	Site	Vent name, background or plume	Sample Type	Year	Depth (m)	Tmax (°C)	Latitude (DecDeg)	Longitude (DecDeg)
FS479	Nikko	North Vent	D	2006	460	96.1	23.0810	142.3255
FS480	Nikko	Tubeworm Hangover	D	2006	447	26.1	23.0791	142.3264
FS481	Nikko	Top Vent	D	2006	413	35.4	23.0798	142.3269
FS482	Nikko	*Background seawater*	B	2006	441	14.9	23.0778	142.3252
FS380	Kasuga-2	Cracked Vent	D	2004	387	27.0	21.6084	143.6363
FS363	NW Eifuku	Ice Fall	D	2004	1643	15.0	21.4877	144.0407
FS364	NW Eifuku	Champagne	HT	2004	1607	103.0	21.4876	144.0415
FS371	NW Eifuku	Sulfur Dendrite	D	2004	1615	55.3	21.4877	144.0415
FS468	NW Eifuku	Cliff House	D	2006	1578	54.1	21.4872	144.0421
FS469	NW Eifuku	Cliff House	D	2006	1578	43.6	21.4872	144.0421
J798-02	NW Eifuku	near Champagne	D	2014	1608	26.5	21.4874	144.0414
J799-33	NW Eifuku	Champagne area	NF	2014	1608	3.0	21.4874	144.0416
V06-10	NW Eifuku	*Background seawater*	B	2014	1000	3.7	21.8168	144.4999
V06-6	NW Eifuku	*Background seawater*	B	2014	1500	2.6	21.8168	144.4999
V10-10	NW Eifuku	*Background seawater*	B	2014	1525	2.6	21.4874	144.0414
V10-14	NW Eifuku	Yellow Cone plume	P	2014	1576	2.5	21.4878	144.0419
V10-6	NW Eifuku	Champagne plume	P	2014	1585	2.4	21.4874	144.0414
FS473	Daikoku	NE Pit	D/P	2006	440	15.6	21.3245	144.1929
FS475	Daikoku	Alka Seltzer	D	2006	414	64.5	21.3250	144.1914
T03-11	Daikoku	Buoyant plume	P	2014	358	14.2	21.3323	144.1993
T03-17	Daikoku	*Background seawater*	B	2014	506	8.8	21.3339	144.2104
T04-5	Daikoku	Buoyant plume	P	2014	384	14.4	21.3249	144.1900
FS357	Maug	Cave Vent	D	2004	145	29.0	20.0234	145.2226
FS349	East Diamante	Barnacle Beach	D	2004	457	13.7	15.9433	145.6828
FS462	East Diamante	Five Towers	D	2006	353	26.0	15.9428	145.6814
FS338	NW Rota-1	Brimstone Pit	D/P	2004	561	29.5	14.6008	144.7753
FS342	NW Rota-1	Shimmering Sands	D	2004	539	69.1	14.6009	144.7754
RotaCTD04	NW Rota-1	*Background seawater*	B	2004	580	6.6	14.5760	144.7978
FS445	NW Rota-1	Brimstone Pit	D/P	2006	561	27.9	14.6009	144.7754
FS446	NW Rota-1	Iceberg complex*	D	2006	534	53.8	14.6009	144.7763
FS447	NW Rota-1	Sandy Saddle	D	2006	521	35.6	14.6012	144.7757
FS448	NW Rota-1	Fault Shrimp	D	2006	586	25.9	14.6009	144.7774
FS449	NW Rota-1	Scarp Top	D	2006	568	17.6	14.6010	144.7776
RotaCTD06	NW Rota-1	*Background seawater*	B	2006	551	6.8	14.3833	144.5667
FS646	NW Rota-1	Mat Mecca	D	2009	538	45.8	14.6007	144.7759
FS652	NW Rota-1	Scarp Top	D	2009	566	18.0	14.6009	144.7775
FS653	NW Rota-1	Iron 599	D	2009	599	8.6	14.6005	144.7773
FS654	NW Rota-1	Iceberg complex*	D	2009	534	39.0	14.6009	144.7761
FS660	NW Rota-1	Brimstone Floc	D	2009	521	19.7	14.6007	144.7753
FS664	NW Rota-1	Flying Floc	D	2009	529	46.3	14.6007	144.7758
FS673	NW Rota-1	Summit Ridge, S side	D	2009	529	33.0	14.6010	144.7754
FS674	NW Rota-1	Floc Rock	D	2009	521	32.2	14.6008	144.7754
RotaCTD09	NW Rota-1	*Background seawater*	B	2009	503	8.0	14.5657	144.8572
FS734	NW Rota-1	Sulfur Crust	D	2010	557	25.9	14.6007	144.7753
FS737	NW Rota-1	Scarp Top	D	2010	564	22.8	14.6009	144.7775
FS741	NW Rota-1	Iceberg complex*	D	2010	533	33.3	14.6009	144.7763
FS745	NW Rota-1	Phantom	D	2010	563	25.6	14.6009	144.7751
FS750	NW Rota-1	Arrowhead	D	2010	546	32.1	14.6010	144.7756
FS753	NW Rota-1	Iceberg complex*	D	2010	532	20.7	14.6008	144.7763
J800-10	NW Rota-1	Iceberg area	NF	2014	584	6.9	14.6005	144.7761
J800-18	NW Rota-1	Iceberg complex*	D	2014	534	19.6	14.6009	144.7762
FS430	Forecast	Bart Vent	HT	2006	1449	155.0	13.3946	143.9201
FS431	Forecast	Snail Scrum	D	2006	1448	16.0	13.3946	143.9201
FS432	Forecast	Homer Vent	D	2006	1448	19.0	13.3953	143.9199
FS433	Forecast	Marge Vent	HT	2006	1448	140.0	13.3953	143.9199
FS435	Forecast	*Background seawater*	B	2006	1340	ND	13.3950	143.9200
J797-41	Urashima	Saipanda Horn	D	2014	2845	25.9	12.9222	143.6492
J801-37	Urashima	Golden Horn Spire (top)	D	2014	2922	30.2	12.9224	143.6493
J801-8	Urashima	Golden Horn Spire (base)	D	2014	2930	11.8	12.9224	143.6493

* Iceberg complex includes five individual vent locations within a 20m diameter area that were sampled in four of five sampling years.
Sample types: B=background, D=point source diffuse, D/P=diffuse/plume, P=buoyant plume, NF=near-field, HT=high temperature
ND =No data.

Table 2.4 Supplementary Table S2: Samples not included in the final analysis

Sample	Volcano	Vent name	Year	Depth (m)	Tmax (℃)	Latitude (DecDeg)	Longitude (DecDeg)
Insufficient MiSeq reads returned							
FS467	NW Eifuku	Sulfur Dendrite	2006	1612	52.7	21.4874	144.0417
FS651	NW Rota-1	Brimstone	2009	524	55.1	14.6005	144.7753
J800-13	NW Rota-1	Smokin' Stones	2014	590	10.4	14.6005	144.7762
Unsuccessful amplification of 18SrRNA genes							
FS327	NW Rota-1	Fault Shrimp	2004	583	22.6	14.6007	144.7773
FS330	NW Rota-1	Scarp Top	2004	564	39.9	14.6008	144.7775
FS333	NW Rota-1	Iceberg complex	2004	530	52.3	14.6008	144.7763
FS665	NW Rota-1	The Amphitheater	2009	558	13.5	14.6016	144.7763
FS671	NW Rota-1	Limpet Lair	2009	548	16.5	14.6006	144.7761
FS672	NW Rota-1	Sulfur Slide	2009	571	17.9	14.6008	144.7773

Table 2.5 Supplementary Table S3: Results of 18S rRNA gene sequencing

Samples are listed in the same order as in table S1.

Sample	Total Eukaryote V4 sequences *	% Metazoan sequences	Protist V4 sequences ‡	Nonsingleton Protist OTUs	OTUs not in background	% sample OTUs not in background	Shannon diversity	1/Simpson diversity
FS479	188801	64.3%	66444	852	324	38.0%	4.48	42.25
FS480	41760	63.7%	15076	62	16	25.8%	2.16	4.79
FS481	273049	40.1%	162908	434	168	38.7%	4.00	26.65
FS482	173589	73.3%	45555	1488	*NA*	*NA*	5.20	56.42
FS380	231667	92.3%	17132	770	220	28.6%	4.57	26.88
FS363	187134	94.1%	10727	315	99	31.4%	4.57	50.02
FS364	32612	54.1%	14657	209	58	27.8%	4.03	29.36
FS371	217115	93.8%	12934	247	71	28.7%	3.87	22.66
FS468	156348	59.1%	63597	345	90	26.1%	3.94	26.18
FS469	123464	35.8%	78585	1052	426	40.5%	4.56	39.80
J798-02	167749	99.6%	599	128	29	22.7%	3.86	21.14
J799-33	101054	96.3%	3713	101	25	24.8%	3.07	9.27
V06-10	46859	0.9%	46162	447	*NA*	*NA*	3.74	15.58
V06-6	118301	5.2%	111688	582	*NA*	*NA*	4.54	46.52
V10-10	97860	7.8%	89490	1096	*NA*	*NA*	4.37	33.85
V10-14	59721	51.9%	28131	912	279	30.6%	4.22	24.69
V10-6	86217	53.0%	40377	80	19	23.8%	3.02	15.54
FS473	170125	2.7%	162292	2428	1199	49.4%	4.69	31.34
FS475	42606	28.8%	30222	205	34	16.6%	4.22	38.92
T03-11	30604	0.9%	29823	907	357	39.4%	4.27	31.40
T03-17	126193	2.8%	121855	1180	*NA*	*NA*	4.21	23.54
T04-5	145261	3.5%	139030	1486	642	43.2%	3.94	12.08
FS357	146004	96.8%	4570	188	73	38.8%	3.03	4.79
FS349	184895	94.6%	9572	401	164	40.9%	4.33	24.19
FS462	63591	56.3%	27334	985	403	40.9%	4.55	30.65
FS338	232821	2.0%	224633	2076	915	44.1%	4.30	28.53
FS342	85222	0.2%	83476	501	274	54.7%	3.60	13.43
RotaCTD04	183554	2.9%	175866	2004	*NA*	*NA*	4.33	21.03
FS445	48397	4.1%	46211	167	28	16.8%	3.94	34.67
FS446	1133	1.3%	1113	105	2	1.9%	3.89	32.24
FS447	258958	0.9%	255159	1579	673	42.6%	4.41	39.43
FS448	180258	2.5%	173329	1258	698	55.5%	4.47	42.31
FS449	237974	3.4%	229213	295	77	26.1%	3.71	21.63
RotaCTD06	268667	17.5%	220834	509	*NA*	*NA*	4.11	36.78
FS646	76822	0.3%	76089	923	435	47.1%	4.75	40.55
FS652	133000	2.3%	128632	1799	736	40.9%	4.50	33.17
FS653	141787	3.6%	134847	2301	1211	52.6%	5.26	57.61
FS654	55180	1.3%	54229	575	128	22.3%	4.13	20.80
FS660	6181	0.2%	6137	127	16	12.6%	3.64	21.51
FS664	227473	3.6%	218562	348	128	36.8%	3.42	10.71
FS673	261449	0.8%	258761	173	65	37.6%	2.99	11.08
FS674	57361	2.2%	55675	447	146	32.7%	4.60	43.81
RotaCTD09	75275	1.0%	73351	1527	*NA*	*NA*	3.42	6.03
FS734	23998	1.8%	22872	507	142	28.0%	4.27	28.74
FS737	100524	14.0%	85783	1211	431	35.6%	4.71	50.18
FS741	46324	1.5%	45411	675	193	28.6%	4.72	50.09
FS745	43459	3.2%	41008	117	24	20.5%	3.50	20.95
FS750	126897	2.9%	122430	1242	435	35.0%	4.49	38.15
FS753	156063	1.9%	152418	1052	338	32.1%	4.43	37.01
J800-10	1647	0.2%	1638	64	1	1.6%	2.77	9.41
J800-18	172550	12.5%	150054	807	349	43.2%	4.51	35.63
FS430	10092	1.9%	9848	85	10	11.8%	2.75	7.86
FS431	93776	22.4%	72374	289	81	28.0%	3.98	30.05
FS432	98087	77.3%	22122	130	41	31.5%	3.68	26.03
FS433	20067	12.8%	17446	166	42	25.3%	4.01	36.99
FS435	22899	3.0%	22023	209	*NA*	*NA*	3.93	30.48
J797-41	67806	28.1%	44425	278	86	30.9%	3.82	25.31
J801-37	16549	1.4%	16240	109	37	33.9%	3.08	13.06
J801-8	202651	33.3%	131031	133	50	37.6%	2.92	10.62

* High-quality reads, including metazoans, retained after initial quality filtering and chimera removal.

‡ Shaded protist V4 reads are included in analyses despite reduced numbers of sequence reads.

Table 2.6 Supplementary Table S4: Co-occurrences between protist and environmental variables

	Protist Group Classification Depth						
	Phylum	Class	Order	Family	Genus	R-value	Correlated physiochemical parameter
Alveolata	*Dinophyta*		Syndiniales (MALV-I, clade 4)			-0.62	Lithium
			Syndiniales (MALV-I, clade 6)			-0.62	Lithium
			Syndiniales (MALV-I, clade 7)			0.67	Sulfate
			Syndiniales (MALV-I, clade 8)			-0.61	Hydrogen Sulfide
			Syndiniales (MALV-II, clade 1)			0.64	pH
			Syndiniales (MALV-II, clade 1)		Amoebophrya	-0.65	Silica
			Syndiniales (MALV-II, clade 3)			-0.66	Methane
			Syndiniales (MALV-II, clade 4)			-0.60	Manganese
			Syndiniales (MALV-II, clade 15)			-0.60	Ammonium
			Syndiniales (MALV-II, clade 22)			-0.65	Lithium
						-0.60	Sodium
			Syndiniales (MALV-II, clade 30)			-0.67	Lithium
			Syndiniales (MALV-II, clade 35)			-0.67	Methane
						-0.61	Ammonium
						-0.62	Nitrate
						-0.60	Phosphate
			Syndiniales (MALV-IV)			-0.65	Iron
Opisthokonta	Basidiomycota	Tremellomycetes	Trichosporonales	Trichosporonaceae	Trichosporon	0.63	Depth
Rhizaria	*Cercozoa*	Novel clade Endo-5				0.64	Nitrate
	Radiolaria	Polycystinea	Collodaria			-0.60	Nitrate
	Radiolaria	Polycystinea	Collodaria	Collosphaeridae	Thalassicola	-0.64	Manganese
	Radiolaria	Polycystinea	Nassellaria			0.69	Hydrogen
	Radiolaria	Polycystinea	Nassellaria	Theoperidae	Pterocanium	0.65	Hydrogen
	Radiolaria	Polycystinea	Spumellaria	Ethmosphaeridae	Styptosphaera	-0.68	Manganese
	Radiolaria	Polycystinea	Spumellaria	Actinommidae	Astrosphaera	-0.59	Depth
						-0.65	Manganese
	Radiolaria	RAD-B, Group III				-0.67	Depth
						-0.72	Manganese
	Radiolaria	RAD-C				-0.69	Manganese
Stramenopila		MAST-4				-0.65	Ammonium
						-0.64	Nitrate
						-0.67	Phosphate
		Pelagophyceae				-0.62	Lithium
					Pirsonia	0.61	Hydrogen
	(unknown Eukarya)					0.68	Iron

Table 2.7 Supplementary Table S5: List of indicator organisms used in co-occurrence network

Reference node identity	Reference node number *	Extreme condition or habitat indicator
Archaea (Crenarchaeota)		
fa_Desulfurococcaceae	1	(hyper)thermophile
or_Desulfurococcales	2	(hyper)thermophile
or_Thermoproteales	3	(hyper)thermophile
cl_Thermoprotei	4	(hyper)thermophile
Archaea (Euryarchaeota)		
Archaeoglobus	5	(hyper)thermophile
fa_Archaeoglobaceae	6	(hyper)thermophile
Methanobacterium	7	methanogen
fa_Methanococcaceae	8	methanogen
cl_Methanomicrobia	9	methanogen
or_Methanofastidiosales	10	methanogen
Thermococcus	11	(hyper)thermophile
fa_Thermococcaceae	12	(hyper)thermophile
Aciduliprofundum	13	thermoacidophile
fa_Aciduliprofundaceae	14	thermoacidophile
fa_Thermoplasmataceae	15	thermoacidophile
or_Thermoplasmatales	16	thermoacidophile
cl_Thermoplasmata (MBG-D and DHVEG-1)	17	crustal/subsurface
Archaea *Candidatus* Bathyarchaeota	18	crustal/subsurface
Archaea *Candidatus* Geothermarchaeota	19	crustal/subsurface
Archaea *Candidatus* Lokiarchaeota	20	crustal/subsurface
Bacteria (Acidobacteria)		
Thermotomaculum	21	(hyper)thermophile
Bacteria (Epsilonbacteraeota)		
Arcobacter	22	vent indicator
Sulfurovum	23	vent indicator
Sulfuricurvum	24	vent indicator
Sulfurimonas	25	vent indicator
Thioreductor	26	vent indicator

*Refers to numbered reference nodes in co-occurrence network Figure 3b.

(Below)

Table S6. List of positive pairwise Spearman correlations (R-values > 0.5) between protist OTUs, grouped by taxonomic identity, and prokaryote OTUs from Huber et al 2010. Protist taxonomic identities in italics lack a rank designation in GenBank and are therefore placed in the seemingly appropriate column. Prokaryote taxa are identified to the genus level unless preceded by a 2-letter code indicating the depth of taxonomic identity: sg=supergroup, ph=phylum, cl=class, or=order, fa=family. Phyla (or proteobacterial classes) to which prokaryotic groups belong are in brackets, unless identity is already classified to that level. Those followed by "(*Cand*.)" indicate candidate phyla. Prokaryotic groups with known "extremophilic" members are in bold.

Table 2.8 Supplementary Table S6: Positive Spearman correlations

	Protist Group Classification Depth					No. of		
	Phylum	Class	Order	Family	Genus	OTUs	R-value	Correlated Prokaryote Group
Alveolata	(unknown Alveolata)					46	0.68	Pseudoalteromonas (Gammaproteobacteria)
	Ciliophora					4	0.61	or_Desulfovibrionales (Deltaproteobacteria)
							0.57	sg_Asgardaeota
	Dinophyta					208	0.59	Halomonas (Gammaproteobacteria)
	Dinophyta	Dinophyceae	Coccidiniales	Chytridiniaceae	Chytriodinium	1	0.79	fa_Piscirickettsiaceae (Gammaproteobacteria)
							0.59	fa_Halobacteriaceae (Euryarchaeota)
	Dinophyta	Dinophyceae	Gymnodiniales	Gymnodiniaceae	Gyrodinium	4	0.78	**ph_Bathyarchaeota (*Cand.*)**
							0.57	**fa_Thermococcaceae (Euryarchaeota)**
							0.57	sg_Asgardaeota
							0.57	ph_Thaumarchaeota
	Dinophyta		Syndiniales			34	0.58	fa_Piscirickettsiaceae (Gammaproteobacteria)
							0.56	or_Myxococcales (Deltaproteobacteria)
							0.56	**or_Desulfurococcales (Crenarchaeota)**
			Syndiniales (MALV-I)			68	0.67	Sphingobium (Alphaproteobacteria)
							0.58	**ph_Lokiarchaeota (*Cand.*)**
			Syndiniales (MALV-I, clade 2)			598	0.66	**Aciduliprofundum (Euryarchaeota)**
							0.64	**cl_Thermoprotei (Crenarchaeota)**
			Syndiniales (MALV-I, clade 3)			46	0.72	ph_Thaumarchaeota
							0.65	or_Myxococcales (Deltaproteobacteria)
			Syndiniales (MALV-I, clade 5)			252	0.60	**fa_Aciduliprofundaceae (Euryarchaeota)**
							0.59	fa_Burkholderiaceae (Betaproteobacteria)
			Syndiniales (MALV-I, clade 5)		Karlodinium	15	0.59	**or_Desulfurococcales (Crenarchaeota)**
							0.57	**fa_Thermococcaceae (Euryarchaeota)**
			Syndiniales (MALV-I, clade 6)			10	0.63	*Cand.* Pelagibacter (Alphaproteobacteria)
			Syndiniales (MALV-I, clade 7)			26	0.60	Nitrospina (Nitrospinae)
							0.58	SUP05 clade (Gammaproteobacteria)
			Syndiniales (MALV-I, clade 8)			16	0.60	**Thioreductor (Epsilonbacteraeota)**
			Syndiniales (MALV-II, clade 1)			160	0.60	**or_Thermoproteales (Crenarchaeota)**
							0.57	Burkholderia (Betaproteobacteria)
			Syndiniales (MALV-II, clade 3)			22	0.59	SAR86 clade (Gammaproteobacteria)
			Syndiniales (MALV-II, clade 4)			6	0.63	Pseudoalteromonas (Gammaproteobacteria)
							0.60	unknown Archaea
			Syndiniales (MALV-II, clade 5)			19	0.59	fa_Halobacteriaceae (Euryarchaeota)
							0.59	**cl_Thermoplasmata MBG-D & DHVEG-1 (Euryarchaeota)**
			Syndiniales (MALV-II, clades 10/11)			6	0.64	or_Myxococcales (Deltaproteobacteria)
			Syndiniales (MALV-II, clade 12)			8	0.73	**ph_Lokiarchaeota (*Cand.*)**
							0.67	**fa_Methanococcaceae (Euryarchaeota)**
							0.64	**cl_Methanomicrobia (Euryarchaeota)**
							0.64	**ph_Bathyarchaeota (*Cand.*)**
							0.63	Mariprofundus (Zetaproteobacteria)
							0.60	**or_Desulfurococcales (Crenarchaeota)**
							0.57	**Sulfuricurvum (Epsilonbacteraeota)**

	Protist Group Classification Depth					No. of		
	Phylum	Class	Order	Family	Genus	OTUs	R-value	Correlated Prokaryote Group
Alveolata *(continued)*			Syndiniales (MALV-II, clade 15)			28	0.60	fa_Flavobacteriaceae (Bacteroidetes)
							0.57	cl_Alphaproteobacteria
			Syndiniales (MALV-II, clade 16)			16	0.68	fa_Halobacteriaceae (Euryarchaeota)
							0.65	**Methanobacterium (Euryarchaeota)**
							0.64	fa_Piscirickettsiaceae (Gammaproteobacteria)
			Syndiniales (MALV-II, clade 17)			4	0.61	SUP05 clade (Gammaproteobacteria)
							0.59	Pseudoalteromonas (Gammaproteobacteria)
			Syndiniales (MALV-II, clade 26)			4	0.62	*Cand.* Pelagibacter (Alphaproteobacteria)
							0.57	**Thioreductor (Epsilonbacteraeota)**
			Syndiniales (MALV-II, clade 31)			5	0.57	**fa_Aciduliprofundaceae (Euryarchaeota)**
			Syndiniales (MALV-II, clade 35)			5	0.58	fa_Oceanospirillaceae (Gammaproteobacteria)
							0.57	Nitrososphaera (Thaumarchaeota)
			Syndiniales (MALV-II, clade 44)			4	0.74	**cl_Thermoplasmata MBG-D & DHVEG-1 (Euryarchaeota)**
							0.65	or_Myxococcales (Deltaproteobacteria)
							0.57	unknown Archaea
							0.56	**ph_Lokiarchaeota (*Cand.*)**
			Syndiniales (MALV-III)			35	0.59	cl_Gammaproteobacteria
			Syndiniales (MALV-IV)			25	0.58	**Thioreductor (Epsilonbacteraeota)**
			Syndiniales (MALV-V)			14	0.58	Labrenzia (Alphaproteobacteria)
							0.56	*Cand.* Pelagibacter (Alphaproteobacteria)
	Apuozoa			Apusomonadidae	Amastigomonas	2	0.64	**Sulfuricurvum (Epsilonbacteraeota)**
							0.62	**Thermococcus (Euryarchaeota)**
							0.59	**ph_Lokiarchaeota (*Cand.*)**
							0.59	**cl_Thermoplasmata MBG-D & DHVEG-1 (Euryarchaeota)**
Incertae sedis	*Centroheliozoa*					6	0.67	Pseudoalteromonas (Gammaproteobacteria)
							0.62	fa_Piscirickettsiaceae (Gammaproteobacteria)
							0.61	**Methanobacterium (Euryarchaeota)**
	Haptophyta	Haptophyceae	Prymnesiales			3	0.71	Albirhodobacter (Alphaproteobacteria)
							0.69	fa_Rhodanobacteraceae (Gammaproteobacteria)
							0.65	Thalassospira (Alphaproteobacteria)
							0.60	Sphingomonas (Alphaproteobacteria)
							0.59	Ilumatobacter (Actinobacteria)
	Haptophyta	Haptophyceae	Prymnesiales	Prymnesiaceae		3	0.70	**Thioreductor (Epsilonbacteraeota)**
							0.57	**Methanobacterium (Euryarchaeota)**
	Picozoa					12	0.65	Sphingobium (Alphaproteobacteria)
							0.63	or_Xanthomonadales (Gammaproteobacteria)
							0.59	Nitrososphaera (Thaumarchaeota)
	(unknown Opisthokonta)					7	0.70	cl_Thermoplasmata (Euryarchaeota)
							0.66	sg_Asgardaeota
							0.62	or_Bacteroidales (Bacteroidetes)
							0.57	**ph_Lokiarchaeota (*Cand.*)**
	Basidiomycota	Tremellomycetes	Trichosporonales	Trichsporonaceae	Trichosporon	8	0.59	Burkholderia (Betaproteobacteria)
	Basidiomycota	Tremellomycetes				1	0.58	*Cand.* Nitrosopelagicus (Thaumarchaeota)

	Protist Group Classification Depth					No. of		
	Phylum	Class	Order	Family	Genus	OTUs	R-value	Correlated Prokaryote Group
Rhizaria	*Cercozoa*	Novel clade Endo-5				2	0.67	**or_Thermoproteales (Crenarchaeota)**
							0.61	fa_Rubritaleaceae (Verrucomicrobia)
							0.59	fa_Flavobacteriaceae (Bacteroidetes)
							0.58	Rubritalea (Verrucomicrobia)
	Cercozoa	Novel clade Endo-4				15	0.69	or_Chromatiales (Gammaproteobacteria)
							0.57	or_Myxococcales (Deltaproteobacteria)
							0.57	**or_Thermoplasmatales (Euryarchaeota)**
	Radiolaria					56	0.60	*Cand.* Pelagibacter (Alphaproteobacteria)
	Radiolaria	Polycystinea	Collodaria			24	0.74	fa_Piscirickettsiaceae (Gammaproteobacteria)
							0.61	**cl_Thermoplasmata MBG-D & DHVEG-1 (Euryarchaeota)**
							0.56	cl_Campylobacteria (Epsilonbacteraeota)
	Radiolaria	Polycystinea	Collodaria	Collosphaeridae	Thalassicolla	2	0.58	fa_Piscirickettsiaceae (Gammaproteobacteria)
							0.58	*Cand.* Pelagibacter (Alphaproteobacteria)
	Radiolaria	Polycystinea	Nassellaria			79	0.72	SUP05 clade (Gammaproteobacteria)
							0.57	**Methanobacterium (Euryarchaeota)**
	Radiolaria	Polycystinea	Nassellaria	Theoperidae	Eucyrtidium	4	0.74	fa_Piscirickettsiaceae (Gammaproteobacteria)
							0.67	fa_Halobacteriaceae (Euryarchaeota)
							0.60	**ph_Lokiarchaeota (*Cand.*)**
							0.57	**Methanobacterium (Euryarchaeota)**
	Radiolaria	Polycystinea	Nassellaria	Theoperidae	Pterocanium	14	0.76	SUP05 clade (Gammaproteobacteria)
							0.60	Elusimicrobium (Elusimicrobia)
	Radiolaria	Polycystinea	Nassellaria	Trissocyclidae	Ceratospyris	12	0.71	**ph_Bathyarchaeota (*Cand.*)**
							0.63	cl_Thermoplasmata (Euryarchaeota)
							0.60	**ph_Lokiarchaeota (*Cand.*)**
	Radiolaria	Polycystinea	Spumellaria			298	0.59	**fa_Thermoplasmataceae (Euryarchaeota)**
	Radiolaria	Polycystinea	Spumellaria	Actinommidae	Arachnosphaera	19	0.72	Pseudoalteromonas (Gammaproteobacteria)
							0.67	**Methanobacterium (Euryarchaeota)**
							0.63	Thiomicrospira (Gammaproteobacteria)
							0.60	**Thioreductor (Epsilonbacteraeota)**
							0.57	or_Alteromonadales (Gammaproteobacteria)
	Radiolaria	Polycystinea	Spumellaria	Actinommidae	Astrosphaera	54	0.79	fa_Piscirickettsiaceae (Gammaproteobacteria)
							0.67	fa_Halobacteriaceae (Euryarchaeota)
	Radiolaria	Polycystinea	Spumellaria	Ethmosphaeridae	Styptosphaera	4	0.75	fa_Halobacteriaceae (Euryarchaeota)
							0.67	fa_Piscirickettsiaceae (Gammaproteobacteria)
							0.63	**cl_Thermoplasmata MBG-D & DHVEG-1 (Euryarchaeota)**
	Radiolaria	Polycystinea	Spumellaria	Spongodiscidae	Spongopyle	2	0.61	Albirhodobacter (Alphaproteobacteria)
							0.61	Vibrio (Gammaproteobacteria)
	Radiolaria	Polycystinea	Spumellaria	Spongodiscidae/Coccodiscidae		37	0.83	fa_Piscirickettsiaceae (Gammaproteobacteria)
							0.70	fa_Halobacteriaceae (Euryarchaeota)
							0.61	**Methanobacterium (Euryarchaeota)**
	Radiolaria	RAD-A				56	0.70	**Arcobacter (Epsilonbacteraeota)**
							0.70	fa_Helicobacteraceae (Epsilonbacteraeota)
							0.66	**cl_Thermoplasmata MBG-D & DHVEG-1 (Euryarchaeota)**

	Protist Group Classification Depth					No. of		
	Phylum	Class	Order	Family	Genus	OTUs	R-value	Correlated Prokaryote Group
Rhizaria *(continued)*	Radiolaria	RAD-A *(continued)*					0.65	**Sulfurovum (Epsilonbacteraeota)**
							0.62	fa_Saprospiraceae (Bacteroidetes)
							0.58	Roseobacter (Alphaproteobacteria)
							0.58	Campylobacter (Epsilonbacteraeota)
							0.58	Cocleimonas (Gammaproteobacteria)
							0.57	**Thioreductor (Epsilonbacteraeota)**
	Radiolaria	RAD-B				24	0.60	Pseudoalteromonas (Gammaproteobacteria)
							0.57	**cl_Methanomicrobia (Euryarchaeota)**
	Radiolaria	RAD-B, Group I				116	0.56	or_Myxococcales (Deltaproteobacteria)
	Radiolaria	RAD-B, Group III				2	0.83	fa_Piscirickettsiaceae (Gammaproteobacteria)
							0.65	fa_Halobacteriaceae (Euryarchaeota)
	Radiolaria	RAD-C				20	0.75	ph_Thaumarchaeota
							0.64	or_Myxococcales (Deltaproteobacteria)
Stramenopila	(unknown Stramenopiles)					80	0.72	Nitrososphaera (Thaumarchaeota)
		Labyrinthulomycetes	Labyrinthulales	Labyrinthulaceae		5	0.60	ph_Thaumarchaeota
							0.58	fa_Piscirickettsiaceae (Gammaproteobacteria)
		MAST-1				7	0.63	fa_Rhodanobacteraceae (Gammaproteobacteria)
							0.58	*Cand.* Pelagibacter (Alphaproteobacteria)
		MAST-3				27	0.62	fa_Piscirickettsiaceae (Gammaproteobacteria)
							0.59	**cl_Thermoplasmata MBG-D & DHVEG-1 (Euryarchaeota)**
							0.57	fa_Halobacteriaceae (Euryarchaeota)
		MAST-4				4	0.59	Thalassospira (Alphaproteobacteria)
							0.58	Pseudoalteromonas (Gammaproteobacteria)
		MAST-7				8	0.63	*Cand.* Pelagibacter (Alphaproteobacteria)
							0.58	or_Xanthomonadales (Gammaproteobacteria)
		MAST-8				9	0.64	*Cand.* Pelagibacter (Alphaproteobacteria)
							0.58	fa_Halobacteriaceae (Euryarchaeota)
					Pirsonia	9	0.81	or_Desulfobacterales (Deltaproteobacteria)
							0.64	**cl_Methanomicrobia (Euryarchaeota)**
							0.58	**Sulfurimonas (Epsilonbacteraeota)**

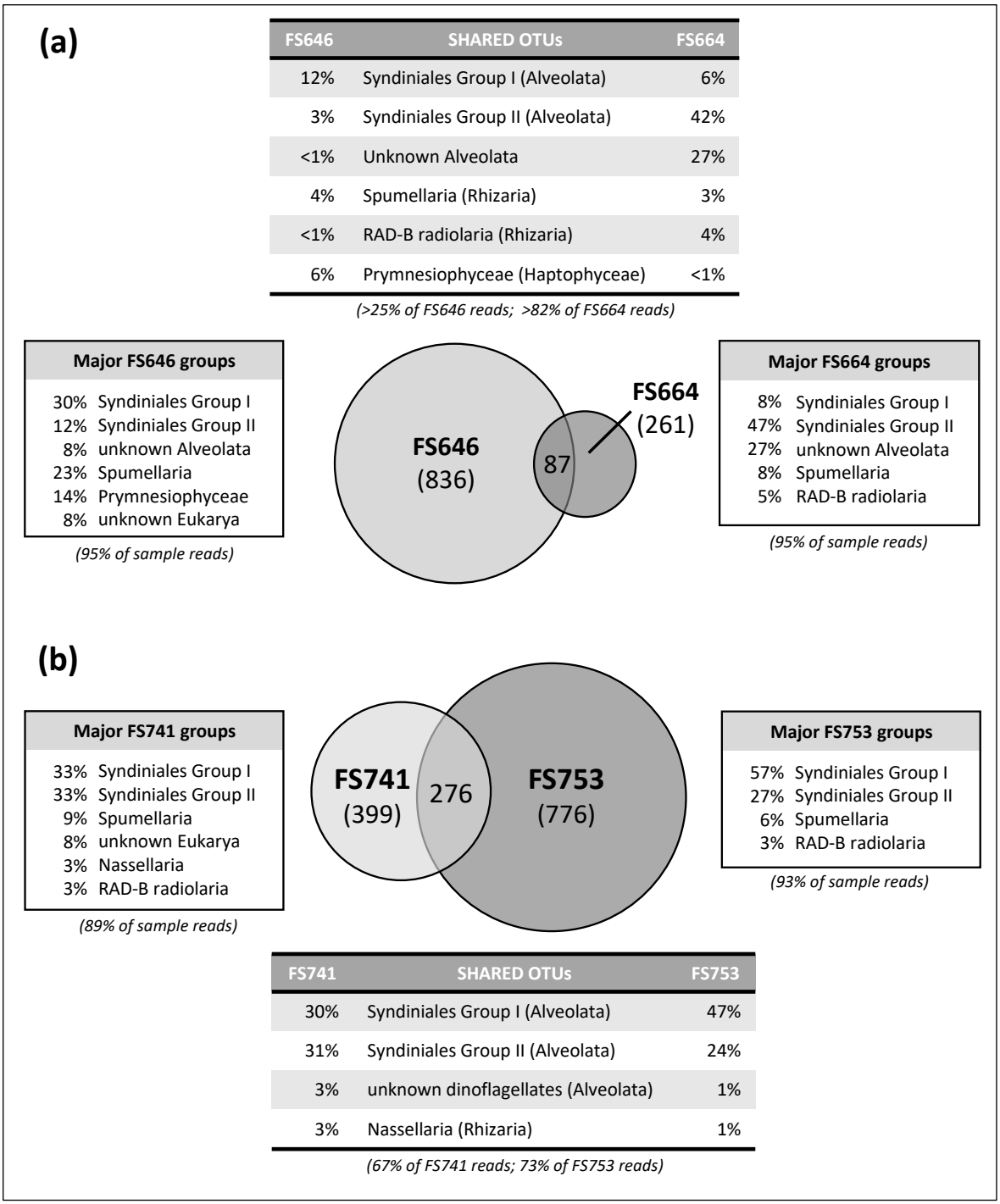


Figure 2.8 Supplementary Figure S3: Close proximity pairs comparison

Samples collected in 2009 **(a)** and 2010 **(b)**. Percentages are given as proportion of sequence reads recovered from that sample.

Chapter 3. Emergent 'core communities' of microbes, meiofauna and macrofauna at hydrothermal vents

Chapter 3 is a manuscript that has been accepted for publication in the open-access journal *ISME Communications.* The present version differs slightly from what was included in the original dissertation submission defended on 8 June 2021. Differences from the defended version include clarification of the two major objectives of the study and a restructuring of the discussion to more clearly reflect those objectives. Sheryl Murdock produced all microbial data, tables and figures and performed all analyses. Faunal species identifications and counts were provided by Rachel Boschen-Rose, Verena Tunnicliffe, and their associated lab members. Manuscript text was written by Sheryl Murdock with editorial contributions and direction from Kim Juniper and Verena Tunnicliffe. Due to word limitations for *ISME Communications*, many of the methodological details and more minor results reside in the article supplementary.

Abstract

Assessment of ecosystem health entails consideration of species interactions within and between size classes to determine their contributions to ecosystem function. Elucidating microbial involvement in these interactions requires tools to distill diverse microbial information down to relevant, manageable elements. We used covariance ratios (proportionality) between pairs of species and patterns of enrichment to identify 'core communities' of likely interacting microbial (<64μm), meiofaunal (64μm to 1mm) and macrofaunal (>1mm) taxa within assemblages hosted by a foundation species, the

hydrothermal vent tubeworm *Ridgeia piscesae*. Compared with samples from co-located hydrothermal fluids, microbial communities within *R. piscesae* assemblages are hotspots of taxonomic richness and are high in novelty (unclassified OTUs) and in relative abundance of Bacteroidetes. We also observed a robust temperature-driven distinction in assemblage composition above and below ~25°C that spanned micro to macro size classes. The core high-temperature community included eight macro- and meiofaunal taxa and members of the Bacteroidetes and Epsilonbacteraeota, particularly the genera *Carboxylicivirga*, *Nitratifractor* and *Arcobacter*. The core low-temperature community included more meiofaunal species in addition to Alpha- and Gammaproteobacteria, and Actinobacteria. Inferred associations among high-temperature core community taxa suggest increased reliance on species interactions under more severe hydrothermal conditions. We propose refinement of species diversity to 'core communities' as a tool to simplify investigations of relationships between taxonomic and functional diversity across domains and scales by narrowing the taxonomic scope.

3.1 Introduction

From microbes to megafauna, interactions among species and with their environments shape biological assemblages (Wisz et al., 2013) and, by extension, ecosystem function. Disentangling the outcomes of interactions can be a monumental task, particularly when attempting to include interactions involving the staggering diversity of microorganisms (Thompson et al., 2017). Growing recognition of the vital roles microorganisms play in marine ecosystems (Cho and Azam, 1988; Rousk and Bengtson, 2014) underscores the importance of integrating microbial and higher organism relationships when characterizing

taxonomic and functional communities. A fundamental knowledge of the core players in these communities and how they interact can guide effective management strategies where human activities may compromise the ecosystem (Sherman et al., 2005; Guilhon et al., 2020).

A multi-scale approach to ecosystem characterization (i.e., one that incorporates multiple organismal size classes) can reveal covariances among organisms with direct links to shifts in ecosystem properties. For example, in marine sediments, multi-scale approaches revealed links between shifts in microbial abundance, faunal diversity, and ecosystem properties including biomass/productivity ratios, trophic structure, and nitrogen cycling (Passarelli et al., 2012; Baldrighi et al., 2013; Foshtomi et al., 2015). In coastal ecosystems, the numerous and varied contributions of microphytobenthos to functionality (Hope et al., 2020) illustrate the importance of multi-scale interactions and feedbacks within ecological networks, and the vital roles of microbial players. However, moving from broad links between size classes to identifying species-level interactions requires methods for distilling the vast diversity of microbes down to those that are most likely interacting beyond the microbial realm.

One useful, first-order approach for identifying candidate multi-scale and interspecies associations is the analysis of co-occurrence patterns (Lima-Mendez et al., 2015). However, the use of co-occurrences to infer ecological interactions has been questioned (Blanchet et al., 2020), mainly because of inappropriate data handling and the potential for spurious correlations (Pearson, 1897; Jackson, 1997; Gloor and Reid, 2016). Proportionality, which considers covariance of species or OTU counts transformed to ratios (Lovell et al., 2015), is an appropriate alternative for inferring associations. Covariance of microbial and faunal species may signal shared environmental preferences/tolerances or

key inter-organism interactions (e.g., food webs, symbioses) and, therefore, be used to focus exploration of functional interactions on those species with the greatest potential importance to the ecosystem.

Hydrothermal vent ecosystems, where diverse non-photosynthetic prokaryotes are the exclusive primary producers, offer a potentially instructive case study for developing new approaches to integrating microbial and macrofaunal ecology. Microbial chemosynthetic primary production underpins the formation and maintenance of high-biomass faunal assemblages at low to moderate temperature (<100°C, "diffuse") hydrothermal vents (Sievert and Vetriani, 2012) that result from mixing of high-temperature hydrothermal fluids and cold seawater within oceanic crust. These diffusively discharging hydrothermal fluids carry abundant microbes from subseafloor chemosynthetic primary production (Karl et al., 1980; Huber et al., 2002; Meyer et al., 2013) and fuel additional chemosynthesis at and above the seafloor (Orcutt et al., 2011). The resulting microbial biomass and metabolisms support grazing and deposit/suspension feeding animals and symbioses with animal partners (Dubilier et al., 2008). Current understanding of hydrothermal vent fauna and microbes reflects more than four decades of research that has, with the exception of symbiotic associations, largely treated them as separate entities. Trophodynamics studies at vents have generalized food web links using stable isotope ratios (Van Dover and Fry, 1989; Colaço et al., 2002; Bergquist et al., 2007; Yamanaka et al., 2015) and lipid profiles (Van Dover, 2002; Colaço et al., 2007; Limen et al., 2008) but have addressed neither the potential for incidental supply of organic material from fauna to support microbial heterotrophy nor the metabolic energy dynamics among microbes. Furthermore, these studies have not examined the composition or diversity of microbial food

sources, only their biogeochemical signatures. Coordinated analyses of species-level variation in cohabiting vent fauna and microbes could be used to reveal potentially important interspecies associations beyond microbial-to-faunal food web links.

Habitats structured by foundation species increase physical stability and biodiversity (Dayton, 1972; Bruno and Bertness, 2001; Lamy et al., 2020) and, by extension, have greater potential for functional interactions across taxonomic groups. Siboglinid tubeworms, common foundation species in chemosynthesis-based vent habitats, physically augment access to chemosynthetic energy resources for microbes, and provide habitat structures for abundant associated fauna at density and biomass levels that rival other highly productive marine ecosystems (see review, (Tunnicliffe and Cordes, 2020)). Within tubeworm aggregations, the tube structures and their adhering microbial biofilms (López-García et al., 2002; Pagé et al., 2004; Rincon-Tomas et al., 2020) slow the dispersion of hydrothermal fluids and so control physico-chemical gradients for other metazoa and microbes (Govenar, 2010). These broad co-dependencies provide direction for exploring potential functional relationships of micro- and macro-scale community members within the entire assemblage.

Ridgeia piscesae is a foundation siboglinid tubeworm species at vents on the mid-ocean ridges of the NE Pacific. The wide range of physico-chemical conditions occupied by *R. piscesae* (Tunnicliffe et al., 2014) provides different habitats for associated faunal assemblages that vary in accordance with diffuse fluid flux and sulfide concentrations (Sarrazin and Juniper, 1999; Sarrazin et al., 1999; Govenar et al., 2002). Non-endosymbiotic microbial composition across the tubeworm's habitat range is less defined, although preliminary characterization spanning parts of this range suggests predictable patterns (Forget and Juniper, 2013). Full characterization of the microbiomes within these faunal

assemblages also requires comparison with microbiomes of the hydrothermal fluids and seawater that flow through them. An integrated study of microbes and fauna in *Ridgeia*-hosted assemblages can advance understanding of how faunal assemblages contribute to augmenting microbial taxonomic diversity at vents — a relevant goal given the importance of biodiversity to ecosystem stability in the deep-sea (Danovaro et al., 2008). Recent assessments of hydrothermal vent microbiomes reveal a tendency to consider only symbiotic microbes within faunal assemblages (Dick, 2019; Lee et al., In press), ignoring the potential selective or enriching effects of habitat conditions on non-endosymbiotic microbial composition within faunal assemblages.

Our objective is to develop and evaluate an approach for identifying 'core communities' of likely interacting organisms within the diverse assemblages associated with hydrothermal vent tubeworm habitat. In particular, the study uses species-level diversity, enrichment and covariance to identify whether such a community transcends micro, meio, and macro size classes spanning a range of *R. piscesae* habitat conditions. Additionally, we assess tubeworm-assemblage microbial richness relative to corresponding hydrothermal fluids and background seawater to evaluate how faunal assemblages contribute to shaping microbial taxonomic diversity within the hydrothermal vent ecosystem. This detailed structural characterization of a hydrothermal vent microbiome represents an essential, yet nontrivial, first step toward identification of functional interactions with relevance to the ecosystem.

3.2 Methods

To remove ambiguity in the use of terms, we define the following: *assemblage* - cohabiting

species whose level of interaction is unknown, *community* - cohabiting, covarying species that likely interact directly and may respond collectively to system changes. Methods are briefly described here with additional details in *Supplementary Material*.

3.2.1 Sample collection

Microbial and faunal assemblages associated with the tubeworm *Ridgeia piscesae* (*Supplementary Figure 1*) were sampled over a range of vent discharge intensities and supporting substrata (basalt and sulfide) using a manipulator arm on the remotely-operated vehicle ROPOS. These tubeworm grab samples came from vents on two segments of the Juan de Fuca Ridge (NE Pacific) - Endeavour (Main Endeavour & Clam Bed fields) and Middle Valley (Dead Dog and Bent Hill fields). Prior to collecting seven of the thirteen samples, we used the ROPOS suction sampler to collect 2 Litres of hydrothermal fluids venting through the sampled tubeworm bushes. Background seawater was collected from each vent field in 5L Niskin bottles. For all tubeworm grab samples, we first measured fluid temperatures inside the bush at the base. Details of sample location and habitat are in Table 3.1.

Shipboard, approximately half of each tubeworm grab sample was placed in a bucket of cold (4˚C) artificial seawater and held at 4˚C until processing of the microbial component. The remaining half was preserved in either 75% ethanol or 7% buffered formalin (*Supplementary Figure 2*). Later, tubeworms were removed from the preserved half, and the residue sieved to separate macrofauna (>1mm) from meiofauna (>64μm to <1mm). See *Supplementary Material* for additional details of sample collection and faunal characterization. The microbial component associated with each grab sample was detached either by gently agitating the worm tubes in the bucket or, when possible, directly removing masses of adhering biofilm into a sterile 50ml tube and agitating; thus, internal worm

Table 3.1 Sample information.

BOLD sample names were characterized form micro to macro

			Location			Temperature (°C)		Balance of microbial domains ‡			Observed faunal richness		
Sample name	Sample Type	Substratum	Vent Field	Latitude (DecDeg)	Longitude (DecDeg)	Worm base	Fluid above worms	% Bacteria (± std err)	% Archaea (± std err)	% Bacteria (± std err)	No. meiofaunal species	No. macrofaunal species	Total No. unique species†
EMw1	Tubeworm bush	Sulfide	Main Endeavour	47.9501	-129.0971	33.2		66.3 ± 2.6	27.3 ± 2.1	6.5 ± 0.5	5	7	11
EMw2	Tubeworm bush	Sulfide	Main Endeavour	47.9501	-129.0961	5.3		87.2 ± 0.2	8.2 ± <0.1	4.7 ± 0.2	na	na	na
EMw3	Tubeworm bush	Basalt	Main Endeavour	47.9500	-129.0971	3.4		82.1 ± 1.2	9.5 ± 0.4	8.4 ± 0.9	18	21	34
EMw4	Tubeworm bush	Sulfide	Main Endeavour	47.9500	-129.0970	37.1		98.2 ± <0.1	0.2 ± <0.1	1.6 ± <0.1	6	9	13
EMw5	Tubeworm bush	Sulfide	Main Endeavour	47.9500	-129.0969	13.4		83.4 ± 0.6	14.8 ± 0.6	1.8 ± <0.1	17	11	24
EMw6	Tubeworm bush	Basalt	Main Endeavour	47.9501	-129.0972	2.8		83.3 ± 0.7	4.8 ± 0.3	11.9 ± 0.8	na	na	na
EMw7	Tubeworm bush	Sulfide	Main Endeavour	47.9497	-129.0983	33.0 [a]		99.0 ± <0.1	0.1 ± <0.1	0.9 ± <0.1	na	na	na
EMw8	Tubeworm bush	Sulfide	Main Endeavour	47.9495	-129.0985	37.4		95.9 ± 0.3	3.4 ± 0.3	0.7 ± <0.1	4	8	11
ECw9	Tubeworm bush	Sulfide	Clam Bed	47.9630	-129.0915	15.0		34.7 ± 0.3	63.2 ± 0.3	2.2 ± <0.1	na	na	na
ECw10	Tubeworm bush	Sulfide	Clam Bed	47.9630	-129.0915	31.0		95.5 ± <0.1	1.0 ± 0.1	3.5 ± 0.1	4	6	8
ECw11	Tubeworm bush	Basalt	Clam Bed	47.9631	-129.0917	5.2		93.9 ± <0.1	1.6 ± 0.1	4.5 ± 0.1	17	23	32
MVw12	Tubeworm bush	Sulfide	Middle Valley	48.4303	-128.6821	27.4		98.1 ± <0.1	0.8 ± <0.1	1.1 ± 0.1	1	5	6
MVw13	Tubeworm bush	Sulfide	Middle Valley	48.4553	-129.7086	21.1		84.5 ± 0.8	6.9 ± 0.8	8.5 ± <0.1	8	23	28
EMf1	Diffuse fluid	na	Main Endeavour	47.9501	-129.0971		5.0	92.0 ± 0.1	5.4 ± 0.1	2.6 ± 0.2			
EMf2	Diffuse fluid	na	Main Endeavour	47.9501	-129.0961		4.6	86.9 ± 0.7	3.1 ± 0.3	10.0 ± 0.4			
EMf3	Diffuse fluid	na	Main Endeavour	47.9500	-129.0971		4.1	75.2 ± 0.9	13.0 ± 0.3	11.8 ± 1.2			
EMf7	Diffuse fluid	na	Main Endeavour	47.9497	-129.0983		12.0	94.3 ± 0.3	1.4 ± <0.1	4.3 ± 0.3			
ECf9	Diffuse fluid	na	Clam Bed	47.9630	-129.0915		2.6	76.8 ± 0.7	12.5 ± 0.1	10.7 ± 0.6			
ECf11	Diffuse fluid	na	Clam Bed	47.9631	-129.0917		6.1	86.6 ± 0.6	7.0 ± 0.1	6.4 ± 0.6			
MVf12	Diffuse fluid	na	Middle Valley	48.4303	-128.6821		23.0	95.8 ± 0.2	3.3 ± 0.2	0.9 ± <0.1			
bkEC	Background fluid	na	Clam Bed	47.9631	-129.0915		na	75.9 ± 8.5	13.7 ± 3.5	10.4 ± 0.4			
bkEM	Background fluid	na	Main Endeavour	47.9501	-129.0972		na	74.4 ± 0.6	15.5 ± 0.6	10.1 ± 1.1			
bkMV	Background fluid	na	Middle Valley	48.4303	-128.6821		na	89.2 ± 9.5	4.6 ± 0.6	6.2 ± 0.2			

[a] Minimum recorded temperature. Subsequent slow increase suggested probe penetration into the substratum.

‡ Determined by quantitative PCR.

† Unique means that juveniles of species occurring in the macrofaunal fraction were not included in this count.

na = not available/not applicable

symbionts were not included. Two samples were subjected to both treatments to assess potential methodological bias (method replicates). For both methods, bucket water or biofilm was then passed through a 64µm sieve to remove meiofauna and serially filtered onto 20µm polycarbonate (micro size-fraction) and 0.2µm Sterivex filters (pico/nano size-fraction) for DNA extraction. Cells from vent and background fluids were collected on 0.2µm Sterivex filters onboard the ship using a peristaltic pump. All filters were stored at -80°C until further processing.

3.2.2 Microbial DNA extraction and sequencing

Following Sogin *et al*. (Sogin et al., 2006), we extracted DNA from filtered cells associated with seven diffuse and three background fluid samples, and 13 tubeworm grab samples that included micro and pico/nano size fractions, plus two methodological replicates in each size fraction. Extractions included a sterile filter as a control. DNA quality and concentrations were measured on a NanoDrop 1000 (Thermo Scientific). Paired-end sequencing (2x300bp) of 16S/18S rRNA genes from 40 DNA extracts was completed on Illumina MiSeq at either the Laboratory for Advanced Genome Analysis (Vancouver Prostate Centre) or the Integrated Microbiome Resource Facility (Dalhousie University), using primers targeting Bacteria, Archaea, and microeukaryotes.

Paired-end sequence reads were merged and quality filtered using the *iu-merge-pairs* script from the Illumina-Utils package (Eren et al., 2013) with the *–enforce-Q30-check* flag and *min-qual-score* set to 20. Chimera removal, clustering into operational taxonomic units (OTUs), and taxonomic identification was performed using Mothur v1.42.3 (Schloss et al., 2009) following the 'MiSeq SOP' analysis example (https://mothur.org/wiki/MiSeq_SOP,

accessed July 2019). Sequences were compared using the average neighbor pairwise distance method and clustered into OTUs using the standard 97% similarity threshold for bacteria and archaea and 98% for microeukaryotes (Caron et al., 2009b). Taxonomy was assigned to OTUs using the silva_nr_132 reference database (Quast et al., 2013) for bacteria and archaea and the PR2 v4.11.1 reference database (Guillou et al., 2013) for microeukarya. Singleton OTUs, those with only one sequence in the entire dataset, were removed from further analysis.

3.2.3 Quantitative PCR and microbial OTU scaling

Relative abundances of bacteria, archaea and microeukarya in each sample were determined by quantitative PCR (qPCR) of 16S and 18S rRNA genes. qPCR standards were created from a combination of clones for each microbial domain and used in ten-fold dilution series to create standard curves. Sample reactions were performed in triplicate on a CFX96 Real-Time PCR Detection System (Bio-Rad), and each run included a standard curve and a no-template control. Samples were analyzed with DNA from the micro and pico/nano size fractions combined. Gene copies per ng of DNA were calculated based on concentrations of the DNA extracts.

OTU counts were converted to relative abundances within each domain and then combined and scaled according to the qPCR results. Each OTU was then expressed as abundance relative to all OTUs. This approximation of the natural microbial assemblage was subjected to the same tests as was each domain separately.

3.2.4 Microbial composition and diversity

Data analysis followed the compositional approach as recommended by Gloor and Reid (2016) and Quinn *et al.* (Quinn et al., 2019). Zero counts were replaced by imputed values using the count zero multiplicative method in the zCompositions R package (Martín-Fernández et al., 2011; Palarea-Albaladejo and Martin-Fernandez, 2015), and data were transformed by centered log-ratio (Aitchison, 1986). To identify patterns in microbial assemblage composition using Analysis of Similarities (ANOSIM), hierarchical clustering and Nonmetric Multidimensional Scaling (NMDS), we calculated Aitchison distances between samples (Aitchison et al., 2000) using the coda.base R package (Comas-Cufí, 2019).

Diversity calculations and compositional difference tests were performed using the vegan R package (Oksanen et al., 2015). Diversity was calculated for each DNA extract and sequenced domain using the Inverse Simpson metric after singleton removal. The function ***metaMDS*** was used to create NMDS plots based on Aitchison distance matrices, and the ***envfit*** function was used to test for significant effects of temperature, substratum and location on NMDS ordinations. The ***hclust*** function was used for hierarchical cluster analysis.

ANOSIM was performed to identify potential biases introduced in our sampling methods and differences between samples related to environmental parameters or sample type. Details of bias tests are in *Supplementary Material*. Tests were performed on individual microbial domains and on the combined qPCR-balanced microbial assemblage. We compared compositions between samples from different substrata (sulfide, basalt), locations (Main Endeavour, Clam Bed, Middle Valley) and sample types (tubeworm grab, diffuse fluid, background fluid). OTUs responsible for significant compositional differences were identified using the ALDEx2 Bioconductor package v1.6.0 (Fernandes et al., 2013) and

we report those with expected effect size differences ≥1, indicating relative enrichment in one category over the other.

3.2.5 Size class congruence and taxonomic covariance

Using species and OTU counts for the macro, meio and micro size classes from nine grab samples, we explored covariance of species associated with *R. piscesae* across its range of habitat conditions. Microbes were assessed as both individual domains and the qPCR-balanced microbial assemblage. Prior to NMDS and hierarchical cluster analysis, macro- and meiofaunal species counts were treated with the same compositional data transformations as the microbial data (see above). Procrustes analysis (function ***procrustes***) on pairs of NMDS ordinations determined the level of congruence between size classes and microbial domains. Significance of the results was determined using the function ***protest*** with 1000 permutations. Both functions applied a symmetric analysis.

To identify covariance in ratios between microbial taxa and macro- and meiofaunal species, we performed proportionality analysis using a symmetric modification of the ρ metric in the propr R package (Quinn et al., 2017). Positive and negative ρ values between pairs of taxa indicate coincident occurrence or niche separation. Microbial OTUs with shared taxonomic identity (as assigned by Mothur) were aggregated into 719 unique taxonomic identities. To minimize the false discovery rate, taxa with counts of <2 in seven or more samples were excluded, leaving 335 microbial taxa, 14 macrofaunal and 17 meiofaunal species that were subjected to proportionality analysis. Taxon pairings with absolute ρ values >0.75 were plotted using Cytoscape v3.7.2 (Shannon et al., 2003).

3.3 Results

3.3.1 Composition and diversity

We produced 15.4 million (M) paired-end sequence reads of 16S and 18S rRNA genes from three microbial domains. Merged and quality filtered sequences consisted of 1.8 M bacterial, 2.7 M archaeal, and 2.3 M microeukaryal reads that clustered into 40 389 bacterial, 3 436 archaeal, and 16 731 microeukaryal nonsingleton OTUs. Sample ECw11 produced <100 archaeal reads that were excluded from further analyses.

Inverse Simpson diversity showed significant differences (Mann-Whitney tests) between microbes in fluid (background and diffuse) and grab samples; similarly, differences arose for microbes and fauna between grab samples with basal temperatures above and below *ca.* 25°C, hereafter termed 'highT grabs' (27-37°C) and 'lowT grabs' (3-21 °C; Figure 3.1). Bacterial diversity was greatest in lowT grabs and spanned a broad range of values, while in Archaea, diversity was greater in highT grabs. Microeukaryote and meiofaunal diversities were elevated in lowT versus highT grabs. Macrofaunal diversity did not differ between highT and lowT grabs (Figure 3.1), although lowT samples had a wider range of diversity values, and highT showed consistent evenness values.

Bacteria dominated the microbial fraction in qPCR analysis (Table 3.1), with values ranging from 66-99% of the microbial 16S/18S copies for most samples. One anomalous lowT grab sample from Clam Bed (ECw9) was dominated by archaea (63% of 16S/18S copies). Otherwise, the average abundances of archaea and microeukarya were 7.2% and 5.8%, respectively, but ranged from below 1% up to 27% for archaea and near 12% for microeukarya. The balanced microbial assemblages reflecting these proportions were used for further analyses. For taxonomic breakdown of sequences before q-PCR balancing, see

Supplementary Figure 3.

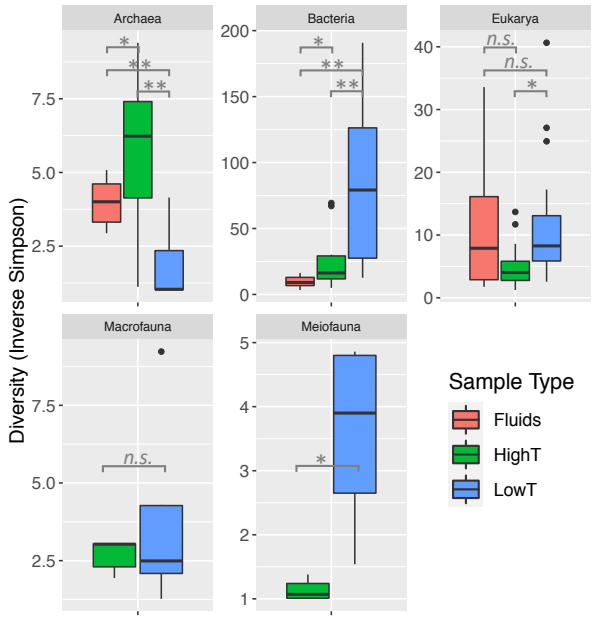


Figure 3.1 Inverse Simpson diversity

Diversity of microbes (all 40 DNA extracts, post removal of singleton OTUs), macrofauna and meiofauna (nine grab samples). Grab samples are divided into 'highT' and 'lowT' (basal temperature above and below 25°C, respectively). Significance values of Mann-Whitney tests: $p<0.01$ (**), $p<0.05$ (*), not significant (*n.s.*). Note varying y-axis scales.

Tests of potential biases on microbial composition introduced by our sample processing methods indicated minimal effects (see *Supplementary Materials).* Data from the two size fractions from each sample were therefore combined, as were reads from the two harvesting methods for samples EMw1 and EMw6.

The qPCR-balanced microbial assemblages of diffuse fluid and background seawater samples grouped together in NMDS and hierarchical clustering analyses (Figure 3.2a, *Supplementary Figure 4*). Compositional differences between these two fluid types are well-known (e.g., (Huber et al., 2002, 2003; Akerman et al., 2013)) and, therefore, are not discussed here. HighT grabs from Main Endeavour and Clam Bed clustered separately from lowT grabs. Among the tested environmental variables (temperature, substratum, location,

sample type), only temperature explained variation in the NMDS (envfit r^2=0.34, p=0.017). NMDS ordinations of individual domains (not shown) revealed that the influence of temperature was strongest for archaea (envfit r^2=0.60, p=0.001) followed by bacteria (envfit r^2=0.49, p=0.001), whilst sample type (diffuse fluid, background fluid, tubeworm grab) emerged as a secondary predictor for archaea (envfit r^2=0.45, p=0.001). None of the tested environmental variables showed significant influence on microeukaryote composition.

ANOSIM tests run on individual domains returned similar results when using the 'highT' and 'lowT' designations – significant compositional differences based on temperature for archaea (R=0.48, p=0.001) and bacteria (R=0.41, p=0.015). There were no significant distinctions based on substratum, location or sample type, using either individual microbial domains or the balanced microbial assemblage.

Broadly, compositional differences between highT and lowT grabs occurred in four major bacterial groups (Figure 3.2a). Epsilonbacteraeota and Bacteroidetes generally accounted for the majority of sequence reads in highT grabs and Gamma- and Alphaproteobacteria sequences were more numerous in lowT grabs. The highT grab from Middle Valley (MVw12) was the exception, being dominated by alphaproteobacterial sequences. For details of major contributors to these four bacterial groups, *see Supplementary Figure 5*. Tubeworm grabs were primarily distinguished from fluids by greater numbers of Bacteroidetes sequences.

Tubeworm-associated fauna included 58 distinct macro- and meiofaunal taxa (*Supplementary Tables 1 & 2*). Patterns of macro- and meiofaunal composition in NMDS ordinations were strongly influenced by temperature (*Supplementary Figure 6*). Overall, dirivultid copepods comprised about 80% of the meiofauna with species identity shifting

between highT and lowT; only lowT grabs returned nematodes and nemerteans (Figure 3.2b). About 13% of overall meiofaunal numbers were juvenile macrofauna, although they were rare in highT grabs. Macrofaunal species abundances were heavily skewed with 90% of total counts represented by three species: two gastropods (*Lepetodrilus fucensis,*

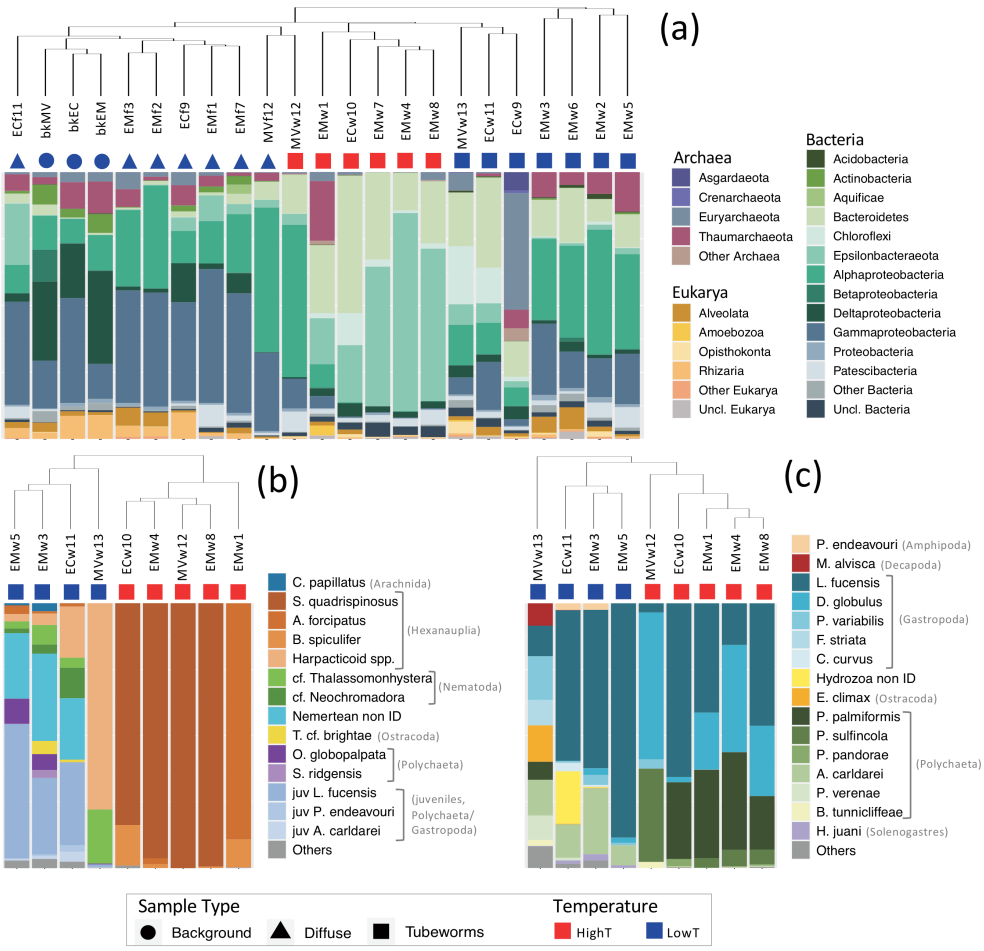


Figure 3.2 Taxonomic composition

Composition of major groups of (a) microorganisms and (b) meio- and (c) macrofaunal species. Dendrograms show hierarchical clustering of samples based on Aitchison distances calculated from centered log-ratio transformed species and OTU counts. "Other" includes taxa that were always <1% relative abundance in (a) and always <2% in (b) and (c). For major contributors to Gammaproteobacteria, Alphaproteobacteria, Bacteroidetes, and Epsilonbacteraeota, see *Supplementary Material Figure 3*.

Depressigyra globulus) and one polychaete (*Paralvinella palmiformis*). The limpet, *L. fucensis* was the only species to occur in every grab sample. Macrofauna relative abundances were highly variable at lowT (Figure 3.2c). The 21°C Middle Valley sample returned ten species not found among Endeavour grabs. Combined faunal rarefaction (*Supplementary Figure 7*) shows rapid levelling of species numbers at highT whereas two lowT grabs indicate larger samples would capture greater diversity. Further details of tubeworm-associated fauna are in *Supplementary Material*.

3.3.2 Congruent variation between size classes and microbial domains

NMDS ordinations of macrofauna, meiofauna, and microbes (three domains, analyzed separately and combined) (*Supplementary Figure 6*) showed strong Procrustes congruence between macro- and meiofauna (Table 3.2) followed by bacteria and microeukarya, then meiofauna with bacteria. The archaeal NMDS was not congruent with those of other microbial domains or size classes. Macro-meiofaunal relationships were consistently strong across samples, indicated by low Procrustes residual values (*Supplementary Table 3*), except one sample (MVw12). Although the combined microbial assemblage was not congruent with

Table 3.2 Procrustes analysis

Correlations between size classes and microbial domains determined by pairwise Procrustes analysis of NMDS plots. Values indicate correlations (sum of squares). **BOLD** values are significant ($p<0.05$).

	Macrofauna	Meiofauna	Bacteria	Archaea
Meiofauna	**0.86 (0.26)**			
Bacteria	0.59 (0.65)	**0.63 (0.61)**		
Archaea	0.60 (0.64)	0.38 (0.86)	0.32 (0.89)	
Microeukarya	0.31 (0.91)	0.47 (0.78)	**0.81 (0.34)**	0.22 (0.95)
Balanced microbes	0.42 (0.83)	0.52 (0.73)	-	-

either macro- or meiofauna (Table 3.2), there was more congruence in the higher temperature samples (Supplementary Table 3). Similarly, Archaea showed some congruence with meiofauna in highT samples.

3.3.3 Building communities

Potentially interacting taxa in highT and lowT *R. piscesae* habitats were identified using the combined results of enrichment and covariance (proportionality). In light of the strong temperature-driven distinction in faunal composition and congruence between macro- and meiofauna, we used highT- and lowT-enriched fauna as a starting point around which we built 'core communities' of covarying taxa. Enriched meiofaunal species included distinct highT and lowT copepod associations and a nematode in only LowT (*Supplementary Table 4*). Among macrofauna, highT-enriched species included three alvinellid polychaetes, a polynoid polychaete and a snail, while at lowT ampharetid polychaetes and a solenogastre were relatively abundant.

The 333 taxa included in proportionality analyses broadly separated into two clusters in the network diagram (Figure 3.3), with faunal placement consistent with their enrichment in either highT or lowT grabs (see *Supplementary Table 4*). Taxa with no absolute ρ values $\geq$ 0.75 or with strictly/mostly negative ρ values were not included in the network diagram. Plotting only macro- and meiofauna (not shown) further partitioned highT-enriched species into clusters highT 1 and highT 2 (Figure 3.3). Cluster 1 fauna had distinctly different relative abundances in highT and lowT samples while those in cluster 2 were more consistent over wider temperature ranges (*Supplementary Figure 8*). LowT macro- and meiofauna had fewer overall network connections (9.8 ± 5.0) but more connections among faunal species ($5.7 \pm$

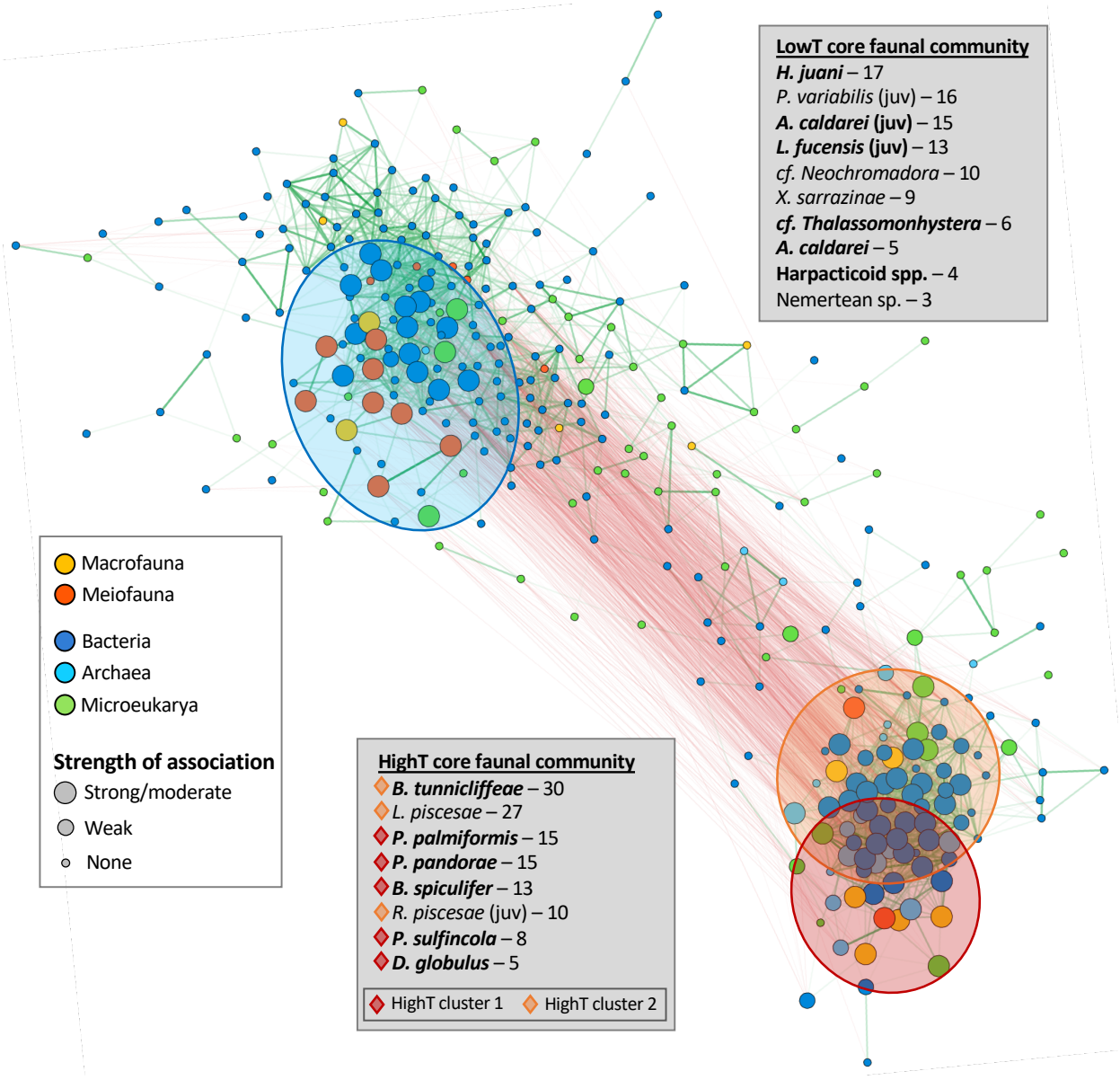


Figure 3.3 Co-occurrence network

Network of invertebrates and microbes from nine tubeworm grab samples showing proportionality (ρ) values of > 0.75 (green lines) or < -0.75 (red lines). Enriched faunal species (listed in **BOLD**) and taxa with strong to moderate associations to them comprise 'core communities' and are indicated by large nodes. Taxa with weak associations (minimal covariance or enrichment only) are indicated by intermediate-sized nodes. HighT taxa formed two clusters (red and orange ellipses). Numbers after listed taxa indicate the total number of positive connections within the network. Details of large and intermediate node taxonomies can be found in Table 3.3 and *Supplementary Table 5*.

2.2) relative to highT with 15.4 ± 8.8 overall connections and 1.5 ± 0.9 among-fauna connections. Enriched faunal species and co-varying taxa (proportionality $\rho \geq 0.75$) with strong to moderate association with highT or lowT clusters (Table 3.3) are indicated by large nodes in the network diagram (Figure 3.3) and considered members of core communities. Taxa with weaker associations (enrichment only or covariance with a single faunal species) with highT or lowT samples (*Supplementary Table 5*) are indicated as intermediate-sized nodes in Figure 3.3 and considered as potentially transient community members.

Assignment of microbial taxa (OTUs binned by identical taxonomic assignments) to core communities was primarily based on covariance with highT- or lowT-enriched fauna. Enrichment of individual OTUs within those taxa in the same temperature regime—highT or lowT—provided additional evidence and strengthened the assignment. Numerous microbial OTUs were notably enriched in some sample types (Figure 3.4, *Supplementary Table 4*). Due to frequent low relative abundances of archaeal and microeukaryal rRNA genes, many OTUs identified as enriched in individual domain ALDEx2 tests (see *Supplementary Figure 9*) did not remain so in the combined microbial assemblage results. HighT-enriched microbes included various Epsilonbacteraeota and microeukaryotes belonging to the Excavata and Fungi. Actinobacteria, Alpha- and Gammaproteobacteria, and Amoebozoa were enriched only in lowT grabs. Microbial groups with mixed highT and lowT enrichment included Bacteroidetes, Deltaproteobacteria, and the superphylum Patescibacteria.

The microbial taxa for which there was the strongest enrichment and covariance evidence for inclusion in the highT core community were the genera *Nitratifractor* and *Arcobacter* (Epsilonbacteraeota) and *Carboxylicivirga* (Bacteroidetes), followed by Bacteroidetes belonging to the Sphingobacteriales and BD2-2 and VC2.1 Bac22 clades (Table

Table 3.3 Microbial covariance and enrichment

Microbial taxa with strong (covariance with fauna and enrichment) and moderate (covariance with multiple faunal species) associations with highT or lowT grab samples.

		Covariance (proportionality)				Relative enrichment †	
Supergroup/Phylum/ Class	Microbial taxa ‡	Cytoscape cluster *	Total no. positive ρ values	No. positive ρ with macro/meio	Average ρ with macro/meio	vs. LowT grabs	vs. HighT diffuse
Strong associations with highT grabs							
Bacteroidetes	B_Carboxylicivirga	HighT 1/2	39	5	0.81	**X**	**X**
Epsilonbacteraeota	B_Nitratifractor	HighT 1	34	3	0.81	**X**	**X**
Epsilonbacteraeota	B_Arcobacter	HighT 1	28	2	0.83	**X**	**X**
Bacteroidetes	B_cl_Bacteroidetes VC2.1 Bac22	HighT 2	43	1	0.77	**X**	**X**
Bacteroidetes	B_or_Bacteroidales BD2-2	HighT 2	34	1	0.77	**X**	**X**
Bacteroidetes	B_or_Sphingobacteriales	HighT 2	26	1	0.75	**X**	**X**
Asgardaeota	A_sg_Asgardaeota	HighT 2	13	1	0.76	**X**	
Bacteroidetes	B_fa_Lentimicrobiaceae	HighT 2	45	2	0.78	**X**	
Bacteroidetes	B_Ichthyobacterium	HighT 1	17	2	0.80	**X**	
Bacteroidetes	B_Maritimimonas	HighT 2	27	1	0.76	**X**	
Deltaproteobacteria	B_fa_Desulfobulbaceae	HighT 2	40	2	0.77	**X**	
Epsilonbacteraeota	B_Hydrogenimonas	HighT 2	30	1	0.87	**X**	
Epsilonbacteraeota	B_or_Campylobacterales	HighT 1	44	1	0.77	**X**	
Excavata	E_or_Jakobida	HighT 1/2	24	2	0.82		**X**
Opisthokonta	E_or_Pezizomycotina	HighT 2	9	1	0.78		**X**
Moderate associations with highT grabs							
	A_do_Archaea	HighT 1/2	33	3	0.81		
Euryarchaeota	A_fa_Aciduliprofundaceae	HighT 1	25	3	0.81		
Euryarchaeota	A_Methanocaldococcus	HighT 1	33	2	0.85		
Euryarchaeota	A_Thermococcus	HighT 1/2	31	4	0.81		
Aquificae	B_fa_Desulfurobacteriaceae	HighT 1	24	2	0.81		
Aquificae	B_Thermosulfidibacter	HighT 2	14	3	0.84		
Bacteroidetes	B_Draconibacterium	HighT 1/2	40	5	0.80		
Bacteroidetes	B_fa_Marinilabiliaceae	HighT 1/2	48	5	0.81		
Bacteroidetes	B_or_Bacteroidales	HighT 2	35	2	0.83		
Calditrichaeota	B_Calorithrix	HighT 2	30	3	0.86		
Deltaproteobacteria	B_Desulfuromonas	HighT 1	30	3	0.85		
Deltaproteobacteria	B_Dissulfurimicrobium	HighT 2	49	2	0.87		
Deltaproteobacteria	B_fa_Desulfuromonadaceae	HighT 1/2	36	4	0.82		
Deltaproteobacteria	B_or_Desulfobacterales	HighT 2	36	3	0.82		
Epsilonbacteraeota	B_fa_Nitratiruptoraceae	HighT 1/2	43	5	0.82		
Epsilonbacteraeota	B_Nitratiruptor	HighT 1/2	41	3	0.82		
Thermodesulfobacteria	B_fa_Thermodesulfobacteriaceae	HighT 2	42	2	0.86		
Amoebozoa	E_Vermistella vermiformis	HighT 1	5	2	0.78		
Opisthokonta	E_fa_Dothideomycetes	HighT 2	6	2	0.77		
Opisthokonta	E_fa_Eurotiomycetes	HighT 2	11	2	0.86		
						vs. HighT grabs	vs. LowT diffuse
Strong associations with lowT grabs							
Actinobacteria	B_fa_Sva0996 marine group	LowT	23	4	0.79	**X**	**X**
Gammaproteobacteria	B_Halioglobus	LowT	30	3	0.79	**X**	**X**
Alphaproteobacteria	B_fa_Devosiaceae	LowT	16	1	0.78	**X**	**X**
Bacteroidetes	B_Euzebyella	LowT	25	1	0.79	**X**	**X**
Cyanobacteria	B_cl_Sericytochromatia	LowT	23	1	0.76	**X**	**X**
Gammaproteobacteria	B_Dasania	LowT	13	1	0.79		**X**
Moderate associations with lowT grabs							
Actinobacteria	B_cl_Acidimicrobiia	LowT	14	2	0.83		
Actinobacteria	B_Iamia	LowT	40	3	0.81		
Cand. Peregrinibacteria	B_cl_Candidatus Peribacteria	LowT	34	3	0.79		
Deltaproteobacteria	B_fa_Oligoflexaceae	LowT	27	3	0.78		
Deltaproteobacteria	B_Nannocystis	LowT	30	3	0.80		
Gammaproteobacteria	B_Marimicrobium	LowT	29	3	0.79		
Gammaproteobacteria	B_or_Cellvibrionales	LowT	40	3	0.77		
Cercozoa	E_Thalassomyxa lineage	LowT	23	2	0.81		
Ciliophora	E_Holosticha	LowT	28	6	0.83		
Ciliophora	E_or_Hypotrichia	LowT	5	3	0.80		

‡ Multiple OTUs binned by identical taxonomic assignment. Taxa are preceded by single letter code indicating microbial domain (A=Archaea, B=Bacteria, E=Microeukarya) and two letter code indicating taxonomic depth of classification, if other than genus (sg=supergroup, do=domain, ph=phylum, cl=class, or=order, fa=family).

* Cluster association in network figure 3.

† Relative enrichment determined by ALDEx2 pairwise tests.

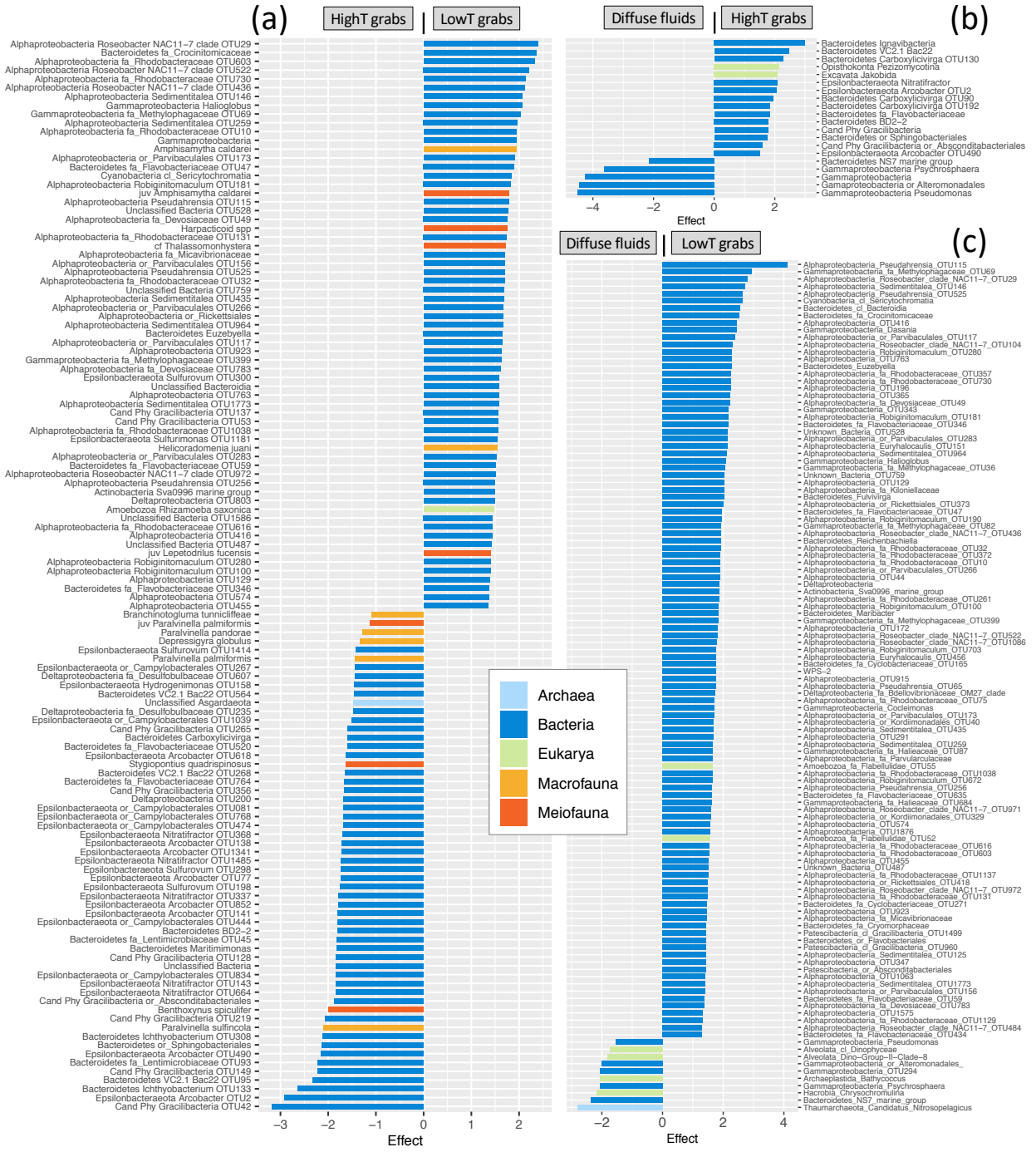


Figure 3.4 Enriched taxa

Bacteria, archaea, microeukarya, macrofauna and meiofauna responsible for significant differences between (a) high temperature (>25°C) and low temperature (<25°C) grab samples, (b) high temperature grab samples and associated diffuse fluids, and (c) low temperature grab samples and associated diffuse fluids. Effect size differences indicate relative enrichment in one sample type over the other. Only effect size differences ≥1 are shown. OTUs enriched in diffuse fluids include only those that were also enriched relative to background fluids.

3.3). Those covarying with highT fauna but with more limited enrichment included *Hydrogenimonas* and unclassified Campylobacterales (Epsilonbacteraeota), *Maritimimonas*, *Ichthyobacterium*, and Lentimicrobiaceae (Bacteroidetes), Desulfobulbaceae (Deltaproteobacteria), unclassified Asgard group Archaea, jakobid nanoflagellates, and fungi in the order Pezizomycotina. Moderate association with highT samples (covariance only) occurred for various members of the Archaea, Aquificae, Bacteroidetes, Calditrichaeota, Deltaproteobacteria, Epsilonbacteraeota, Thermodesulfobacteria, Amoebozoa and Fungi (Table 3.3).

Enrichment and covariance results most strongly supported the placement of Sva0996 marine group Actinobacteria and *Halioglobus* in the lowT core community, followed by *Euzebyella* (Bacteroidetes), the family Devosiaceae (Alphaproteobacteria), a basal cyanobacteria, and, with limited enrichment, *Dasania* (Gammaproteobacteria). Microbial LowT community members with moderate associations included members of the Actinobacteria, Delta- and Gammaproteobacteria, the candidate phylum Peregrinibacteria, ciliates and cercozoa (Table 3.3).

3.3.4 Fluid and tubeworm-hosted microbial assemblages

Comparisons of the seven diffuse fluid and associated tubeworm grab samples revealed elevated microbial OTU richness in the tubeworm grabs consistently for bacteria and occasionally for microeukaryotes. Archaeal richness in the grabs was rarely elevated relative to fluids. The greatest increases in the relative proportion of OTUs per sample from fluids to tubeworm grabs occurred in the taxa shown in Figure 3.5a. Collectively, these ten groups increased from 26(±5)% (fluid) to 72(±6)% (tubeworm grab) of the total OTUs per sample.

Although these increases were greatest for the Bacteroidetes and Alphaproteobacteria, much smaller percentage increases for the apicomplexans and lobose amoeba amounted to nearly 50- and 90-fold changes, respectively, in the numbers of OTUs per sample (Figure 3.5b). For details of the major contributors to richness increases, see *Supplementary Figure 10*.

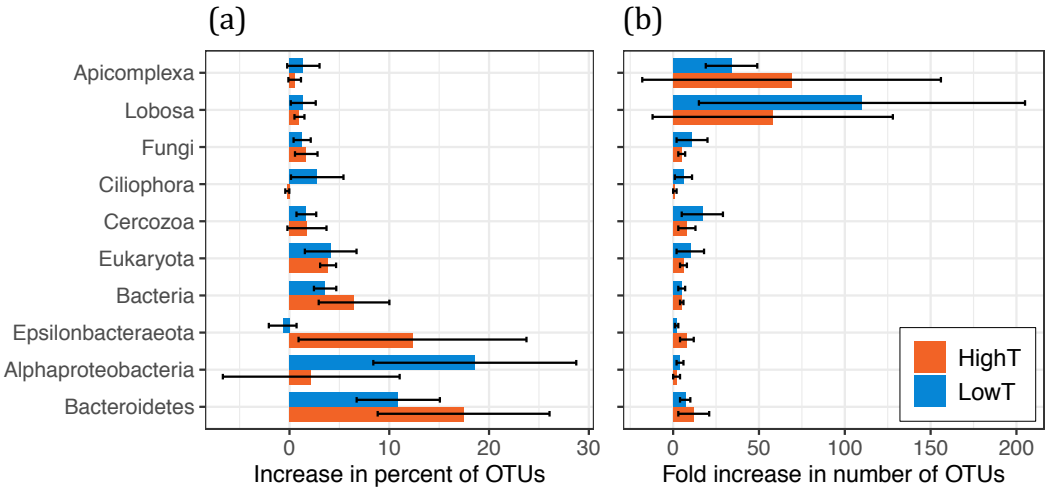


Figure 3.5 Increased diversity of major microbial groups

Increased diversity of Bacteria and Eukarya in highT and lowT tubeworm grabs relative to diffuse fluids. Increases shown as (a) the average change in the percentage of total OTUs per sample, and (b) the average fold change in the numbers of OTUs per sample. Error bars indicate one standard deviation.

Microbial taxa that were absent from diffuse and background fluids but present in at least three of these seven associated tubeworm grabs included various Amoebozoa (10 OTUs), the candidate bacterial phyla Eremiobacteraeota (17 OTUs) and Hydrogenedentes (1 OTU), Hilomonadea (Apusozoa; 6 OTUs), and the class Breviatea (68 OTUs). Many were low abundance OTUs with limited distributions, but some within *Ca.* Eremiobacteraeota and Breviatea were widely distributed across our grab samples. *Ca.* Eremiobacteraeota OTUs

occurred across lowT and one highT samples (EMw1), whereas Breviatea OTUs only occurred in grab samples with basal temperatures above 15°C.

3.3.5 Data availability

All 16S/18S rRNA sequence data have been deposited in the NCBI Sequence Read Archive under project number PRJNA665742. Many faunal identifications were complemented with COI and/or 18S sequences (GenBank accession numbers MZ197534-MZ197773). Data and R code are available at http://github.com/smurdock-UVic/Emergent-communities.

3.4 Discussion

Our coordinated analysis of a broad size spectrum of organisms associated with the foundation species *Ridgeia piscesae* has defined a robust temperature-driven distinction of communities. While faunal compositional differences are known among *R. piscesae* assemblages (Sarrazin and Juniper, 1999; Tsurumi and Tunnicliffe, 2003), our study greatly extends the characterization. Differences between assemblages with basal temperatures above and below ~25°C were evident in all three size classes comprising characteristic highT and lowT fauna and microbes. The thermal regime is likely one of several covarying controlling factors as the chemical milieu of the habitat also affects resident species. For example, temperature is a proxy for sulfide concentration (Butterfield et al., 1990), which is beneficial to sulfide-oxidizing primary producers, including faunal symbionts, but may inhibit species sensitive to its toxic effects. Similarly, dissolved oxygen concentrations at vents, inversely related to temperature (Johnson et al., 1986), are another driver of species distribution. Thus, we recognize that the highT/lowT designation may reflect geochemical

differences not addressed in this study.

Detailed structural evaluation of these faunal assemblages was a necessary first step towards understanding interactions among species and eventual ecosystem characterization. We discuss assemblage structure, including consideration of the fluid-associated microbial context in which they reside, and then make functional inferences based on identified core communities that can be explored in more detail in the future using metagenomic and other approaches.

3.4.1 Structural characterization

Taxonomic definition of the core communities associated with two primary habitat states—highT and lowT—involved distinguishing microbial residents of assemblages from those both in diffuse fluids venting through them and in background fluids. Elevated microbial OTU richness within tubeworm matrices relative to hydrothermal fluids discharging through them (Figure 3.5) indicates a facilitative effect of faunal assemblages on microbial diversity. Faunal assemblages contribute to habitat stability, some remaining nearly unchanged for at least a decade (Du Preez and Fisher, 2018). At Endeavour, one estimated age of the tubeworms supporting a low temperature assemblage was 30 years (Urcuyo et al., 2007). We hypothesize that these relatively stable tubeworm habitats aggregate microbial diversity over years to decades from an ever-changing supply of microbial taxa in discharging hydrothermal fluids (Huber et al., 2002, 2003; Perner et al., 2009). This aggregating effect broadens the importance of tubeworm habitats to the biodiversity of hydrothermal ecosystems from prior work on faunal diversity (Govenar, 2010).

Further, we identify tubeworm assemblages as hotspots of microbial novelty. Relative to fluids, tubeworm assemblages had increased richness of OTUs with shallow classification (domain, phylum, subphylum, class; *Supplementary Figure 10*). Large numbers of OTUs classifiable only as Bacteria or Eukarya occurred exclusively in tubeworm-hosted assemblages. Richness enhancements in unclassified Alphaproteobacteria, Bacteroidia, Pezizomycotina, Apicomplexa, Cercozoa and Spirotrichea, highlight the degree to which such hydrothermal vent habitats have been under-sampled. These highly novel taxa represent unknown ecosystem contributions, interspecies interactions, and potential benefits beyond the hydrothermal ecosystem through undiscovered genetic resource potential (Orcutt et al., 2020). This increased microbial richness in *R. piscesae*-hosted assemblages, with the substantial included novelty, suggests that previous assessments of hydrothermal vent microbiomes (Dick, 2019) may have missed the largest reservoir of diversity by not including contributions from non-symbiotic microbes in faunal assemblages.

Although we identify two primary habitat states, there is evidence of sub-components within the highT core community, suggesting even more specialized structuring and interaction of microbes and metazoans, under severe hydrothermal conditions. We propose that such structuring reflects core community members inhabiting different environmental niches within highT tubeworm grabs. Two highT sub-clusters (Figure 3.3) contrasted in faunal relative abundance distributions across the sampled range, with highT-specialist fauna in cluster 1 and broader-range fauna in cluster 2 (*Supplementary Figure 8*). We also noted non-overlapping covariance patterns between extreme-tolerant fauna and microbes. The copepod *Benthoxynus spiculifer* has a high oxygen affinity hemoglobin adaptation (Gollner et al., 2010a) to severely hypoxic conditions (Kalanetra and Nelson, 2010) and

covaries with obligate and facultatively anaerobic (hyper)thermophiles (e.g., *Nitratiruptor*, *Thermococcus, Methanocaldococcus*) and acidophiles (e.g., Aciduliprofundaceae, *Nitratifractor, Hydrogenimonas*), bespeaking extreme (high temperature, low pH, oxygen-starved) habitat conditions. In contrast, the polychaete *Paralvinella sulfincola*, which is the most thermally tolerant aquatic metazoan on record (Girguis and Lee, 2006), covaried with more mesophilic and facultatively anaerobic microbes indicating potential interactions in less extreme portions of its tolerance range.

Microbial taxa with limited enrichment—in highT or lowT grabs relative either to each other or to diffuse fluids—may indicate either a subsurface origin or general preference for any type of tubeworm habitat. Putative subsurface microbes—those enriched in highT versus lowT tubeworm grabs but also relatively abundant in diffuse fluids—included unclassified Asgard Archaea, members of the Bacteroidetes and Epsilonbacteraeota and a family of Deltaproteobacteria. A similar enrichment pattern did not occur in lowT grabs, suggesting decreased subsurface influence on microbial composition. Microbes broadly enriched in grab samples relative to diffuse fluids included jakobids and fungal Pezizomycotina (microeukaryotes) and *Dasania* (Gammaproteobacteria). These microeukaryotes are potential vent endemics (Burgaud et al., 2009; Murdock and Juniper, 2019) for which we specify tubeworm assemblages as a likely habitat. *Dasania* and other Cellvibrionales are mesophilic, heterotrophic, obligate aerobes likely belonging to the heterotrophic 'belt' surrounding diffuse vents (Meier et al., 2016). Mixed affinity for highT and lowT assemblages in taxa belonging to Bacteroidetes, Deltaproteobacteria, and Patescibacteria may indicate generally important roles played by species occupying different environmental or metabolic niches.

3.4.2 Inferred functional interactions among core community taxa

Refining the full taxonomic composition down to core communities offers opportunities for inferring functional relationships among enriched and covarying species. Increased congruence among size classes and decreased diversity of meiofauna, bacteria, and microeukaryotes in highT grabs suggest specialization of a few taxa, either by environmental selection or reliance on interspecies interactions. We identify eight faunal species (see Figure 3.3) and microbial taxa primarily belonging to the Epsilonbacteraeota and Bacteroidetes as the core community favoring highT habitats. Some taxa within this core community are known to form functional associations in other habitats or laboratory settings, providing legitimacy to our method of identifying potentially interacting taxa. For example, tight coupling of carbon and nitrogen cycling may occur between Epsilonbacteraeota and Bacteroidetes, as in hydrothermal vent biofilms (Stokke et al., 2015). Bacteroidetes may also benefit from associations with core highT faunal species *D. globulus* and *P. palmiformis*, both mucus-secreting species potentially providing fresh organic material for these heterotrophic bacteria. Amoeboflagellates in the class Breviatea were among the taxa absent from fluids but associated with tubeworms; they are deeply-branching anaerobes/microaerophiles (Adl et al., 2019) that possess mitochondria-like hydrogenosomes (Brown et al., 2013) and form mutualisms involving hydrogen transfer with *Arcobacter* and other Epsilonbacteraeota (Hamann et al., 2016). These and other potential partnerships may be explored further using metagenomic approaches focused on specific core community taxa (e.g., *Carboxylicivirga*, *Arcobacter*, *Nitratifractor*, etc.), but the prevalence of potential mutualisms among highT core taxa suggests interspecies connections may be crucial to survival in more extreme hydrothermal habitats.

Functional links among lowT core taxa can be inferred but are fewer and less apparent. The core lowT faunal community was more weighted towards meiofauna, possibly reflecting lower thermal tolerance (Gollner et al., 2010b) and/or suggesting potentially relevant functional connections of meiofauna to lowT assemblages (Sarrazin et al., 2015). The inclusion of juveniles of three species in this core community suggests that lowT assemblages may serve as nursery areas for macrofauna. Juvenile macrofauna are included with meiofauna as they undoubtedly have different functional interactions within the community than their adult forms. For example, juvenile *L. fucensis* were very abundant in low T while the adults occur over the full range of habitats where they develop a symbiosis with a γ-proteobacterium, the extent of which depends on fluid vigour (Bates et al., 2011). Recent interest in congruence among macrofauna, meiofauna, and microbes highlights the undervalued meiofaunal contributions to ecosystem processes that link the micro and macro worlds (Cronin-O'Reilly et al., 2018; Schratzberger and Ingels, 2018). Examples of inferred links involving meiofauna in the lowT core community include nematodes fueling strictly aerobic alpha- and gammaproteobacterial heterotrophs through secretion of organic-rich mucus (Reimann and Schrage, 1978), and hypotrich ciliates, which covaried with several meiofaunal species, possibly acting as a preferred prey item. These and other potential meiofaunal interactions have been largely overlooked in hydrothermal habitats and may require creative experimental approaches to confirm (Léveillé et al., 2005; Gollner et al., 2010b).

Ecosystem characterization aims to identify the key players and define their contributions to maintaining ecosystem function. Characterization of microbial, meio- and macrofaunal diversity across a range of habitats of the foundation species *Ridgeia piscesae*

has revealed core highT and lowT communities with inferred interactions between species and highlighted the importance of faunal assemblages as hotspots of microbial richness in hydrothermal ecosystems. We propose definition of core communities as an important first step in moving from structural descriptions of biological assemblages to informed consideration of relevant functional interactions within complex and diverse ecosystems. Identification of core communities has also suggested potential functional contributions of small organisms, including meiofauna, which have been undervalued in deep-ocean monitoring and conservation strategies (Ingels et al., 2020). We foresee additional applicability of the core communities in vulnerability assessments for hydrothermal vent ecosystems in relation to future stressors such as seabed mining (Thompson et al., 2018). This work exposes the rich complexity and microbial underpinning of hydrothermal vent ecosystems to managers who, while familiar with phytoplankton-zooplankton-fish connections in the water column, may have a less established framework for extending inclusive, ecosystem-based approaches to vents and other deep-sea ecosystems.

3.5 Article Supplementary Material

Supplementary Methods

Sample collection

Locations were visually assessed to determine the likelihood of retrieving a tubeworm bush with associated fauna intact. Hydrothermal fluids venting through assemblages were collected prior to assemblage sampling using the ROPOS suction sampler with the intake positioned as close as possible to the faunal assemblage without disturbing it. Afterwards, the manipulator claw gently encompassed part, or all, of the tubeworm bush (at the base) and transferred tubes and associated organisms to a collection box which was then sealed. The entire sample was usually retrieved in a single grab. We reviewed video imagery of the operation to determine losses of mobile fauna, and subsequently added small numbers of scaleworms and a few squat lobsters to final counts.

Shipboard, approximately half of each tubeworm grab sample was placed in a bucket of cold artificial seawater (ASW) and held at 4˚C for processing of the microbial component. The microbial fraction was a mix of loosely associated cells and firmly attached biofilms that had to be detached prior to collection on filters. The collection boxes on the ROV were then drained to ensure all remaining organisms were retained, and samples were preserved in either 75% ethanol or 7% buffered formalin. See *Supplementary Figure 2* for visual representation of processing steps.

Macro- and meiofaunal characterization

To sort the material, we first assessed sample size and, in four cases, split the samples (using only ¼ to ½) to speed processing. In all these splits, species discovery rapidly reached

asymptote (Supplementary Figure 4), thus it is likely we missed few or none. Next, we removed large *Ridgeia pisceae* tubes and processed the remaining material into two size classes on 1mm and 64μm sieves into fresh 75% ethanol. Macrofauna was retained on the larger sieve, while the animals on the 64μm sieve constituted the meiofauna. Examination of material passing the 64μm sieve never revealed animals. Where the adult stages of species occurred on both sieves, all individuals were assigned to the fraction that retained the most numbers. Many juveniles of macrofauna passed the 1mm sieve and are reported separately in the meiofaunal list. As meiofaunal individuals were abundant, we used a plankton splitter to generate smaller portions. For six of the nine samples, 0.25 portions were fixed for counting, but for the remaining samples, finer 0.05 to 0.10 splits were required because of high copepod abundances.

Animals were identified to distinct species for all macrofauna although complete names are not available for all. Among meiofauna, where copepods were highly abundant, subsamples of 200-250 individuals were identified to species. However, insufficient taxonomic information hampered differentiation of harpacticoid taxa, especially as juvenile forms were present. Nematodes were separated into three morphotypes representing at least three species. We calculated relative abundances of species/taxa within each size class for every sample.

To calculate density, tube surface area was determined from images of each tubeworm following the method of Tsurumi and Tunnicliffe (2003). For similarity and cluster analyses, juveniles of macrofauna remained in the meiofauna as we are interested in the functional relations among the five groups in this study. Rarefaction curves were produced in PAST v.3.23 (Hammer et al., 2001).

Microbial DNA extraction and sequencing

Primers used for paired-end sequencing on Illumina MiSeq included 63F (5'-CAGGCCTAACACATGCAAGTC-3') and 519R (5'-GWATTACCGCGGCKGCTG-3') for Bacteria, 956F (TYAATYGGANTCAACRCC) and 1410R (5'-CRGTGWGTRCAAGGRGCA-3') for Archaea, and 572F (5'-CYGCGGTAATTCCAGCTC-3') and 1009R (5'-AYGGTATCTRATCRTCTTYG-3') for Eukarya. Bacteria and Archaea were sequenced from genomic DNA but initial sequencing of Eukarya from genomic DNA returned mostly metazoan sequences. Therefore, sequence data for microeukaryotes reported in this study came from amplicons generated using the primers 18S-EUK581-F (5'-GTGCCAGCAGCCGCG-3') and 18S-EUK1134-R (5'-TTTAAGTTTCAGCCTTGCG-3') (Bower et al., 2004), which have been validated for selectively amplifying non-metazoan DNA (Del Campo et al., 2019). Microeukaryote 18S amplicons were produced on an iCycler (Bio-Rad) with the following conditions: initial denaturation at 94°C for 2 minutes, 27 cycles of 94°C for 30 seconds, 62°C for 45 seconds, and 72°C for 1 minute, and final extension at 72°C for 10 minutes. Each 20µl amplification reaction consisted of 1µl of DNA, 1X buffer (2.5mM $MgCl_2$; Promega), 0.2mmol each deoxyribonucleoside triphosphate (dNTP), 0.25µmol each primer, 1U of GoTaq polymerase (Promega), and DNAse-free water to final volume. Amplicons were cleaned using a QiaQuick PCR purification kit (Qiagen) prior to preparation for sequencing.

Quantitative PCR

Relative abundances of bacteria, archaea and microeukarya in each sample were determined by quantitative PCR (qPCR) of 16S and 18S rRNA genes using established and tested lab protocols, which utilize the primers 331F (5'-TCCTACGGGAGGCAGCAGT-3') and 797R (5'-

GGACTACCAGGGTATCTAATCCTGTT-3') (Nadkarni et al., 2002) for bacteria, 20F (5'-TTCCGGTTGATCCYGCCRG-3') and DW518R (5'-GNTTTACCGCGGCKGCTG-3') (Zaikova et al., 2010) for archaea, and 18S-EUK581-F and 18S-EUK1134-R (Bower et al., 2004) for microeukaryotes. To create qPCR standards, plasmid DNA was extracted from 4 bacterial, 5 archaeal, and 3 microeukaryote environmental and culture clones from a variety of hydrothermal vent locations (in-house collections from Mariana Arc, Kermadec Arc and Juan de Fuca Ridge) using a Miniprep plasmid extraction kit (Qiagen) and further prepared as described in (Zaikova et al., 2010) beginning with DNase treatment. Plasmids were quantified on a Qubit 3 (Invitrogen) using the Qubit dsDNA high-sensitivity assay (Invitrogen) and mixed in equal volumes to produce a standard for each domain, with concentrations on the order of 10^7 molecules/µl.

A ten-fold dilution series of each plasmid mixture, ranging from 10^7 to 10^1, was used to produce standard curves with R^2 values > 0.995. Each 10µl reaction consisted of 5µl SsoFast EvaGreen Supermix (Bio-Rad), 0.5µM each forward and reverse primers, 2µl DNase-free water, and 1µl of DNA template. Reactions were performed in triplicate on a CFX96 Real-Time PCR Detection System (Bio-Rad) and consisted of an initial denature for 2 min at 95°C followed by 45 cycles of 95°C for 5 sec and a 5 sec anneal/extend step and concluding with a melt curve from 65-95°C. Temperatures for the anneal/extend step were 61°C for bacteria/archaea and 62°C for eukarya. Each run included a standard curve and a no-template control. Samples were analyzed with DNA combined from the two size fractions. A dilution series of each DNA sample was analyzed to test for inhibitory reaction effects, and final quantitative reactions were run using a 1:10 dilution of the DNA. CFX Manager 2.0 software (Bio-Rad) was used to analyze the results.

Microbial dissimilarity linked to sampling method

To address whether microbial composition was influenced by different techniques used for harvesting cells from tubeworm grab samples (direct removal versus mild agitation in ASW), we performed Analysis of Similarities (ANOSIM) on two samples (EMw1 and EMw6) with replicates that were subjected to both removal methods. ANOSIM was also used to test for compositional differences between the two size fractions (0.2-20μm, 20-64μm) collected by serial filtration of microbial biomass.

Supplementary Results

Tests of sampling bias on microbial composition

Composition of the multi-domain balanced microbial assemblage was not significantly different (ANOSIM R=0.25, p=0.33) between the two harvesting methods, although when domains were tested separately using ALDEx2, a few bacterial and microeukaryal OTUs were differentially represented between the two methods (not shown). Further ANOSIM performed on all tubeworm-associated extracts found no OTUs that were consistently enriched using one method versus the other. However, removal method did have an effect on microeukaryote diversity, with significantly lower inverse Simpson values when using the direct removal method (Mann-Whitney test, p=0.004).

Serial filtration onto 20 and 0.2 μm filters was aimed at capturing eukaryotes from both micro (20-64μm) and pico/nano (0.2-20μm) size ranges, but our results indicated no significant difference (ANOSIM R=0.02, p=0.23) in composition between size fractions for eukaryotes or bacterial and archaea.

Faunal Character

The projected (assuming full sample sort) ratio of total meiofauna to macrofauna was 3.7, however, the variation among samples was very high (0.6 to 317). Overall, we identified 58 distinct taxa (*Supplementary Tables 1 & 2*), of which the harpacticoid copepods, and possibly the nemerteans, were lumped species. Rarefaction curves (*Supplementary Figure 7*) show rapid saturation in high temperature samples indicating that all species were likely recovered; the two lowest temperature samples with highest species numbers (30 to 35) have lower estimates of full species recovery.

Meiofaunal description

Meiofaunal density (not including juvenile macrofauna) on *Ridgeia* tubes ranged from 3.2 to 68.6 individuals/10cm^2 of tube surface in the Endeavour samples. Dirivultid copepods comprised about 80% of the meiofauna with *Stygiopontius quadrispinosus* dominating highT samples while harpacticoid copepods occurred mostly in lowT samples (Figure 3.2b). Only lowT samples returned nematodes (Chromadorea) and nemerteans (Palaeonemertea) at 6% abundance overall.

Macrofaunal description

Macrofaunal densities ranged from 0.2 to 1.8 individuals/10cm^2 of tube surface in the Endeavour samples. We retrieved 41 macrofaunal species in which abundances were heavily skewed with 90% of total counts represented by three species: two gastropods (*Lepetodrilus fucensis, Depressigyra globulus*) and one polychaete (*Paralvinella palmiformis*) (*Supplementary Table 1*). The limpet, *Lepetodrilus fucensis* was the only species to occur in every sample. Relative abundances were highly variable at lowT (Figure 3.2c).

Supplementary Figures and Tables

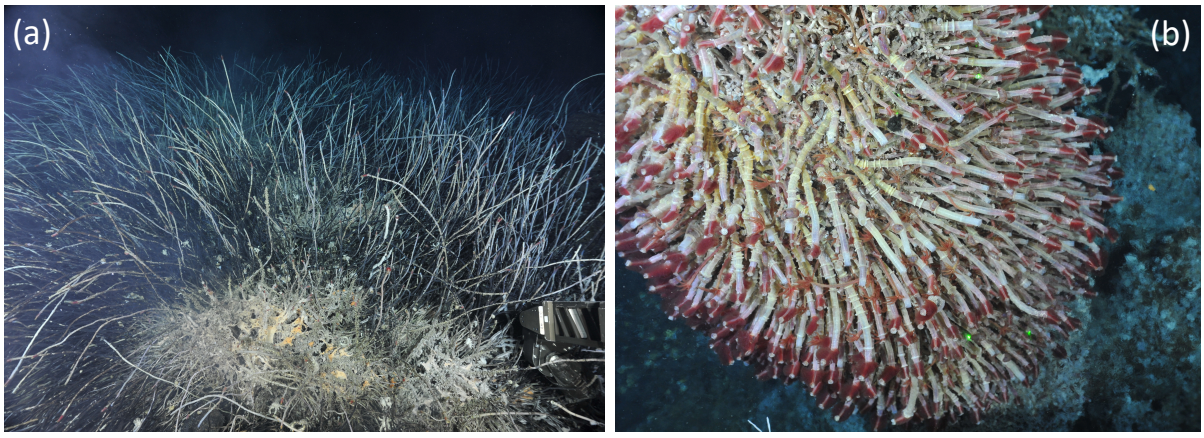


Figure 3.6 Supplementary Figure 1: Tubeworm assemblages

Microbial and faunal assemblages hosted by the Siboglinid tubeworm *Ridgeia piscesae* in (a) low temperature, basalt-hosted and (b) high temperature, sulfide-hosted habitats.

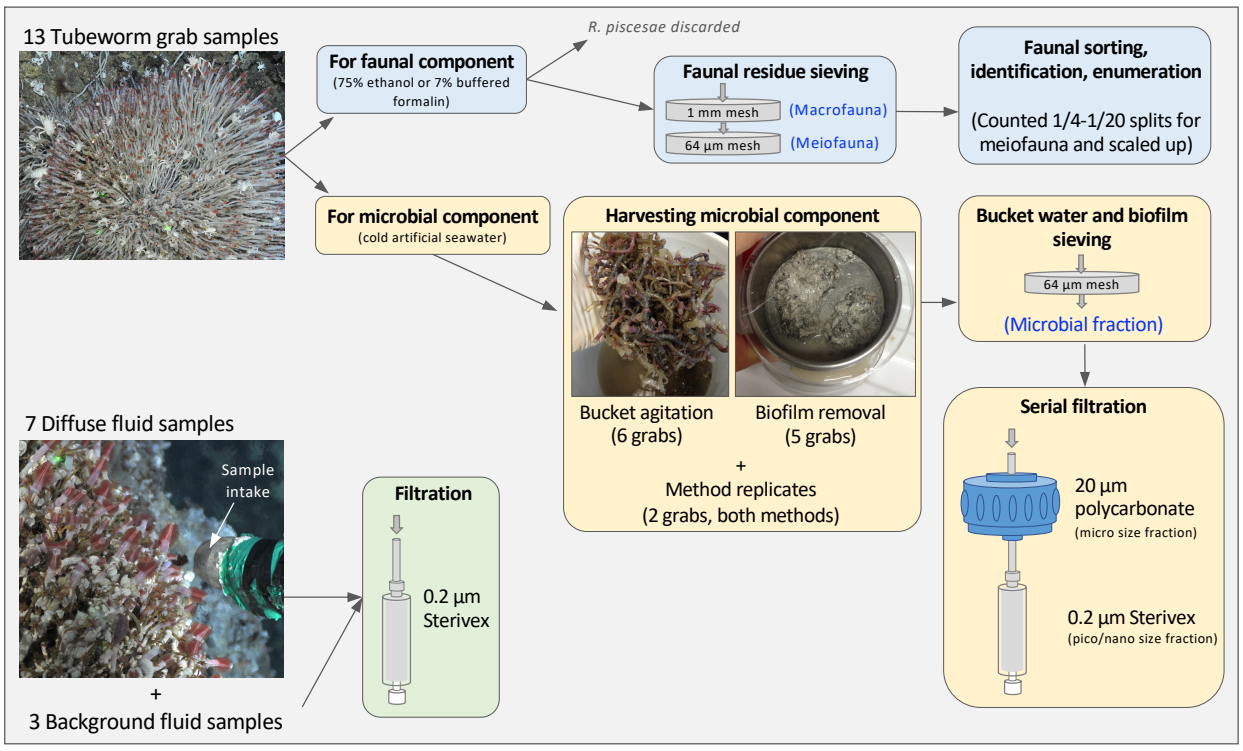


Figure 3.7 Supplementary Figure 2: Sample processing diagram

Sample processing steps from tubeworm grabs and diffuse and background fluids. Tubeworm grabs were divided for processing of fauna (blue boxes) and microbes (yellow boxes), and fluid samples went straight to filtration (green box).

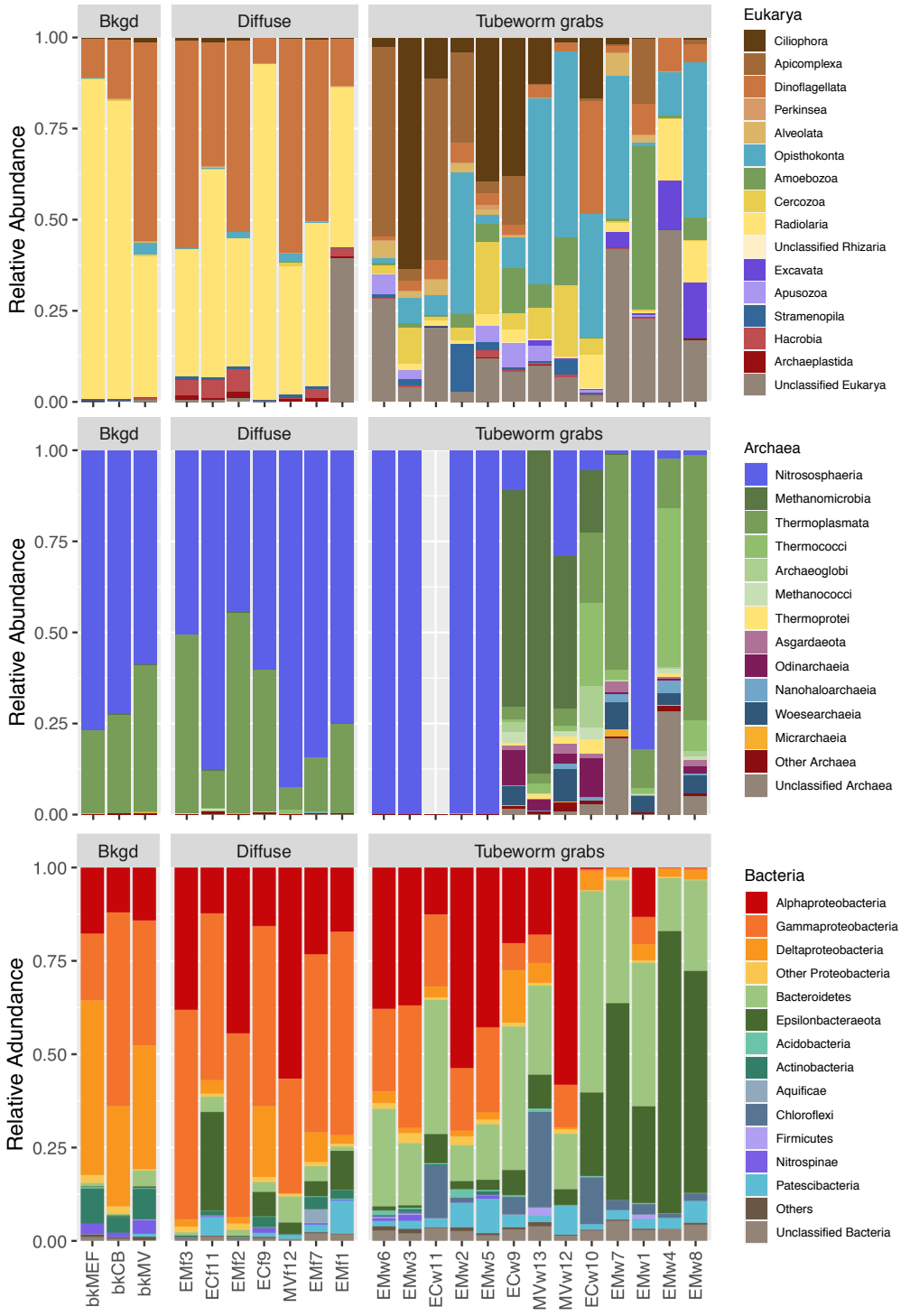


Figure 3.8 Supplementary Figure 3: Barplots of taxonomies by domain

Taxonomic identity and relative abundance of sequences for each domain. Sample ECw11 did not produce sufficient archaeal reads. Bkgd= background fluid samples; Diffuse= diffuse fluid samples; Tubeworm grabs= grab samples. Diffuse fluid and grab samples are arranged by increasing basal temperature from left to right.

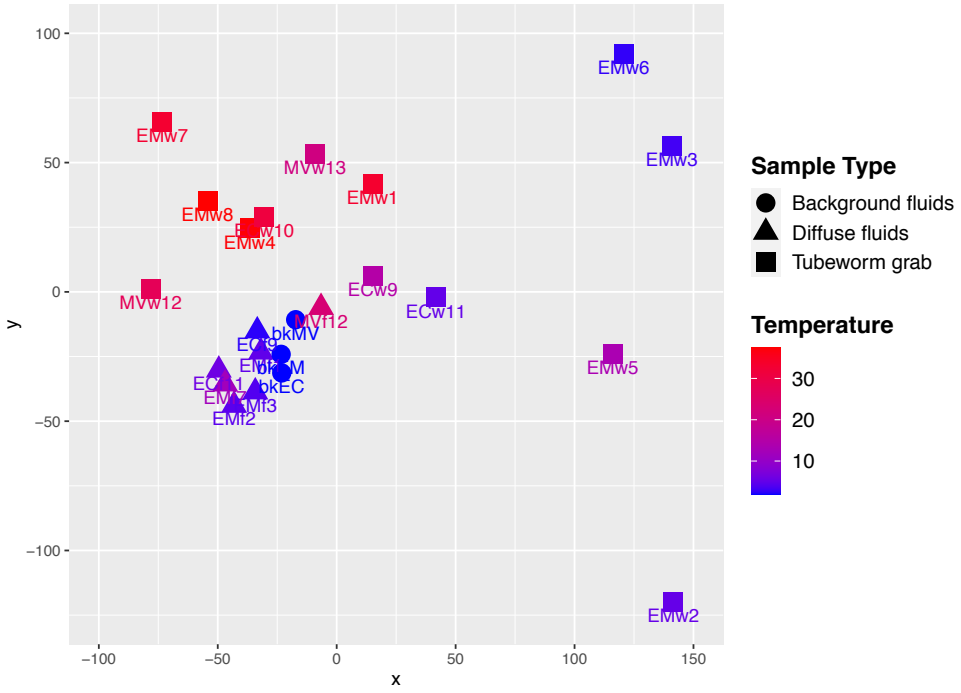


Figure 3.9 Supplementary Figure 4: NMDS of q-PCR balanced microbes

Nonmetric Multidimensional Scaling ordination of the q-PCR balanced microbial assemblage from 13 grab samples, 7 diffuse fluid and 3 background fluid samples. 2-D stress=0.13.

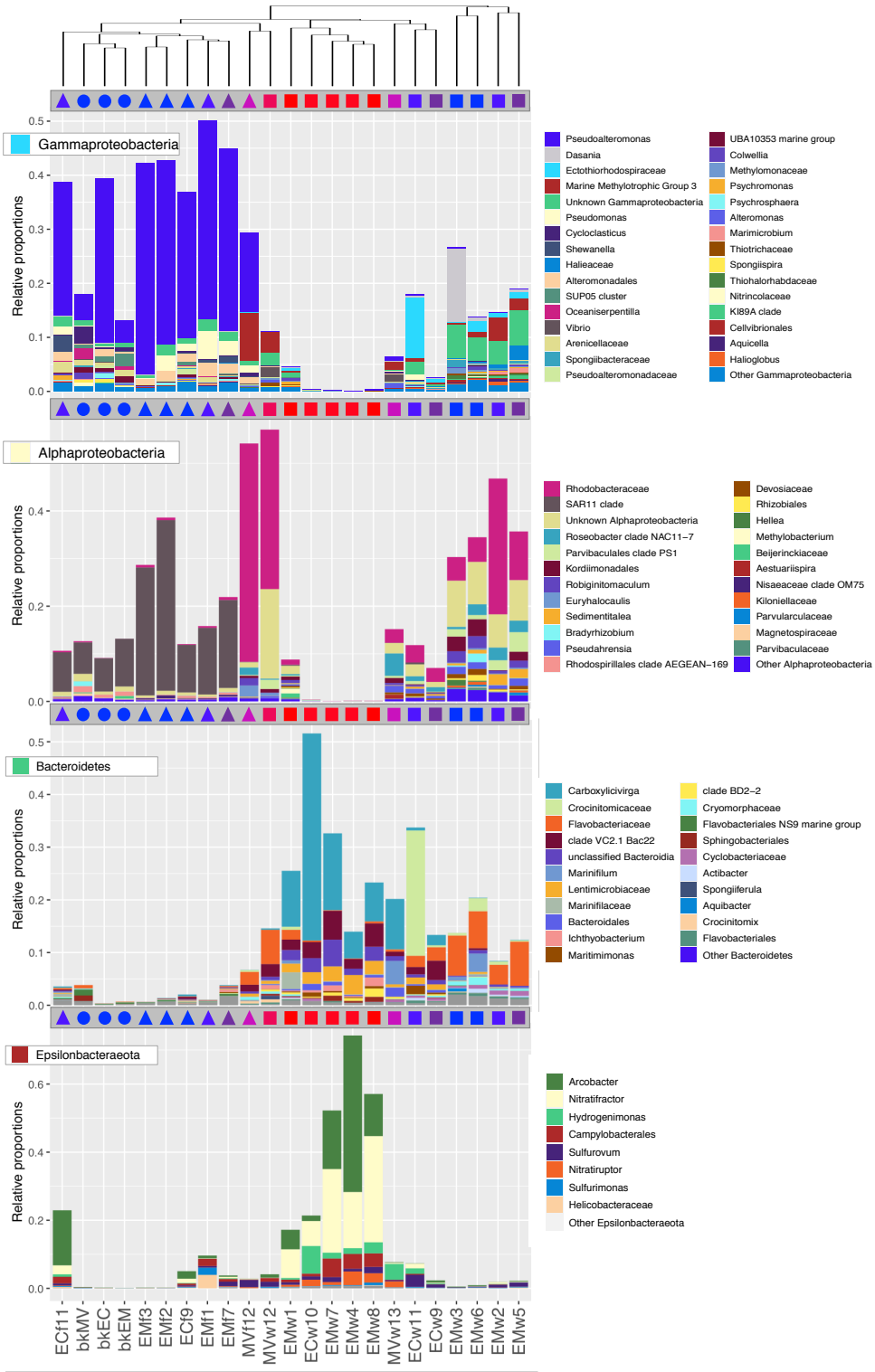


Figure 3.10 Supplementary Figure 5: Taxonomic composition of major groups

Four bacterial groups with the greatest contributions to diversity (from Figure 3.2) across the sample set. Sample order and dendrogram are identical to Figure 3.2a.

Table 3.4 Supplementary Table 1. Macrofaunal species abundances

Counts justified to full sample.

Group	Family	Name	EMw1	EMw3	EMw4	EMw5	EMw8	ECw10	ECw11	MVw12	MVw13
Amphipoda	Ischyroceridae	*Bonnieriella linearis*	0	0	0	0	0	0	1	0	0
Amphipoda	Pardaliscidae	*Pardalisca endeavouri*	0	63	0	0	0	0	52	0	0
Amphipoda	Sebidae	*Seba profunda*	0	2	0	0	0	0	6	0	0
Asteroidea	Xyloplactidae	*Xyloplax* n. sp.	0	46	0	0	0	0	0	0	0
Bivalvia	Benthomodiolinae	*Benthomodiolus erebus*	0	0	0	0	0	0	2	0	11
Bivalvia	Vesicomyidae	unk juveniles	0	0	0	0	0	0	0	0	2
Cnidaria	Actinaria	unk anemone	0	0	0	0	0	0	0	0	22
Decapoda	Munidopsidae	*Munidopsis alvisca*	0	0	0	0	0	0	2	0	103
Demospongiae	Cladorhizidae	Cladorhizidae	0	3	0	0	0	0	0	0	0
Foraminifera		Foraminifera	0	1	0	0	0	0	7	0	0
Gastropoda	Buccinidae	*Buccinum thermophilum*	0	2	0	0	0	0	0	0	0
Gastropoda	Lepetodrilidae	*Clypeosectus curvus*	0	18	0	2	0	0	63	0	0
Gastropoda	Lepetodrilidae	*Lepetodrilus fucensis*	5016	1534	716	10370	1020	17446	1169	6	140
Gastropoda	Peltospiridae	*Depressigyra globulus*	2612	65	1834	234	585	538	4	96	0
Gastropoda	Provannidae	*Provanna variabilis*	8	96	2	104	0	0	9	6	200
Gastropoda	Provannidae	*Provanna cf laevis*	0	0	0	0	0	0	0	0	15
Gastropoda	Pyropeltidae	*Pyropelta musaica*	0	0	0	0	0	0	0	0	3
Gastropoda	Skeneidae	*Fucaria striata*	0	1	0	0	0	0	5	0	117
Hydrozoa		Hydrozoa	0	8	0	0	0	0	407	0	0
Ostracoda	Myodocopida	*Euphilomedes climax*	0	4	0	2	0	0	4	0	167
Polychaeta	Allvinellidae	*Paralvinella dela*	0	0	0	0	0	0	0	0	2
Polychaeta	Allvinellidae	*Paralvinella palmiformis*	4064	0	1676	0	445	7711	0	0	83
Polychaeta	Allvinellidae	*Paralvinella pandorae*	20	0	4	0	10	758	0	0	0
Polychaeta	Allvinellidae	*Paralvinella sulfincola*	424	0	284	0	123	0	0	61	0
Polychaeta	Ampharetidae	*Amphisamytha carldarei*	0	640	4	884	9	0	260	0	164
Polychaeta	cf Archinomidae	sp non ID	0	0	0	0	0	0	0	0	4
Polychaeta	Hesionidae	sp non ID	0	0	0	0	0	0	0	0	9
Polychaeta	Maldanidae	*Nicomache venticola*	0	3	0	4	0	0	0	0	5
Polychaeta	Maldanidae	*Nicomache cf ardwissoni*	0	0	0	0	0	0	0	0	5
Polychaeta	Orbiniidae	*Orbiniella hobsonae*	0	5	0	2	0	0	4	0	0
Polychaeta	Orbiniidae	*Scoloplos* n. sp.	0	3	0	0	0	0	0	0	0
Polychaeta	Phyllodocidae	*Protomystides verenae*	0	0	0	0	0	0	21	0	112
Polychaeta	Polynoidae	*Branchinotogluma tunnicliffeae*	4	0	12	0	3	59	0	4	27
Polychaeta	Polynoidae	*Lepidonotopodium piscesae*	0	0	8	0	9	90	6	0	5
Polychaeta	Polynoidae	*Levensteiniella intermedia*	0	0	0	4	0	0	6	0	0
Polychaeta	Scolicida	sp non ID	0	0	0	0	0	0	0	0	8
Polychaeta	Spionidae	sp non ID	0	0	0	0	0	0	0	0	8
Pycnogonida	Ammotheidae	*Sericosura dissita*	0	1	0	0	0	0	0	0	0
Pycnogonida	Ammotheidae	*Sericosura verenae*	0	2	0	6	0	0	1	0	0
Sipuncula		Sipunculidea	0	5	0	0	0	0	1	0	0
Solenogastres	Simrothiellidae	*Helicoradomenia juani*	0	61	0	106	0	0	25	0	3
		Justified to full sample	12148	2563	4540	11718	2204	26602	2055	173	1215
		Original numbers assessed	3037	2563	2270	5859	2204	10641	2055	173	1215
		# spp present	7	21	9	11	8	6	21	5	23

Table 3.5 Supplementary Table 2: Meiofaunal species abundances

Counts justified to full sample.

Group	Family	Name	EMw1	EMw3	EMw4	EMw5	EMw8	ECw10	ECw11	MVw12	MVw13
Arachnida	Halacaridae	*Copidognathus papillatus*	4	336	0	150	0	0	4	0	0
Hexanauplia	Dirivultidae	*Aphotopontius forcipatus*	9184	96	1088	540	0	0	64	0	0
Hexanauplia	Dirivultidae	*Benthoxynus spiculifer*	1096	0	726	0	312	2408	8	0	0
Hexanauplia	Dirivultidae	*Stygiopontius quadrispinosus*	0	0	47530	84	54796	13186	0	54912	0
Hexanauplia	Dirivultidae	*Collocherides brychius*	0	200	0	0	0	0	0	0	0
Hexanauplia	Harpacticoida	spp	0	480	0	512	0	0	1036	0	10840
Malacostraca	Pardaliscidae	*Pardalisca endeavouri*	0	64	0	0	0	0	132	0	0
Malacostraca	Amphipoda	Amphipod sp (Seba?)	0	20	0	0	0	0	0	0	0
Ostracoda	Cytheruridae	*Xylocethere sarrazinae*	0	4	0	240	0	0	28	0	1
Ostracoda	Pontocyprididae	*Thomontocypris cf brightae*	0	549	0	4	0	0	52	0	0
Gastropoda	Lepetodrilidae	*Lepetodrilus fucensis*	20	3184	12	9456	20	0	1688	0	140
Gastropoda	Peltospiridae	*Depressigyra globulus*	0	0	0	4	0	0	16	0	0
Gastropoda	Provannidae	*Provanna variabilis*	0	96	0	252	0	0	72	0	0
Polychaeta	Alvinellidae	*Paralvinella palmiformis*	0	0	8	0	0	98	0	0	0
Polychaeta	Ampharetidae	*Amphisamytha carldarei*	0	124	0	140	0	0	204	0	30
Polychaeta	Dorvilleidae	*Ophryotrocha globopalpata*	0	662	0	1764	0	0	0	0	42
Polychaeta	Orbiniidae	*Scoloplos* sp	0	8	0	28	0	0	0	0	0
Polychaeta	Phyllodocidae	*Protomystides verenae*	0	0	0	0	0	0	12	0	20
Polychaeta	Siboglinidae	*Ridgeia piscesae*	0	0	4	0	0	59	0	0	10
Polychaeta	Syllidae	*Sphaerosyllis cf ridgensis*	0	325	0	12	0	0	3	0	0
Polychaeta	Polychaete unk	non ID cf hesionid	0	0	0	0	0	0	4	0	0
Nematoda	Monhysteridae	*cf Thalassomonhystera*	0	812	0	504	0	0	204	0	2840
Nematoda	unk	unknown sp	4	56	0	12	0	0	0	0	0
Nematoda	cf Chromadoridae	*cf Neochromadora*	0	376	0	316	0	0	620	0	0
Nemertea	Palaeonemertea	unknown sp	0	3632	0	4608	8	0	1252	0	0
		Justified to full sample	10308	11024	49368	18626	55136	15750	5399	54912	13923
		Original numbers assessed	2577	2756	12342	4657	13784	788	1350	3295	1392
		# spp present	5	18	6	17	4	4	17	1	8

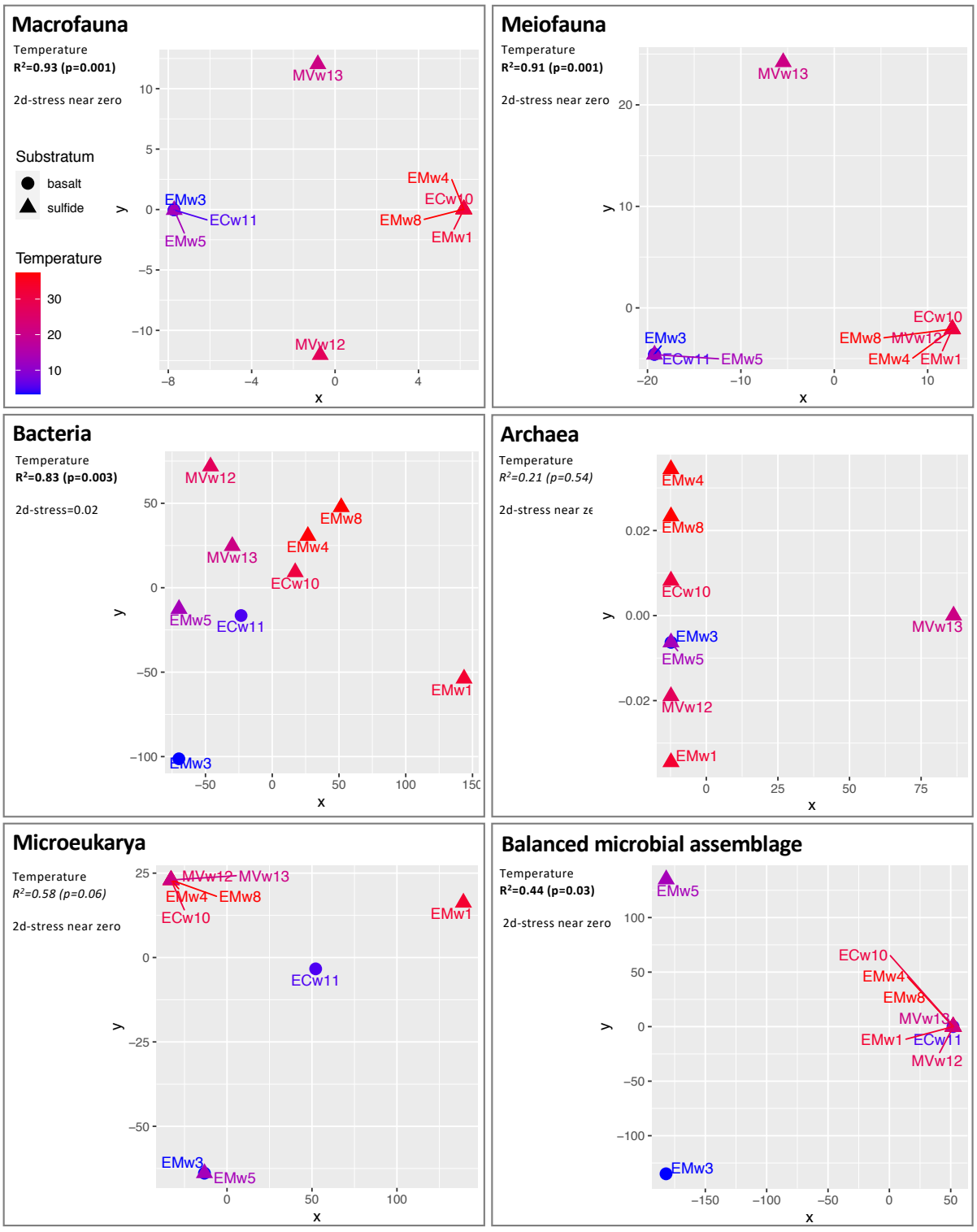


Figure 3.11 Supplementary Figure 6: NMDS used in Procrustes analysis

Ordinations were performed using Aitchison distances between samples with centered log ratio transformed abundances.

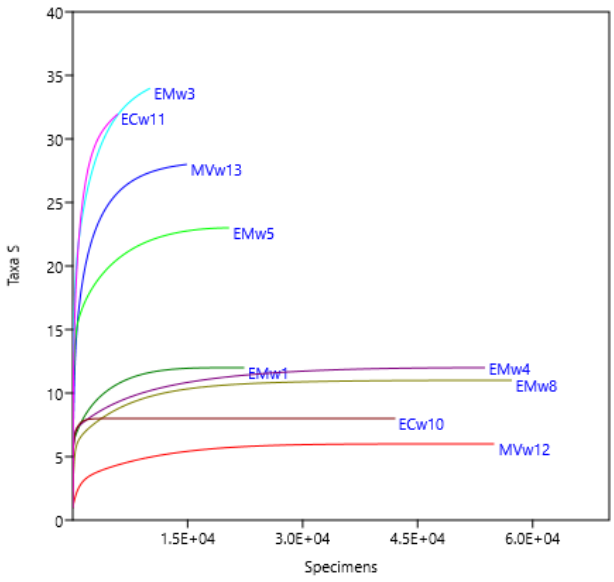


Figure 3.12 Supplementary Figure 7: Macro-/meiofaunal rarefaction curves

Rarefaction of individuals within samples; unique species (i.e., juveniles not included) only. Five highT samples reach saturation rapidly, two more lowT (sulfide) slowly and two lowT (basalt) not at all.

Table 3.6 Supplementary Table 3: Procrustes residuals

Individual sample residual values from pairwise Procrustes analyses. Pairs in **BOLD** showed a significant relationship ($p<0.05$). Green=low residual values, Red=high residual values (relative color scaling is within rows). ND=no data, Macro=macrofauna, Meio=meiofauna, Bact=bacteria, Arch=archaea, Euk=microeukarya, Micro=bacteria, archaea, and microeukarya proportionally combined based on qPCR.

Temperature (°C)	3.4	5.2	13.4	21.1	27.4	31	33.2	37.1	37.4
	EMw3	ECw11	EMw5	MVw13	MVw12	ECw10	EMw1	EMw4	EMw8
Macro-Meio	0.08	0.08	0.08	0.04	0.46	0.10	0.10	0.10	0.10
Bact-Euk	0.21	0.32	0.17	0.11	0.17	0.20	0.16	0.18	0.27
Macro-Bact	0.30	0.20	0.07	0.62	0.15	0.18	0.48	0.12	0.10
Meio-Bact	0.14	0.27	0.22	0.33	0.47	0.27	0.48	0.13	0.10
Meio-Euk	0.12	0.43	0.12	0.35	0.28	0.28	0.63	0.28	0.28
Macro-Euk	0.06	0.40	0.06	0.32	0.67	0.22	0.70	0.22	0.22
Macro-Arch	0.41	ND	0.41	0.42	0.38	0.24	0.24	0.23	0.24
Meio-Arch	0.58	ND	0.58	0.56	0.20	0.20	0.20	0.20	0.20
Arch-Euk	0.27	ND	0.27	0.80	0.27	0.27	0.80	0.27	0.27
Bact-Arch	0.47	ND	0.45	0.74	0.46	0.15	0.50	0.21	0.28
Micro-Macro	0.44	0.45	0.44	0.51	0.51	0.10	0.10	0.10	0.10
Micro-Meio	0.41	0.54	0.41	0.50	0.13	0.13	0.13	0.13	0.13
Mean	0.29	0.34	0.27	0.44	0.35	0.20	0.38	0.18	0.19
Std deviation	0.16	0.14	0.17	0.22	0.16	0.06	0.24	0.06	0.08

Table 3.7 Supplementary Table 4: List of enriched OTUs and species

Identified as relatively enriched by ALDEx2 in at least one sample type. OTUs reported here are from the q-PCR balanced microbial community and are preceded by a letter indicating the microbial domain (A=Archaea, B=Bacteria, E=Microeukarya). Associations with highT or lowT clusters in co-occurrence network (Figure 3.3) are listed on the right.

		Absolute Effect Values (ALDEx2) distinguishing X from Y				
OTU no. or size class	Taxonomy of enriched species or OTUs	HighT from diffuse	HighT from LowT	LowT from diffuse	LowT from highT	Cytoscape cluster
Macrofauna	*Depressigyra globulus* (Gastropoda)		1.34			HighT 1
Macrofauna	*Amphisamytha caldarei* (Polychaeta, Ampharetidae)				1.94	LowT
Macrofauna	*Branchinotogluma tunnicliffeae* (Polychaeta, Polynoidae)		1.09			HighT 2
Macrofauna	*Paralvinella palmiformis* (Polychaeta, Alvinellidae)		1.43			HighT 1
Macrofauna	*Paralvinella pandorae* (Polychaeta, Alvinellidae)		1.28			HighT 1
Macrofauna	*Paralvinella sulfincola* (Polychaeta, Alvinellidae)		2.09			HighT 1
Macrofauna	*Helicoradomenia juani* (Solenogastres)				1.54	LowT
Meiofauna	*Benthoxynus spiculifer* (Hexanauplia, Dirivultidae)		1.99			HighT 1
Meiofauna	Harpacticoid spp. (Hexanauplia, Harpacticoida)				1.75	LowT
Meiofauna	*Stygiopontius quadrispinosus* (Hexanauplia, Dirivultidae)		1.63			
Meiofauna	cf. *Thalassomonhystera* (Nematoda)				1.71	LowT
Meiofauna	juvenile *Lepetodrilus fucensis* (Gastropoda)				1.41	LowT
Meiofauna	juvenile *Amphisamytha caldarei* (Polychaeta, Ampharetidae)				1.79	LowT
Meiofauna	juvenile *Paralvinella palmiformis* (Polychaeta, Alvinellidae)		1.12			
A_OTU32	Archaea;Asgardaeota		1.47			HighT 2
B_OTU419	Bacteria		1.84			
B_OTU487	Bacteria			1.52	1.43	
B_OTU528	Bacteria			2.16	1.77	
B_OTU759	Bacteria			2.05	1.69	
B_OTU1586	Bacteria				1.45	
B_OTU387	Actinobacteria;Sva0996_marine_group			1.88	1.49	LowT
B_OTU295	Bacteroidetes;Bacteroidia			2.55	1.58	HighT 2
B_OTU506	Bacteroidetes;Bacteroidales BD2-2	1.80	1.81			HighT 2
B_OTU90	Bacteroidetes;Bacteroidales;Marinilabiliaceae; Carboxylicivirga	1.96				HighT 1/2
B_OTU130	Bacteroidetes;Bacteroidales;Marinilabiliaceae; Carboxylicivirga	2.27				HighT 1/2
B_OTU192	Bacteroidetes;Bacteroidales;Marinilabiliaceae; Carboxylicivirga	1.84				HighT 1/2
B_OTU292	Bacteroidetes;Bacteroidales;Marinilabiliaceae; Carboxylicivirga		1.59			HighT 1/2
B_OTU95	Bacteroidetes;VC2.1_Bac22	2.46	2.32			HighT 2
B_OTU268	Bacteroidetes;VC2.1_Bac22		1.64			HighT 2
B_OTU564	Bacteroidetes;VC2.1_Bac22		1.46			HighT 2
B_OTU165	Bacteroidetes;Cytophagales;Cyclobacteriaceae			1.77		
B_OTU271	Bacteroidetes;Cytophagales;Cyclobacteriaceae			1.47		
B_OTU121	Bacteroidetes;Cytophagales;Cyclobacteriaceae; Fulvivirga			2.03		
B_OTU860	Bacteroidetes;Cytophagales;Cyclobacteriaceae; Reichenbachiella			1.92		
B_OTU1724	Bacteroidetes;Flavobacteriales			1.45		
B_OTU246	Bacteroidetes;Flavobacteriales;Crocinitomicaceae			2.54	2.36	
B_OTU285	Bacteroidetes;Flavobacteriales;Cryomorphaceae			1.45		
B_OTU47	Bacteroidetes;Flavobacteriales;Flavobacteriaceae			1.95	1.90	
B_OTU59	Bacteroidetes;Flavobacteriales;Flavobacteriaceae			1.37	1.50	
B_OTU346	Bacteroidetes;Flavobacteriales;Flavobacteriaceae			2.17	1.37	
B_OTU434	Bacteroidetes;Flavobacteriales;Flavobacteriaceae			1.30		
B_OTU520	Bacteroidetes;Flavobacteriales;Flavobacteriaceae		1.59			
B_OTU635	Bacteroidetes;Flavobacteriales;Flavobacteriaceae			1.63		
B_OTU764	Bacteroidetes;Flavobacteriales;Flavobacteriaceae	1.83	1.66			
B_OTU547	Bacteroidetes;Flavobacteriales;Flavobacteriaceae; Euzebyella			2.28	1.65	LowT
B_OTU791	Bacteroidetes;Flavobacteriales;Flavobacteriaceae; Maribacter			1.85		
B_OTU482	Bacteroidetes;Flavobacteriales;Flavobacteriaceae; Maritimimonas		1.83			HighT 2
B_OTU133	Bacteroidetes;Flavobacteriales;Ichthyobacteriaceae; Ichthyobacterium		2.63			HighT 1
B_OTU308	Bacteroidetes;Flavobacteriales;Ichthyobacteriaceae; Ichthyobacterium		2.12			HighT 1
B_OTU97	Bacteroidetes;Sphingobacteriales	1.75	2.13			HighT 2
B_OTU45	Bacteroidetes;Sphingobacteriales;Lentimicrobiaceae		1.82			HighT 2
B_OTU93	Bacteroidetes;Sphingobacteriales;Lentimicrobiaceae		2.21			HighT 2
B_OTU153	Bacteroidetes;Ignavibacteria	3.00				
B_OTU551	Cyanobacteria;Sericytochromatia			2.63	1.83	LowT
B_OTU81	Epsilonbacteraeota;Campylobacteria;Campylobacterales		1.68			HighT 1
B_OTU267	Epsilonbacteraeota;Campylobacterales		1.43			HighT 1
B_OTU444	Epsilonbacteraeota;Campylobacterales		1.80			HighT 1
B_OTU474	Epsilonbacteraeota;Campylobacterales		1.69			HighT 1
B_OTU768	Epsilonbacteraeota;Campylobacterales		1.68			HighT 1
B_OTU834	Epsilonbacteraeota;Campylobacterales		1.84			HighT 1
B_OTU1039	Epsilonbacteraeota;Campylobacterales		1.51			HighT 1

OTU no. or size class	Taxonomy of enriched species or OTUs	HighT from diffuse	HighT from LowT	LowT from diffuse	LowT from highT	Cytoscape cluster
B_OTU2	Epsilonbacteraeota;Campylobacterales;Arcobacteraceae; Arcobacter	2.06	2.91			HighT 1
B_OTU77	Epsilonbacteraeota;Campylobacterales;Arcobacteraceae; Arcobacter		1.73			HighT 1
B_OTU138	Epsilonbacteraeota;Campylobacterales;Arcobacteraceae; Arcobacter		1.71			HighT 1
B_OTU141	Epsilonbacteraeota;Campylobacterales;Arcobacteraceae; Arcobacter		1.80			HighT 1
B_OTU490	Epsilonbacteraeota;Campylobacterales;Arcobacteraceae; Arcobacter	1.51	2.16			HighT 1
B_OTU618	Epsilonbacteraeota;Campylobacterales;Arcobacteraceae; Arcobacter		1.62			HighT 1
B_OTU852	Epsilonbacteraeota;Campylobacterales;Arcobacteraceae; Arcobacter		1.78			HighT 1
B_OTU1341	Epsilonbacteraeota;Campylobacterales;Arcobacteraceae; Arcobacter		1.72			HighT 1
B_OTU158	Epsilonbacteraeota;Campylobacterales;Nitratiruptoraceae; Hydrogenimonas		1.45			HighT 2
B_OTU143	Epsilonbacteraeota;Campylobacterales;Sulfurovaceae; Nitratifractor		1.84			HighT 1
B_OTU337	Epsilonbacteraeota;Campylobacterales;Sulfurovaceae; Nitratifractor		1.78			HighT 1
B_OTU368	Epsilonbacteraeota;Campylobacterales;Sulfurovaceae; Nitratifractor		1.70			HighT 1
B_OTU664	Epsilonbacteraeota;Campylobacterales;Sulfurovaceae; Nitratifractor	2.09	1.84			HighT 1
B_OTU1485	Epsilonbacteraeota;Campylobacterales;Sulfurovaceae; Nitratifractor		1.73			HighT 1
B_OTU198	Epsilonbacteraeota;Campylobacterales;Sulfurovaceae; Sulfurovum		1.75			
B_OTU298	Epsilonbacteraeota;Campylobacterales;Sulfurovaceae; Sulfurovum		1.73			
B_OTU300	Epsilonbacteraeota;Campylobacterales;Sulfurovaceae; Sulfurovum				1.58	
B_OTU1414	Epsilonbacteraeota;Campylobacterales;Sulfurovaceae; Sulfurovum		1.41			
B_OTU1181	Epsilonbacteraeota;Campylobacterales;Thiovulaceae; Sulfurimonas				1.54	
B_OTU42	Patescibacteria;Gracilibacteria	1.78	2.57			
B_OTU53	Patescibacteria;Gracilibacteria				1.55	
B_OTU128	Patescibacteria;Gracilibacteria		2.01			
B_OTU137	Patescibacteria;Gracilibacteria				1.57	
B_OTU149	Patescibacteria;Gracilibacteria		2.16			
B_OTU219	Patescibacteria;Gracilibacteria		1.88			
B_OTU265	Patescibacteria;Gracilibacteria		1.70			
B_OTU356	Patescibacteria;Gracilibacteria		1.68			
B_OTU960	Patescibacteria;Gracilibacteria			1.44		
B_OTU1499	Patescibacteria;Gracilibacteria			1.45		
B_OTU849	Patescibacteria;Gracilibacteria;Absconditabacteriales	1.59	1.81			
B_OTU1804	Patescibacteria;Gracilibacteria;Absconditabacteriales			1.43		
B_OTU44	Alphaproteobacteria			1.89		
B_OTU129	Alphaproteobacteria			2.04	1.39	
B_OTU172	Alphaproteobacteria			1.83		
B_OTU196	Alphaproteobacteria			2.26		
B_OTU291	Alphaproteobacteria			1.68		
B_OTU347	Alphaproteobacteria			1.43		
B_OTU365	Alphaproteobacteria			2.24		
B_OTU416	Alphaproteobacteria			2.46	1.43	
B_OTU455	Alphaproteobacteria			1.52	1.36	
B_OTU574	Alphaproteobacteria			1.58	1.37	
B_OTU763	Alphaproteobacteria			2.29	1.58	
B_OTU915	Alphaproteobacteria			1.76		
B_OTU923	Alphaproteobacteria			1.47	1.63	
B_OTU1063	Alphaproteobacteria			1.42		
B_OTU1575	Alphaproteobacteria			1.34		
B_OTU1876	Alphaproteobacteria			1.58		
B_OTU151	Alphaproteobacteria;Caulobacterales;Hyphomonadaceae; Euryhalocaulis			2.14		
B_OTU456	Alphaproteobacteria;Caulobacterales;Hyphomonadaceae; Euryhalocaulis			1.78		
B_OTU100	Alphaproteobacteria;Caulobacterales;Hyphomonadaceae; Robiginitomaculum			1.85	1.40	
B_OTU181	Alphaproteobacteria;Caulobacterales;Hyphomonadaceae; Robiginitomaculum			2.17	1.83	
B_OTU190	Alphaproteobacteria;Caulobacterales;Hyphomonadaceae; Robiginitomaculum			1.95		
B_OTU280	Alphaproteobacteria;Caulobacterales;Hyphomonadaceae; Robiginitomaculum			2.29	1.41	
B_OTU672	Alphaproteobacteria;Caulobacterales;Hyphomonadaceae; Robiginitomaculum			1.65		
B_OTU703	Alphaproteobacteria;Caulobacterales;Hyphomonadaceae; Robiginitomaculum			1.78		
B_OTU332	Alphaproteobacteria;Caulobacterales;Parvularculaceae			1.65		
B_OTU40	Alphaproteobacteria;Kordiimonadales			1.69		
B_OTU329	Alphaproteobacteria;Kordiimonadales			1.59		
B_OTU478	Alphaproteobacteria;Micavibrionales;Micavibrionaceae			1.46	1.70	

OTU no. or size class	Taxonomy of enriched species or OTUs	HighT from diffuse	HighT from LowT	LowT from diffuse	LowT from highT	Cytoscape cluster
B_OTU117	Alphaproteobacteria;Parvibaculales;PS1_clade			2.39	1.65	
B_OTU156	Alphaproteobacteria;Parvibaculales;PS1_clade			1.41	1.70	
B_OTU173	Alphaproteobacteria;Parvibaculales;PS1_clade			1.70	1.91	
B_OTU266	Alphaproteobacteria;Parvibaculales;PS1_clade			1.90	1.66	
B_OTU283	Alphaproteobacteria;Parvibaculales;PS1_clade			2.15	1.52	
B_OTU49	Alphaproteobacteria;Rhizobiales;Devosiaceae			2.23	1.76	LowT
B_OTU783	Alphaproteobacteria;Rhizobiales;Devosiaceae			1.37	1.61	LowT
B_OTU65	Alphaproteobacteria;Rhizobiales;Rhizobiaceae; Pseudahrensia			1.76		
B_OTU115	Alphaproteobacteria;Rhizobiales;Rhizobiaceae; Pseudahrensia			4.12	1.78	
B_OTU256	Alphaproteobacteria;Rhizobiales;Rhizobiaceae; Pseudahrensia			1.63	1.50	
B_OTU525	Alphaproteobacteria;Rhizobiales;Rhizobiaceae; Pseudahrensia			2.64	1.70	
B_OTU10	Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae			1.91	1.95	
B_OTU32	Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae			1.92	1.69	
B_OTU75	Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae			1.72		
B_OTU131	Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae			1.48	1.74	
B_OTU261	Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae			1.87		
B_OTU357	Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae			2.27		
B_OTU372	Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae			1.91		
B_OTU603	Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae			1.54	2.32	
B_OTU616	Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae			1.55	1.43	
B_OTU730	Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae			2.26	2.13	
B_OTU1038	Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae			1.65	1.55	
B_OTU1129	Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae			1.33		
B_OTU1137	Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae			1.51		
B_OTU29	Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Roseobacter_clade_NAC11-7			2.80	2.40	
B_OTU104	Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Roseobacter_clade_NAC11-7			2.31		
B_OTU436	Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Roseobacter_clade_NAC11-7			1.93	2.11	
B_OTU484	Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Roseobacter_clade_NAC11-7			1.31		
B_OTU522	Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Roseobacter_clade_NAC11-7			1.81	2.21	
B_OTU971	Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Roseobacter_clade_NAC11-7			1.60		
B_OTU972	Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Roseobacter_clade_NAC11-7			1.50	1.50	
B_OTU1086	Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Roseobacter_clade_NAC11-7			1.80		
B_OTU125	Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae; Sedimentitalea			1.44		
B_OTU146	Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae; Sedimentitalea			2.73	2.07	
B_OTU259	Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae; Sedimentitalea			1.67	1.97	
B_OTU435	Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae; Sedimentitalea			1.68	1.68	
B_OTU964	Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae; Sedimentitalea			2.13	1.66	
B_OTU1773	Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae; Sedimentitalea			1.41	1.57	
B_OTU134	Alphaproteobacteria;Rhodovibrionales;Kiloniellaceae			2.04		
B_OTU373	Alphaproteobacteria;Rickettsiales			2.00	1.66	
B_OTU418	Alphaproteobacteria;Rickettsiales			1.50		
B_OTU200	Deltaproteobacteria		1.68			
B_OTU575	Deltaproteobacteria			1.88		
B_OTU803	Deltaproteobacteria				1.49	
B_OTU666	Deltaproteobacteria;Bdellovibrionales;Bdellovibrionaceae;OM27_clade			1.74		
B_OTU235	Deltaproteobacteria;Desulfobacterales;Desulfobulbaceae		1.48			HighT 2
B_OTU607	Deltaproteobacteria;Desulfobacterales;Desulfobulbaceae		1.44			HighT 2
B_OTU343	Gammaproteobacteria			2.19	1.94	
B_OTU87	Gammaproteobacteria;Cellvibrionales;Halieaceae			1.65		
B_OTU684	Gammaproteobacteria;Cellvibrionales;Halieaceae			1.62		
B_OTU160	Gammaproteobacteria;Cellvibrionales;Halieaceae; Halioglobus			2.09	2.06	LowT
B_OTU14	Gammaproteobacteria;Cellvibrionales;Spongiibacteraceae; Dasania			2.46		LowT
B_OTU36	Gammaproteobacteria;Nitrosococcales;Methylophagaceae;Marine_Methylotrophic_Grp_3			2.06		
B_OTU69	Gammaproteobacteria;Nitrosococcales;Methylophagaceae;Marine_Methylotrophic_Grp_3			2.94	2.04	
B_OTU82	Gammaproteobacteria;Nitrosococcales;Methylophagaceae;Marine_Methylotrophic_Grp_3			1.93		
B_OTU399	Gammaproteobacteria;Nitrosococcales;Methylophagaceae;Marine_Methylotrophic_Grp_3			1.84	1.63	
B_OTU388	Gammaproteobacteria;Thiotrichales;Thiotrichaceae; Cocleimonas			1.70		
B_OTU277	WPS-2 (Candidate Phylum Eremiobacteraeota)			1.77		
E_OTU52	Amoebozoa;Lobosa;Tubulinea;Leptomyxida;Flabellulidae			1.57		
E_OTU55	Amoebozoa;Lobosa;Tubulinea;Leptomyxida;Flabellulidae			1.65		
E_OTU94	Amoebozoa;Lobosa;Tubulinea;Leptomyxida;Rhizamoeba; R. saxonica				1.48	
E_OTU18	Excavata;Discoba;Jakobida	2.10				HighT 1/2
E_OTU3	Opisthokonta;Fungi;Ascomycota;Pezizomycotina	2.15				HighT 2

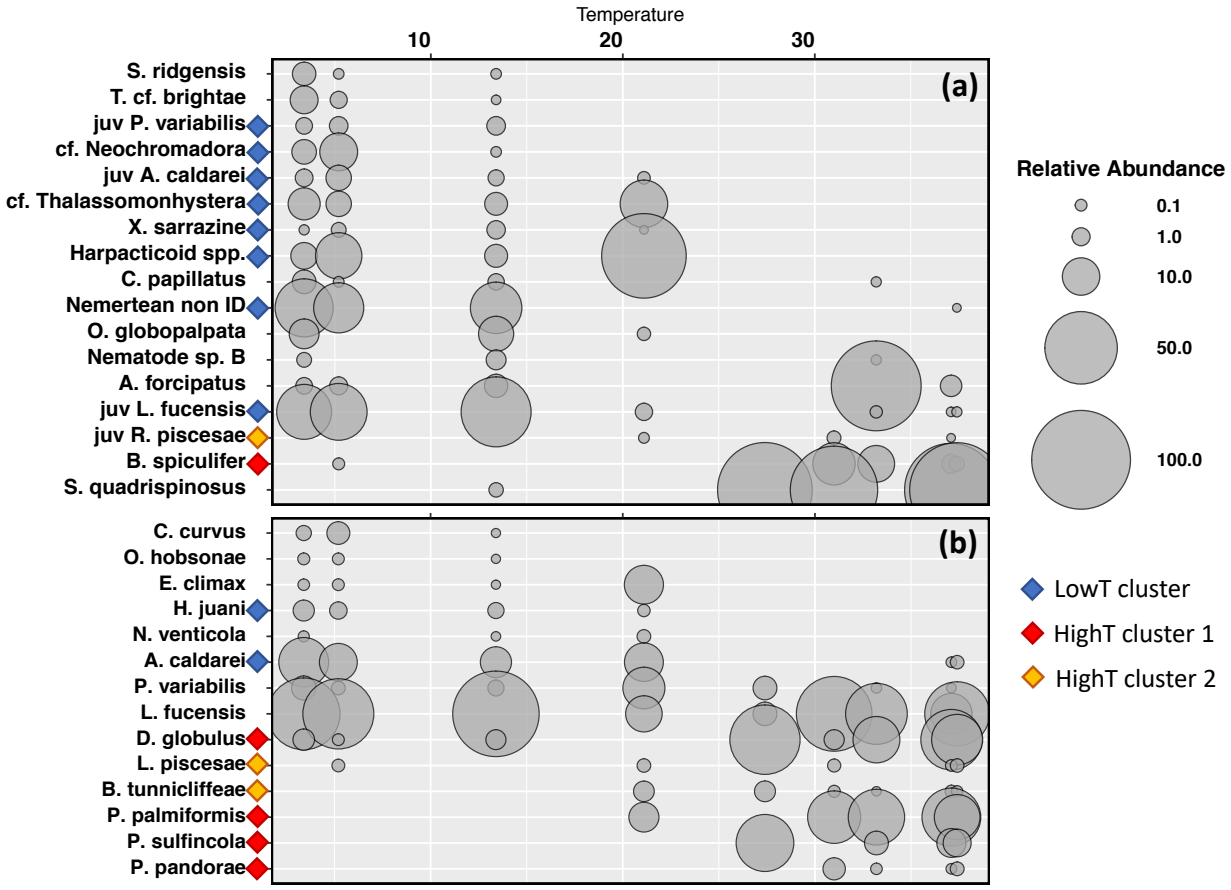


Figure 3.13 Supplementary Figure 8: Macro-/meiofaunal relative abundances

Relative abundance (within each size class) of common (a) meiofauna and (b) macrofauna across the temperature range of sampled *Ridgeia piscesae* assemblages. Taxa with colored diamonds belong to either the lowT or highT community based on enrichment or covariance with an enriched species. Red and orange colors indicate the highT network cluster in which the species falls in Figure 3.3.

Table 3.8 Supplementary Table 5: Microbial taxa with weak associations to grabs

Weak associations to highT or lowT grab samples, indicated by intermediate-sized nodes in network diagram (Figure 3.3).

		Covariance (proportionality)				Relative enrichment in X vs. Y †			
Supergroup/Phylum/ Class	Microbial taxa ‡	*Cytoscape cluster	Total no. positive ρ values	No. positive ρ with macro/meio	Average ρ with macro/meio	HighT vs. LowT grabs	HighT grabs vs. HighT diffuse)	LowT vs. HighT grabs	LowT grabs vs. LowT diffuse
Covariance with a single faunal species									
Crenarchaeota	A_or_Desulfurococcales	HighT 2	14	1	0.84				
Euryarchaeota	A_cl_Thermoplasmata	HighT 1	27	1	0.76				
Euryarchaeota	A_cl_Thermoplasmata Marine Group III	HighT 1	11	1	0.78				
Euryarchaeota	A_Methanocorpusculum	HighT 2	7	1	0.76				
Euryarchaeota	A_Pyrococcus	HighT 2	28	1	0.81				
Nanoarchaeota	A_cl_Nanohaloarchaeia DSEG	HighT 2	28	1	0.82				
Nanoarchaeota	A_cl_Woesearchaeia	HighT 1	5	1	0.95				
Acidobacteria	B_fa_Thermoanaerobaculaceae Subgroup 23	HighT 2	22	1	0.78				
Actinobacteria	B_fa_Microtrichaceae	LowT	27	1	0.77				
Alphaproteobacteria	B_Methylobacterium	HighT 1	4	1	0.75				
Bacteroidetes	B_cl_Bacteroidia	HighT 2	31	1	0.77			X	X
Bacteroidetes	B_ph_Bacteroidetes	HighT 2	29	1	0.85				
Calditrichaeota	B_Calditrichaeota	HighT 2	32	1	0.81				
Chloroflexi	B_cl_Anaerolineae	HighT 2	23	1	0.81				
Deltaproteobacteria	B_Desulfuromusa	HighT 2	27	1	0.85				
Epsilonbacteraeota	B_cl_Campylobacteria	HighT 1	26	1	0.77				
Epsilonbacteraeota	B_fa_Helicobacteraceae	HighT 1	2	1	0.78				
Epsilonbacteraeota	B_fa_Nautiliaceae	HighT 2	24	1	0.80				
Epsilonbacteraeota	B_fa_Sulfurovaceae	HighT 2	45	1	0.77				
Epsilonbacteraeota	B_Thiofractor	HighT 2	29	1	0.79				
Firmicutes	B_fa_Ruminococcaceae	HighT 2	40	1	0.86				
Nitrospinae	B_fa_Nitrospinaceae	LowT	31	1	0.78				
	B_sg_Patescibacteria	HighT 1	27	1	0.77				
Amoebozoa	E_cl_Breviatea NAMAKO-1	HighT 1	8	1	0.76				
Archaeplastida	E_Prasinoderma singularis	HighT 2	6	1	0.76				
Cercozoa	E_cl_Filosa/Imbricatea	HighT 2	3	1	0.76				
Ciliophora	E_fa_Vaginicolidae	LowT	10	1	0.78				
Dinoflagellata	E_or_Dino Group I Clade 2	HighT 2	7	1	0.77				
Enrichment only									
	B_do_Bacteria		4	0	-	X		X	X
Alphaproteobacteria	B_cl_Alphaproteobacteria		24	0	-			X	X
Alphaproteobacteria	B_Euryhalocaulis		19	0	-				X
Alphaproteobacteria	B_Robiginitomaculum		23	0	-			X	X
Alphaproteobacteria	B_fa_Parvularculaceae		33	0	-				X
Alphaproteobacteria	B_or_Kordiimonadales		12	0	-				X
Alphaproteobacteria	B_or_Micavibrionales		21	0	-				X
Alphaproteobacteria	B_fa_Micavibrionaceae		13	0	-			X	
Alphaproteobacteria	B_or_Parvibaculales PS1 clade		17	0	-			X	X
Alphaproteobacteria	B_Pseudahrensia		18	0	-			X	X
Alphaproteobacteria	B_fa_Rhodobacteraceae		16	0	-			X	X
Alphaproteobacteria	B_fa_Rhodobacteraceae NAC11-7 clade		8	0	-			X	X
Alphaproteobacteria	B_Sedimentitalea		20	0	-			X	X
Alphaproteobacteria	B_fa_Kiloniellaceae		20	0	-				X
Alphaproteobacteria	B_or_Rickettsiales		15	0	-			X	X
Bacteroidetes	B_Fulvivirga		17	0	-				X
Bacteroidetes	B_Reichenbachiella		20	0	-				X
Bacteroidetes	B_fa_Crocinitomicaceae		3	0	-			X	X
Bacteroidetes	B_fa_Cryomorphaceae		11	0	-				X
Bacteroidetes	B_fa_Flavobacteriaceae		6	0	-	X	X	X	X
Bacteroidetes	B_Maribacter		22	0	-				X
Cand. Eremiobacteraeota	B_cand_WPS-2		21	0	-				X
Cand. Gracilibacteria	B_cl_Gracilibacteria		3	0	-	X	X	X	X
Cand. Gracilibacteria	B_or_Absconditabacteriales		8	0	-	X	X		X
Deltaproteobacteria	B_cl_Deltaproteobacteria		2	0	-	X		X	X
Epsilonbacteraeota	B_Sulfurovum		7	0	-	X		X	
Epsilonbacteraeota	B_Sulfurimonas		7	0	-			X	
Gammaproteobacteria	B_fa_Halieaceae		23	0	-				X
Gammaproteobacteria	B_cl_Gammaproteobacteria		21	0	-			X	X
Gammaproteobacteria	B_fa_Methylophagaceae Marine Methylotrophic Grp3		13	0	-			X	X
Gammaproteobacteria	B_Cocleimonas		28	0	-				X
Ignavibacteriae	B_cl_Ignavibacteria		12	0	-		X		
Amoebozoa	E_fa_Flabellulidae		9	0	-				X
Amoebozoa	E_Rhizamoeba saxonica		1	0	-				X

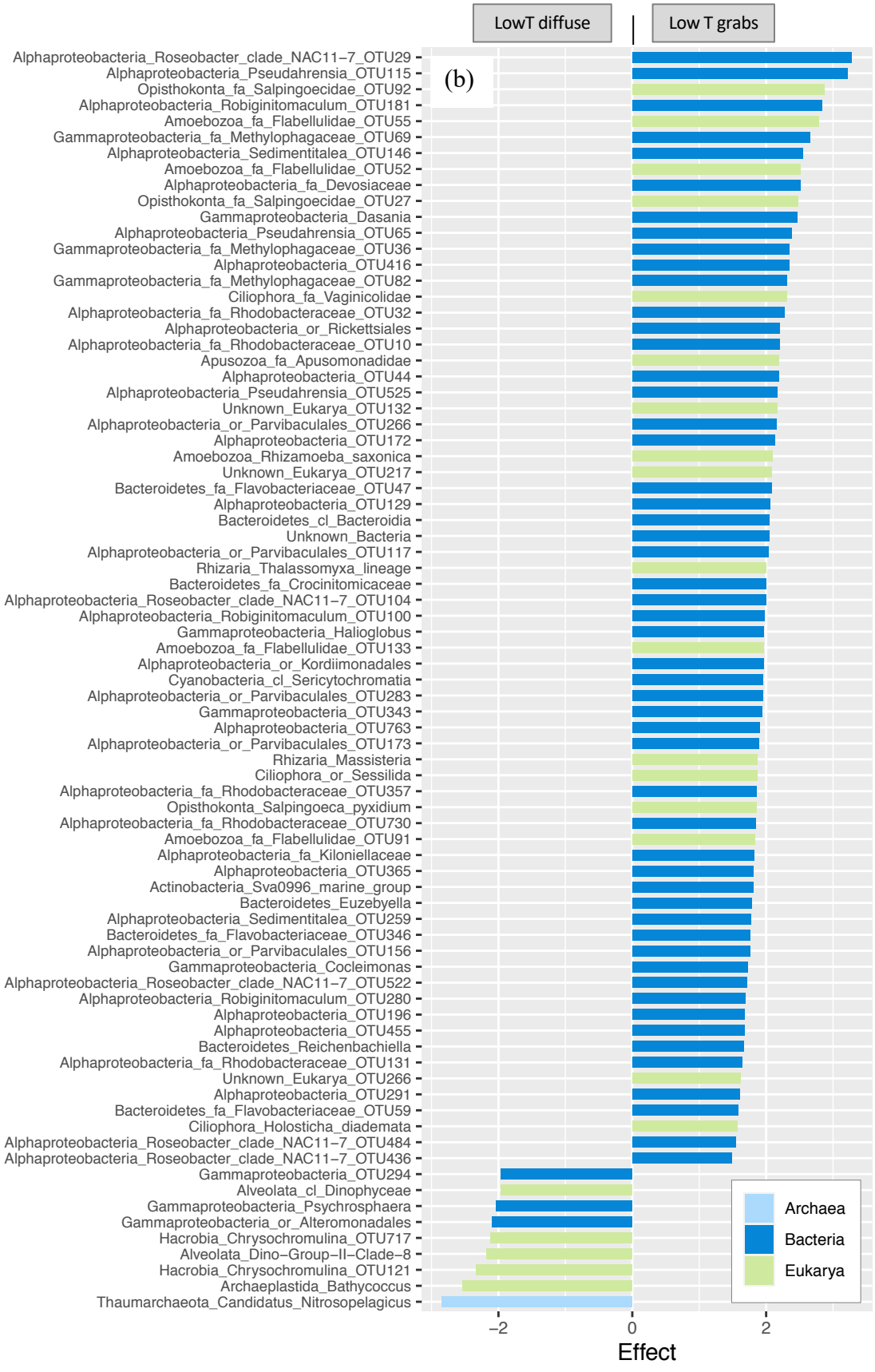
‡ Multiple OTUs binned by identical taxonomic assignment. Taxa are preceded by single letter code indicating microbial domain (A=Archaea, B=Bacteria, E=Microeukarya) and two letter code indicating taxonomic depth of classification, if other than genus (sg=supergroup, do=domain, ph=phylum, cl=class, or=order, fa=family).

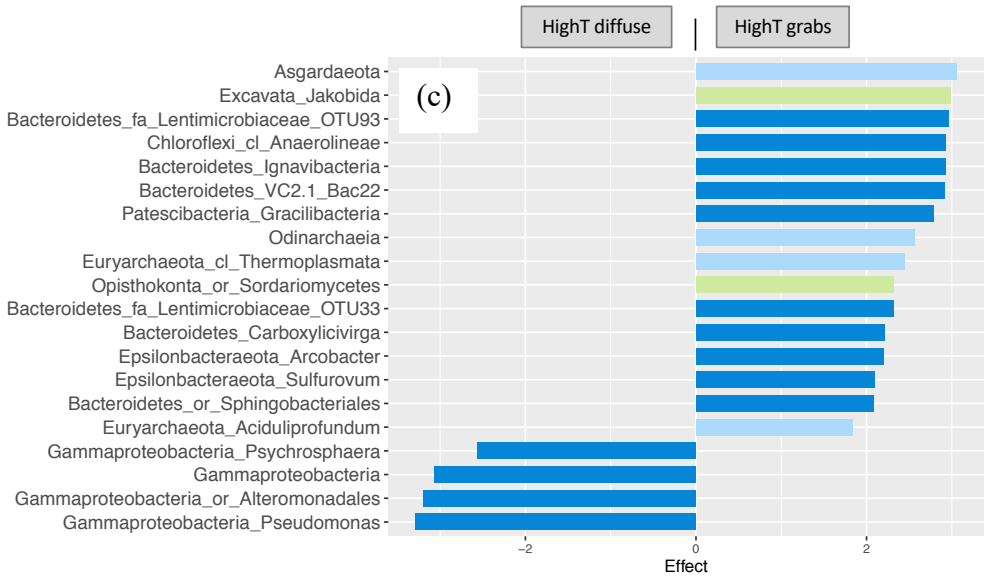
* Cluster association in network figure 3.

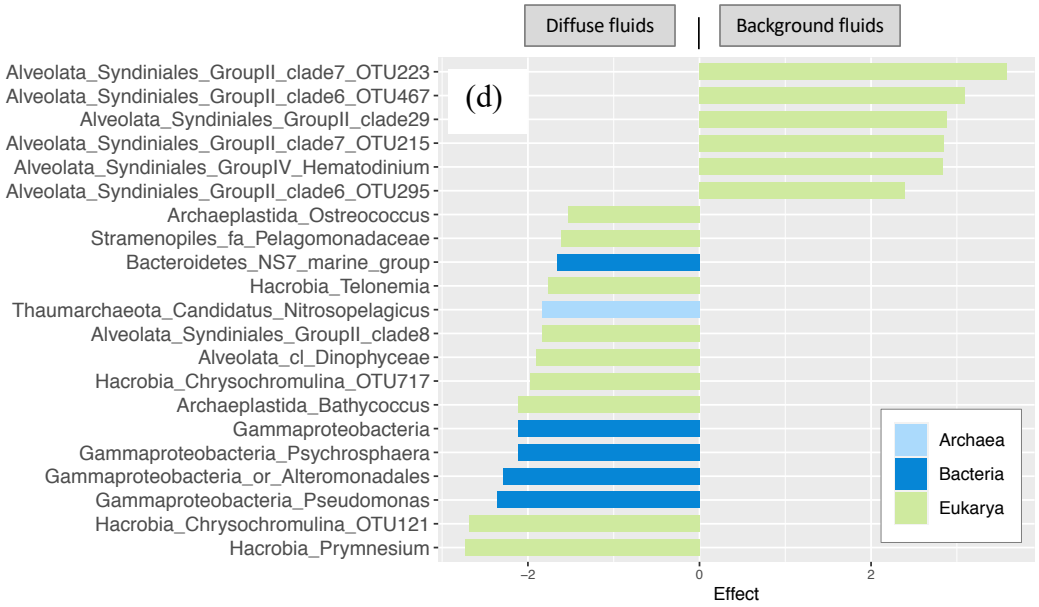
† Relative enrichment determined by ALDEx2 pairwise tests.

Figure 3.14 Supplementary Figure 9: Enriched taxa from individual domain tests

Enriched taxa identified and combined from ALDEx2 tests run on individual microbial domains. Bacteria, archaea, microeukarya, macro- and meiofauna responsible for significant differences between (a) highT and lowT (above and below 25°C, respectively) tubeworm grabs, (b) highT grabs and associated diffuse fluids, (c) lowT grabs and associated diffuse fluids, and (d) diffuse and background fluids. Effect size differences indicate relative enrichment in one sample type over the other. Only effect size differences ≥1 are shown.


LowT diffuse
Low T grabs
(b)
Alphaproteobacteria_Roseobacter_clade_NAC11−7_OTU29
Alphaproteobacteria_Pseudahrensia_OTU115
Opisthokonta_fa_Salpingoecidae_OTU92
Alphaproteobacteria_Robiginitomaculum_OTU181
Amoebozoa_fa_Flabellulidae_OTU55
Gammaproteobacteria_fa_Methylophagaceae_OTU69
Alphaproteobacteria_Sedimentitalea_OTU146
Amoebozoa_fa_Flabellulidae_OTU52
Alphaproteobacteria_fa_Devosiaceae
Opisthokonta_fa_Salpingoecidae_OTU27
Gammaproteobacteria_Dasania
Alphaproteobacteria_Pseudahrensia_OTU65
Gammaproteobacteria_fa_Methylophagaceae_OTU36
Alphaproteobacteria_OTU416
Gammaproteobacteria_fa_Methylophagaceae_OTU82
Ciliophora_fa_Vaginicolidae
Alphaproteobacteria_fa_Rhodobacteraceae_OTU32
Alphaproteobacteria_or_Rickettsiales
Alphaproteobacteria_fa_Rhodobacteraceae_OTU10
Apusozoa_fa_Apusomonadidae
Alphaproteobacteria_OTU44
Alphaproteobacteria_Pseudahrensia_OTU525
Unknown_Eukarya_OTU132
Alphaproteobacteria_or_Parvibaculales_OTU266
Alphaproteobacteria_OTU172
Amoebozoa_Rhizamoeba_saxonica
Unknown_Eukarya_OTU217
Bacteroidetes_fa_Flavobacteriaceae_OTU47
Alphaproteobacteria_OTU129
Bacteroidetes_cl_Bacteroidia
Unknown_Bacteria
Alphaproteobacteria_or_Parvibaculales_OTU117
Rhizaria_Thalassomyxa_lineage
Bacteroidetes_fa_Crocinitomicaceae
Alphaproteobacteria_Roseobacter_clade_NAC11−7_OTU104
Alphaproteobacteria_Robiginitomaculum_OTU100
Gammaproteobacteria_Halioglobus
Amoebozoa_fa_Flabellulidae_OTU133
Alphaproteobacteria_or_Kordiimonadales
Cyanobacteria_cl_Sericytochromatia
Alphaproteobacteria_or_Parvibaculales_OTU283
Gammaproteobacteria_OTU343
Alphaproteobacteria_OTU763
Alphaproteobacteria_or_Parvibaculales_OTU173
Rhizaria_Massisteria
Ciliophora_or_Sessilida
Alphaproteobacteria_fa_Rhodobacteraceae_OTU357
Opisthokonta_Salpingoeca_pyxidium
Alphaproteobacteria_fa_Rhodobacteraceae_OTU730
Amoebozoa_fa_Flabellulidae_OTU91
Alphaproteobacteria_fa_Kiloniellaceae
Alphaproteobacteria_OTU365
Actinobacteria_Sva0996_marine_group
Bacteroidetes_Euzebyella
Alphaproteobacteria_Sedimentitalea_OTU259
Bacteroidetes_fa_Flavobacteriaceae_OTU346
Alphaproteobacteria_or_Parvibaculales_OTU156
Gammaproteobacteria_Cocleimonas
Alphaproteobacteria_Roseobacter_clade_NAC11−7_OTU522
Alphaproteobacteria_Robiginitomaculum_OTU280
Alphaproteobacteria_OTU196
Alphaproteobacteria_OTU455
Bacteroidetes_Reichenbachiella
Alphaproteobacteria_fa_Rhodobacteraceae_OTU131
Unknown_Eukarya_OTU266
Alphaproteobacteria_OTU291
Bacteroidetes_fa_Flavobacteriaceae_OTU59
Ciliophora_Holosticha_diademata
Alphaproteobacteria_Roseobacter_clade_NAC11−7_OTU484
Alphaproteobacteria_Roseobacter_clade_NAC11−7_OTU436
Gammaproteobacteria_OTU294
Alveolata_cl_Dinophyceae
Gammaproteobacteria_Psychrosphaera
Gammaproteobacteria_or_Alteromonadales
Hacrobia_Chrysochromulina_OTU717
Alveolata_Dino−Group−II−Clade−8
Hacrobia_Chrysochromulina_OTU121
Archaeplastida_Bathycoccus
Thaumarchaeota_Candidatus_Nitrosopelagicus
Archaea
Bacteria
Eukarya
−2
0
2
Effect


HighT diffuse
HighT grabs
(c)
Asgardaeota
Excavata_Jakobida
Bacteroidetes_fa_Lentimicrobiaceae_OTU93
Chloroflexi_cl_Anaerolineae
Bacteroidetes_Ignavibacteria
Bacteroidetes_VC2.1_Bac22
Patescibacteria_Gracilibacteria
Odinarchaeia
Euryarchaeota_cl_Thermoplasmata
Opisthokonta_or_Sordariomycetes
Bacteroidetes_fa_Lentimicrobiaceae_OTU33
Bacteroidetes_Carboxylicivirga
Epsilonbacteraeota_Arcobacter
Epsilonbacteraeota_Sulfurovum
Bacteroidetes_or_Sphingobacteriales
Euryarchaeota_Aciduliprofundum
Gammaproteobacteria_Psychrosphaera
Gammaproteobacteria
Gammaproteobacteria_or_Alteromonadales
Gammaproteobacteria_Pseudomonas
−2
0
2
Effect


Diffuse fluids
Background fluids
(d)
Alveolata_Syndiniales_GroupII_clade7_OTU223
Alveolata_Syndiniales_GroupII_clade6_OTU467
Alveolata_Syndiniales_GroupII_clade29
Alveolata_Syndiniales_GroupII_clade7_OTU215
Alveolata_Syndiniales_GroupIV_Hematodinium
Alveolata_Syndiniales_GroupII_clade6_OTU295
Archaeplastida_Ostreococcus
Stramenopiles_fa_Pelagomonadaceae
Bacteroidetes_NS7_marine_group
Hacrobia_Telonemia
Thaumarchaeota_Candidatus_Nitrosopelagicus
Alveolata_Syndiniales_GroupII_clade8
Alveolata_cl_Dinophyceae
Hacrobia_Chrysochromulina_OTU717
Archaeplastida_Bathycoccus
Gammaproteobacteria
Gammaproteobacteria_Psychrosphaera
Gammaproteobacteria_or_Alteromonadales
Gammaproteobacteria_Pseudomonas
Hacrobia_Chrysochromulina_OTU121
Hacrobia_Prymnesium
Archaea
Bacteria
Eukarya
−2
0
2
Effect

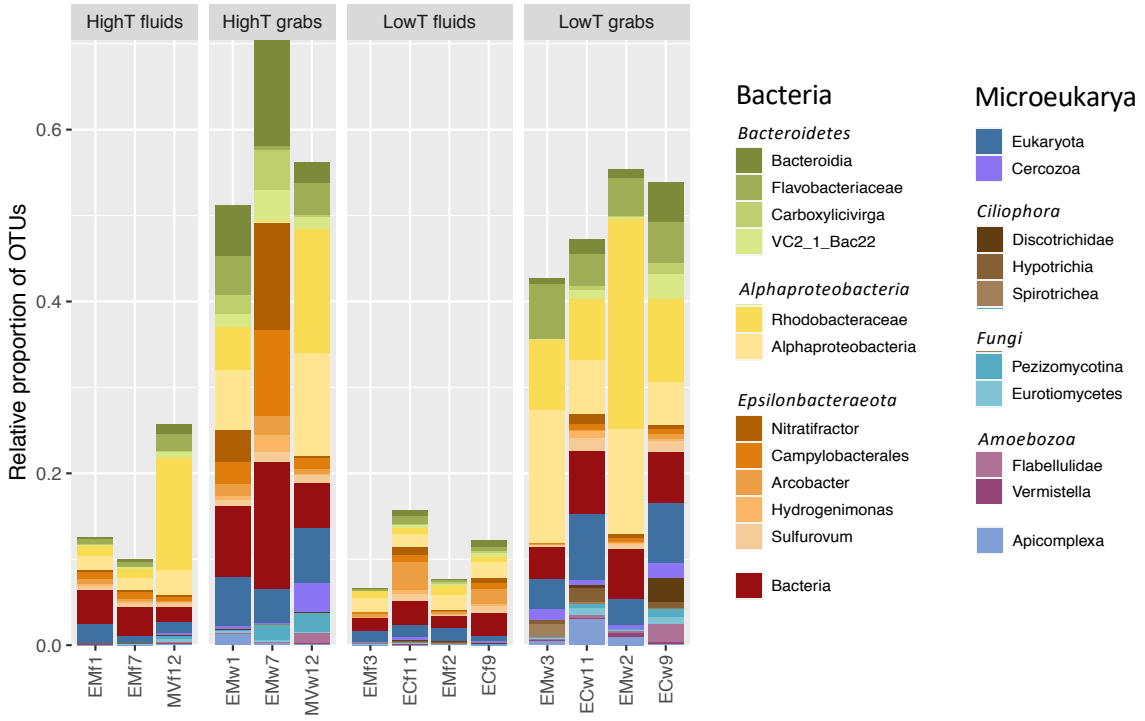


Figure 3.15 Supplementary Figure 10: Taxa contributing to increased richness

Largest microbial contributors to OTU richness differences between fluid and grab samples.

Chapter 4. Community functional characterization of tubeworm-associated microbial assemblages

4.1 Introduction

Understanding how ecosystems function is necessary for defining the normal range of ecosystem states, recognizing the effects of environmental stress, and deciding if and how to implement management and protection measures. Growing awareness of the vital roles microorganisms play in ecosystem function and in shaping natural systems in general, the latter largely driven by increasing knowledge of the connections between human microbiomes and health (Young, 2017), has led to increased efforts to define microbial roles in ecosystem health (Handley, 2019). The vast majority of microbial species in nature remain uncultured (Amann et al., 1995), thus limiting our ability to characterize and understand microbial processes and explore their contributions within ecosystems. Metagenomics offers a culture-independent alternative characterization method that can reveal details of microbial functional potential at the community level (Riesenfeld et al., 2004; Schloss and Handelsman, 2005), and even at species or strain levels through genome binning and creation of metagenome-assembled genomes (MAGs) (Hugerth et al., 2015).

Metagenomic studies from hydrothermal vent habitats have provided insight into potential microbial involvement in various elemental biogeochemical cycles. These studies have evolved, from metagenomic characterizations of functional potential (Brazelton and Baross, 2009; Xie et al., 2011; Anderson et al., 2014; Reveillaud et al., 2016) to more meaningful identifications of the active microbial players involved in specific processes by

incorporating metatranscriptomics, metaproteomics and genome binning (Urich et al., 2014; Fortunato and Huber, 2016; Meier et al., 2016; Fortunato et al., 2018; Galambos et al., 2019; Trembath-Reichert et al., 2019). For example, an early metagenome study described a sulfide chimney-associated microbial community with the potential for coupling sulfur oxidation to nitrate reduction to fuel carbon fixation (Xie et al., 2011). Later studies from diffuse fluids and hydrothermal sediments confirmed the activity of autotrophic Epsilonbacteraeota involved in sulfur and nitrogen cycling and carbon fixation via the reductive tricarboxylic acid (rTCA) cycle and archaeal expression of genes involved in methanogenesis (Urich et al., 2014; Fortunato and Huber, 2016).

To date, so called 'omics research at hydrothermal vents has directed little effort at characterizing functions of microbes associated with faunal assemblages. This is likely attributable to a tendency to view faunal assemblage microbiomes as symbiont-dominated (Dick, 2019), and a general trend in vent 'omics research to focus on autotrophic processes and overlook microbial heterotrophy. This predominant view of faunal assemblage microbiomes has limited exploration of an extremely rich reservoir of microbial diversity in hydrothermal systems, such as the high levels of non-endosymbiotic microbial richness in faunal assemblages demonstrated in Chapter 3. Vent 'omics research has also been inclined to focus on microbes in diffuse fluids owing to the foundational role of subsurface chemolithoautotrophs, abundant in diffuse venting fluids, in supporting life in these systems (Sievert and Vetriani, 2012), the convenience of sampling microbes in fluids as opposed to separating them from faunal aggregations and rock surfaces, and the challenging heterogeneity of benthic habitat around vent openings.

Findings in Chapter 3 indicate an important role for vent habitat in shaping non-

symbiotic microbial assemblages. Indeed, even in a fluids-only study Meier et al. (2016) identified shifts in the taxonomic composition and functional capacity of microbes along a dilution gradient away from vent openings where Epsilonbacteraeota-dominated autotrophs gave way to Gamma- and Alphaproteobacteria-dominated heterotrophs. This parallels the findings of Chapter 3 that revealed not only a shift from Epsilonbacteraeota to Alpha- and Gammaproteobacteria between highT and lowT assemblages, but also an abundance of heterotrophic Bacteroidetes in both habitats as opposed to their respective fluids. This suggests a shift in the autotrophy-heterotrophy balance over the environmental range of these faunal habitats and the potential for faunal assemblages to play a major role in organic matter remineralization in these systems. Together, these observations highlight the need for a functional assessment of microbes within hydrothermal vent faunal habitats to determine their potential roles in ecosystem function.

Chapter 4 aims to develop approaches for understanding the functional roles of non-endosymbiotic microbes living in faunal assemblages at deep-sea hydrothermal vents based on metagenomic sequencing and comparative analyses of metagenomes. Both exploratory and targeted function-related approaches are used to develop a baseline characterization of microbial functional potential within Juan de Fuca Ridge tubeworm assemblages. The exploratory approach emphasizes functional processes that distinguish diffuse fluids and highT and lowT tubeworm habitats by identifying genes enriched in representative samples. This approach takes into account both taxonomic and functional annotation of genes to assess the relative contributions of core and non-core community taxa to habitat-distinguishing functional pathways. The targeted approach aims to create metrics of community functional capacity within a few selected biogeochemical cycles that may

represent various ecosystem states and can be compared between sites or at the same site over time.

4.2 Methods

Harvesting of non-endosymbiotic microbial assemblages from tubeworm bushes and DNA extraction methods are detailed in Chapter 3.

4.2.1 Metagenomic sequencing and data handling

Shotgun metagenomic sequencing of 20 genomic DNA extracts from diffuse fluids and tubeworm-associated assemblages was performed at the Integrated Microbiome Resource Facility (Dalhousie University) on an Illimina NextSeq550. Samples were spread over three separate sequencing runs to ensure deep coverage of highly diverse microbial communities. Raw sequence data were filtered to remove low-quality reads using BBDuk (sourceforge.net/projects/bbmap/). High-quality short reads were assembled into longer contigs using MEGAHIT (Li et al., 2015) and submitted to the Joint Genome Institute's Integrated Microbial Genomes and Metagenomes (IMG/M) pipeline for Open Reading Frame (ORF) identification and taxonomic and functional annotation (Chen et al., 2017). To determine the number of reads (coverage) belonging to each ORF in the annotated metagenomes, short reads were mapped onto ORFs using Bowtie2 (Langmead and Salzberg, 2012).

4.2.2 Comparing functional compositions

Gene compositions of the microbial assemblages were compared among samples as in

Chapter 3 with zero-count replacement, centered log-ratio transformation of the data, and distances between samples calculated using the Aitchison metric. Relationships between samples were visualized by NMDS and hierarchical clustering. Explanatory habitat variables (temperature, substratum, location, and sample type) were tested using ***envfit*** on the NMDS ordination and using ANOSIM tests. Functional diversity was quantified by the Inverse Simpson diversity metric. Differences in functional diversity between diffuse fluid and grab samples and between highT and lowT grabs were assessed by Mann-Whitney tests.

Congruence of functional and taxonomic composition of samples was measured by Procrustes analysis (***procrustes*** function in R) using the functional NMDS ordination and the taxonomic NMDS ordination from chapter 3. Significance of the result was determined using the ***protest*** function with 1000 permutations.

4.2.3 Identifying enriched genes and taxa

Enrichment is an expression of differential relative abundance in one group over another. Enrichment tests were performed using the ALDEx2 R package (Fernandes et al., 2013) in two phases. Phase 1 tested for enriched genes that differentiated between sample types (diffuse fluids, highT grabs, lowT grabs). Phase 2 tested for enriched taxa that contributed to differences between sample types within specific functional pathways. In phase 1, a sample-by-gene-count matrix was used to identify enriched genes in three different pairwise comparisons—between highT and lowT grabs and between each grab type and their associated diffuse fluids. In phase two, enriched genes belonging to selected functional pathways were binned by taxonomic assignment and compiled into pathway specific sample-by-taxa matrices. Enrichment tests were run on these pathway specific matrices

comparing only diffuse fluid versus grab samples and highT versus lowT grab samples. The selection process involved in prioritizing pathways for phase 2 tests is detailed below in the section entitled '*Community functional patterns across habitat types*'.

Some variability is expected in the results between ALDEx2 runs due to random sampling of the data. To quantify this variability, the same data input was run six times comparing highT (n=6) and lowT (n=5) grab samples. To assess the potential bias in enrichment results when using groups of unequal size, ALDEx2 results comparing highT (n=6) to lowT (n=5) assemblages were compared with results from six additional runs with systematic removal of each of the highT samples (i.e., with n=5 for both highT and lowT).

4.2.4 Community functional patterns across habitat types

Given the overwhelming amount of information in metagenomes, the path from genes to meaningful functional distinctions between the three sample types required a mechanism to prioritize the potential importance of functional pathways in each sample type. "Enrichment leanings" were developed as a metric of the tendency of functional pathways towards enrichment in highT or lowT grabs or diffuse fluids. Figure 4.1 provides an example and detailed explanation of the calculation and application of enrichment leanings. The calculation of enrichment leanings relies on the nature of results from pairwise ALDEx2 tests, which are positive when enriched in one group and negative when enriched in the other. Therefore, pathways containing genes with mixed enrichment in different sample types will have weaker average enrichment scores than pathways with consistent enrichment of genes in one sample type. This pathway prioritization is not intended as a statistically rigorous assessment of pathway relevance, but rather as a refinement tool to

create a hierarchical framework for pursuing more in-depth functional explorations.

The importance of the 'core community' prokaryote taxa identified in Chapter 3 to the selected functional pathways was evaluated using Principal Components Analysis (PCA; ***prcomp*** function in R) on the gene counts from each taxon in each of the three sample types. In other words, for each included pathway, each core community prokaryote taxon had a gene abundance for highT grabs, lowT grabs, and diffuse fluids. In cases where a core community member had no sequenced genome in the reference database, genes identified to all genera within the family were included. Likewise, for core community members that were only identified at the family level. Core community members identified to order or higher were not included in this comparison. Several additional taxa were included that often contributed a large portion of the genes in the pathway but were not part of the core community. A heatmap of taxonomic contributions to pathways in each of the three sample types was created with the R ***heatmap*** function using the same gene count data used in PCA.

4.2.5 Genes for community functional metrics

Two metrics were created to explore the usefulness of community-level descriptors for defining the range of microbial ecosystem states in faunal habitat. The first metric aimed to quantify the relative balance of autotrophy to heterotrophy and was expressed as the ratio of gene abundances involved in all six major carbon fixation pathways (details below) to gene abundances coding for proteases, lipases, chitinases, and glucosidases. For this metric genes were normalized to the total numbers of genes with a predicted product in each sample to create a meaningful metric. The second set of community level metrics assessed the balance of genes involved in the major process modules of nitrogen (N) cycling (N

fixation, nitrification, denitrification, and dissimilatory and assimilatory nitrate reduction), sulfur cycling (thiosulfate oxidation and dissimilatory and assimilatory sulfate reduction), and carbon fixation (Calvin-Benson-Bassham cycle, reductive tricarboxylic acid cycle, dicarboxylate-hydroxybutylate cycle, hydroxypropionate-hydroxybutylate cycle, 3-hydroxypropionate bi-cycle, and reductive acetyl-CoA pathway). For this metric, the clr-transformed gene abundances were used.

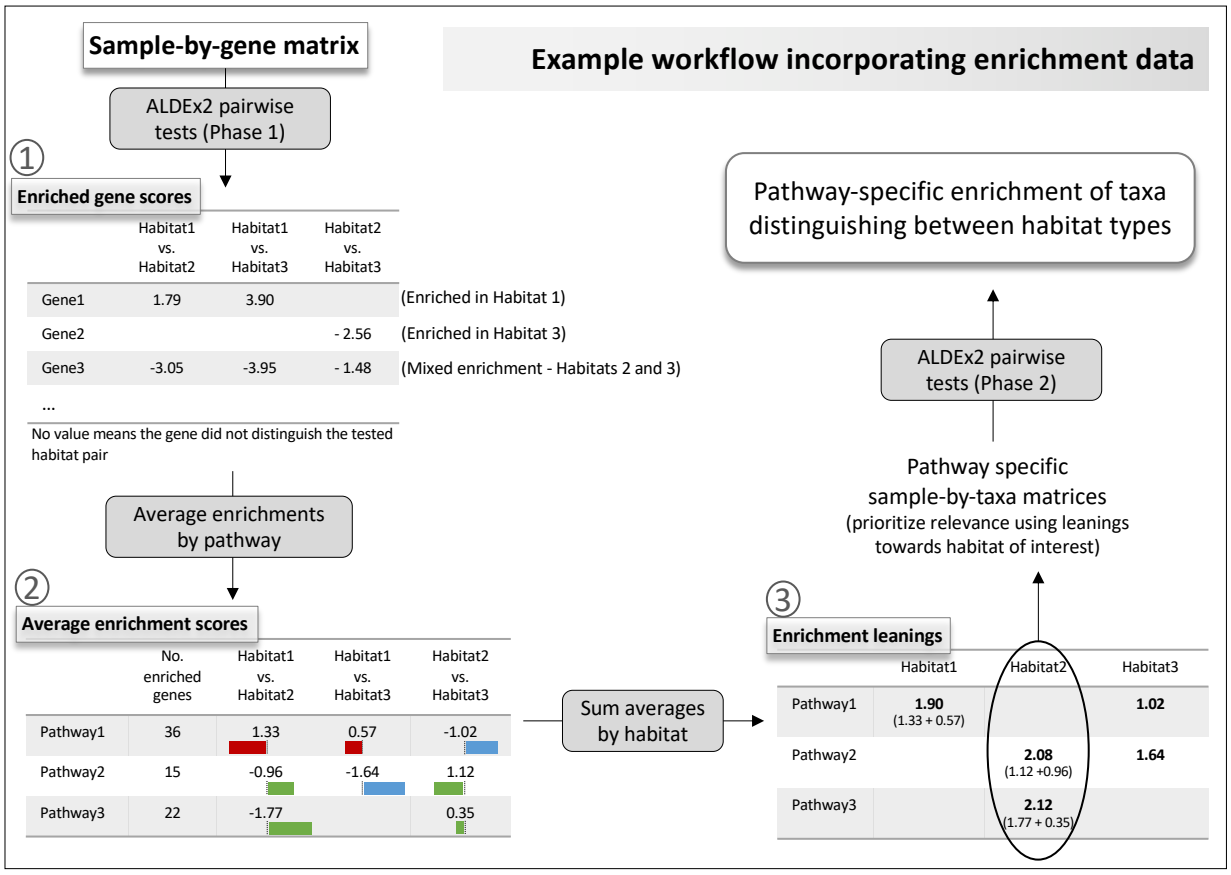


Figure 4.1 Enrichment example

Methods used to merge gene-level enrichment scores into enrichment leanings as a tool to prioritize the potential importance of functional pathways in different habitat types. Using the list of enriched genes (1) compiled from the three pairwise enrichment tests, average scores (2) were calculated from all enriched genes involved in each functional pathway—one average for each of the three pairwise tests. In some cases, there was no enrichment of genes in a pathway in one or two of the three tests. Enrichment leanings towards each of the three habitats (3) are the summed absolute values of the average enrichment scores from the two tests involving each habitat.

Table 4.1 Metagenome assembly and annotation metrics

					Assembled Scaffolds									With Coverage		
										Genes with product name		Genes identified with KEGG				
Sample	Sample Type	Million reads (before assembly)	Assembled scaffolds	Assembled bases (million)	Average scaffold length	GC%	Total genes	RNA genes (rRNA + tRNA)	Protein coding genes	No.	% *	No.	% *	Total genes	Protein coding genes	Genes per million reads
EMw1	HighT grab	155.45	1178704	805.88	684	40.5%	1580995	12628	1568367	868369	54.9%	547104	34.6%	10896347	10665667	70097
EMw2	LowT grab	109.60	1214268	710.74	585	43.7%	1544148	8547	1535601	884685	57.3%	559415	36.2%	6479652	6345888	59122
EMw4	HighT grab	160.22	1009811	793.03	785	39.1%	1517541	12163	1505378	810362	53.4%	488549	32.2%	15329283	14951136	95678
EMw5	LowT grab	117.40	922966	495.45	537	40.2%	973697	6015	967682	477063	49.0%	302235	31.0%	6059715	5882173	51614
EMw7	HighT grab	93.64	746791	612.14	820	41.0%	1130151	10713	1119438	648311	57.4%	396357	35.1%	9642508	9370274	102976
EMw8	HighT grab	96.62	913696	647.31	708	40.7%	1232360	10808	1221552	663330	53.8%	406327	33.0%	8714258	8471734	90187
ECw9	LowT grab	132.95	1258268	750.07	596	38.3%	1128240	4333	1123907	385425	34.2%	238947	21.2%	6178235	6056677	46470
ECw10	HighT grab	149.88	1074121	853.57	795	42.1%	1582049	11626	1570423	897512	56.7%	544024	34.4%	13469964	13163334	89869
ECw11	LowT grab	96.55	833127	558.08	670	43.5%	1146712	7368	1139344	675536	58.9%	422711	36.9%	7197206	7063145	74547
MVw12	HighT grab	99.88	565054	565.29	1000	43.0%	966569	8847	957722	582605	60.3%	361830	37.4%	11098717	10780896	111116
MVw13	LowT grab	146.38	1470432	902.42	614	42.7%	1874488	10887	1863601	1040012	55.5%	623069	33.2%	9509495	9306997	64966
EMf1	diffuse fluids	131.94	738170	494.23	670	37.1%	1034508	10217	1024291	646752	62.5%	438154	42.4%	14131892	13746450	107105
EMf2	diffuse fluids	110.23	926448	596.61	644	39.5%	1305325	11958	1293367	849628	65.1%	595662	45.6%	7438544	7290756	67485
EMf3	diffuse fluids	130.78	903436	550.60	609	38.4%	1239919	10679	1229240	781262	63.0%	543599	43.8%	7912557	7740539	60501
EMf7	diffuse fluids	144.70	337632	179.68	532	38.5%	442145	3515	438630	290864	65.8%	200926	45.4%	3849441	3811911	26603
ECf9	diffuse fluids	111.19	859146	563.94	656	40.5%	1196606	10539	1186067	743649	62.1%	486221	40.6%	9126099	8900337	82073
ECf11	diffuse fluids	155.20	961336	560.53	583	37.8%	1287057	8258	1278799	804285	62.5%	532898	41.4%	8205663	8046131	52870
MVf12	diffuse fluids	140.56	476437	370.20	777	41.0%	647066	5041	642025	363836	56.2%	231605	35.8%	12089471	11577758	86009

* Anomalously low percentage in red.

4.3 Results

4.3.1 Metagenome summary

Metagenomes were produced for all 20 diffuse fluid and tubeworm-associated samples, however, two lowT samples (EMw3 and EMw6) failed the IMG/M annotation step and will not be discussed further here. After quality filtering, the remaining 18 metagenomes each averaged ca. 127 million (M) reads (Table 4.1) and assembled contigs from each sample contained an average 1.2 M protein coding genes. Functionally annotated protein coding genes had an average per sample prediction rate of 59 (±5)% (Table 4.1) with the exception of sample ECw9, from which only 34% of genes had a predicted product. Functional annotation utilizes multiple reference databases to predict the products of protein-coding genes. Data interpretations will focus on genes identified with Kyoto Encyclopedia of Genes and Genomes (KEGG) Orthology terms.

Significant differences occurred between diffuse fluids and grab samples in some metagenome annotation metrics. Specifically, grab samples had lower percentages of genes with an identified product relative to diffuse fluids (Figure 4.2a,b), and higher G+C content (percentage of guanine and cytosine nucleotides) (Figure 4.2c). Greater than 90% of protein coding genes in each sample had a bacterial origin, and archaeal genes contributed 3-7% in diffuse fluid samples. Viral genes, although they made up only ca. 0.1% of genes on average, were also more abundant (0.4-0.8%) in diffuse fluid samples. The few genes belonging to Eukarya were mostly related to metazoans.

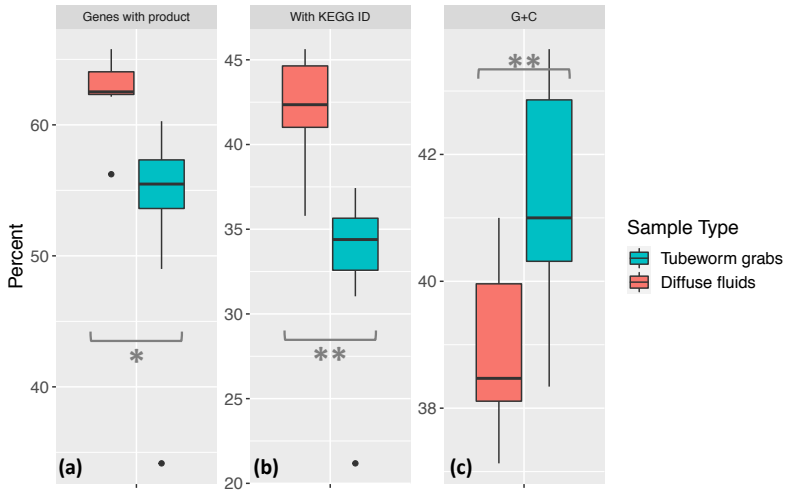


Figure 4.2 Metagenome annotation metrics

Differences in metagenome annotation metrics between diffuse fluid and grab samples. (a) Percent of genes with an identified product, (b) percent of genes identified with the KEGG Orthology database, and (c) nucleotide GC content. Significance values from Mann-Whitney tests: * $p<0.05$, ** $p<0.005$.

4.3.2. Functional composition and diversity

The functional composition of the diffuse fluid metagenomes was generally distinct from the grab samples with the exception of the highT sample from Middle Valley (MVf12), which clustered with the associated grab sample (MVw12) in cluster analysis and NMDS (Figure 4.3). Four of the five remaining highT assemblages clustered together but sample EMw1 was more similar to lowT grabs from Middle Valley and Clam Bed. Clustering of the grab sample metagenomes was less driven by temperature than the taxonomic composition reported in Chapter 3, although Procrustes analysis indicated significant congruence (correlation=0.55, p=0.009) between the function-based and taxonomy-based NMDS. The functional composition of ECw9 was distinct from all other samples.

Two habitat variables had a significant influence on the functional composition of samples, but the strength of influence differed between the results of the ***envfit*** and ANOSIM tests. Temperature and sample type (fluids or grabs) were both significant predictors of

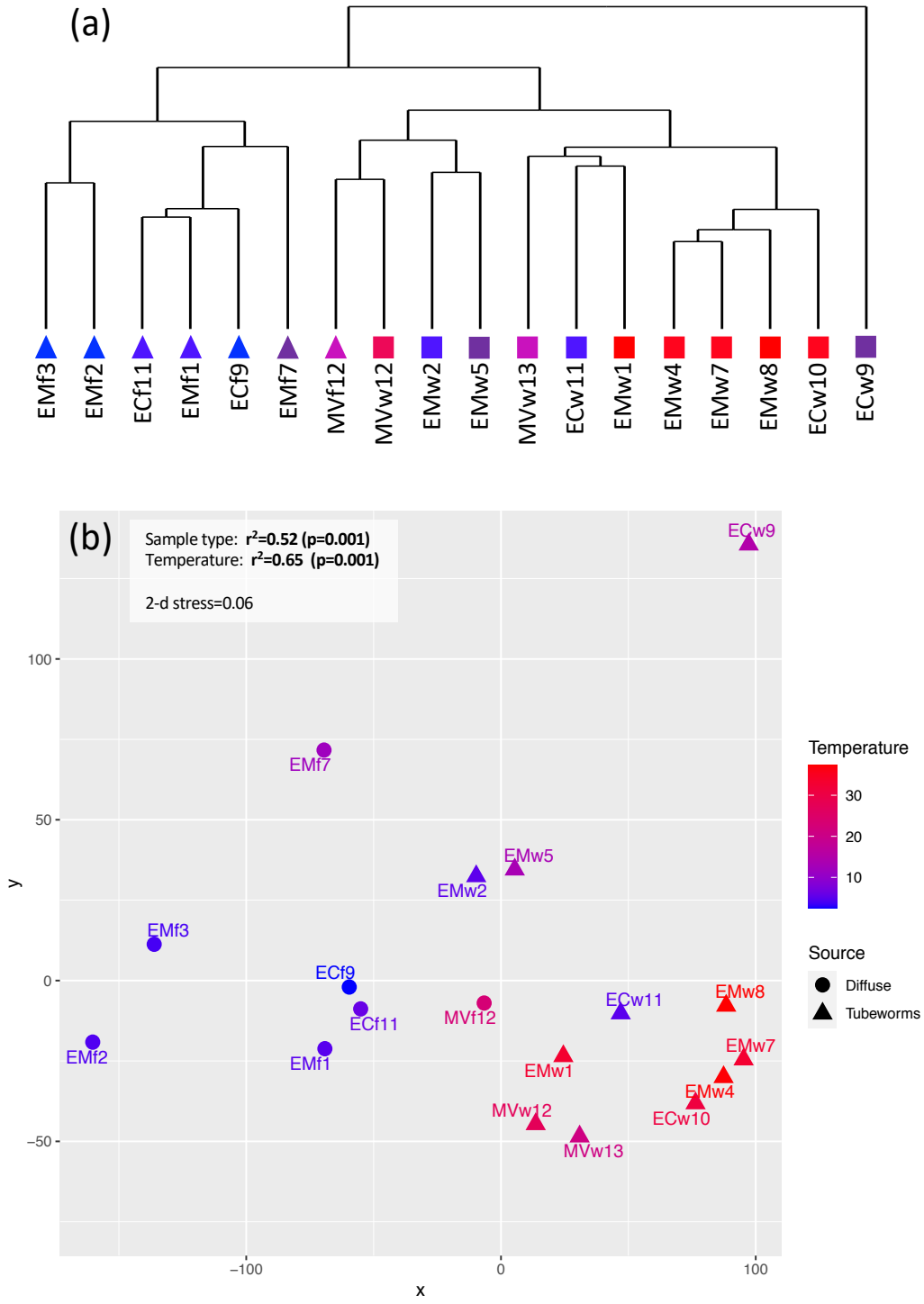


Figure 4.3 Clustering and NMDS

Clustering of samples based on the full gene composition using a) hierarchical clustering and b) Nonmetric Multidimensional Scaling. The strength of fit of sample type and temperature variables to the NMDS are inset in b).

functional composition among samples, but ***envfit*** reported temperature as the strongest predictor of sample clustering in the NMDS plot (r^2=0.65, p=0.001) followed by sample type (r^2=0.52, p=0.001) (Figure 4.3b). ANOSIM, which compares full functional compositions rather than only relationships between samples that can be represented in two-dimensions, reported sample type as the stronger predictor of functional composition (sample type R=0.60, p=0.001; highT/lowT R=0.29, p=0.013). Substratum and location were not significant predictors of functional composition.

Inverse Simpson diversity of gene composition had a slightly lower median value in highT grabs than in either lowT grabs or fluids, but none of the three sample types differed significantly from the others (Figure 4.4).

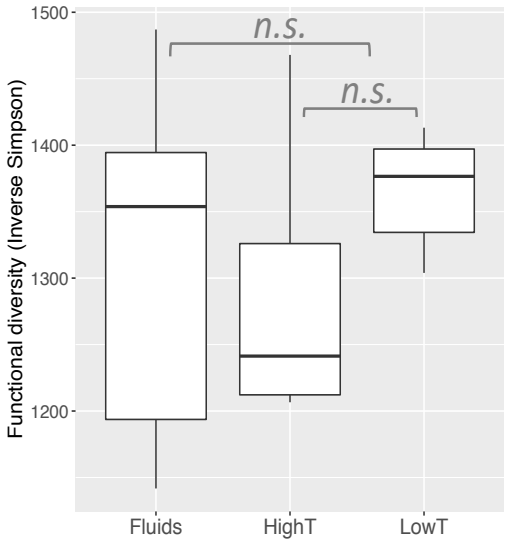


Figure 4.4 Inverse Simpson diversity

Functional diversity values among the three sample types. Mann-Whitney tests indicated no significant differences (*n.s.*) between highT and lowT grabs or between grabs and diffuse fluids.

4.3.3 Functional categories

Out of 19,163 KEGG Orthology terms used by the IMG/M annotation pipeline (note: there are 4,675 KEGG terms missing from the current version of the annotation tool), 10,339 were present in the metagenomes (Figure 4.5a). Of the genes present, 29% were only shallowly characterized under the BRITE hierarchy system, which classifies genes to broad groups of proteins (e.g., transporters). A further 13% fell into categories of unclassified genes. The remaining 5,993 genes fell into four major KEGG categories: metabolism (53%), genetic information processing (17%), environmental information processing (23%), and cellular processes (16%). Many genes are involved in multiple processes so gene affiliations with categories sum to greater than 100%.

Within some KEGG sub-categories, nearly all of the database genes were represented in the metagenomes (Figure 4.5a). The most thoroughly represented group of genes was the Poorly Characterized category with 97% of databases genes present. The next most represented categories were metabolic, including nucleotide metabolism, with 90% of genes present, and carbohydrate, energy, and amino acid metabolisms each with 85-87% of genes present. Transcription genes were also well represented in the metagenomes, with 84% of genes present.

4.3.4 Enriched genes in functional categories

Tests of identical repeat ALDEx2 runs revealed variation of 5.8(±0.9)% in the enriched genes identified. Genes that were not consistently enriched across all runs had an average effect size of 1.07 (± 0.08). To account for this variation, a value of 1.15 was implemented as the lower threshold score for reliable enrichment. Variation in enriched genes was much

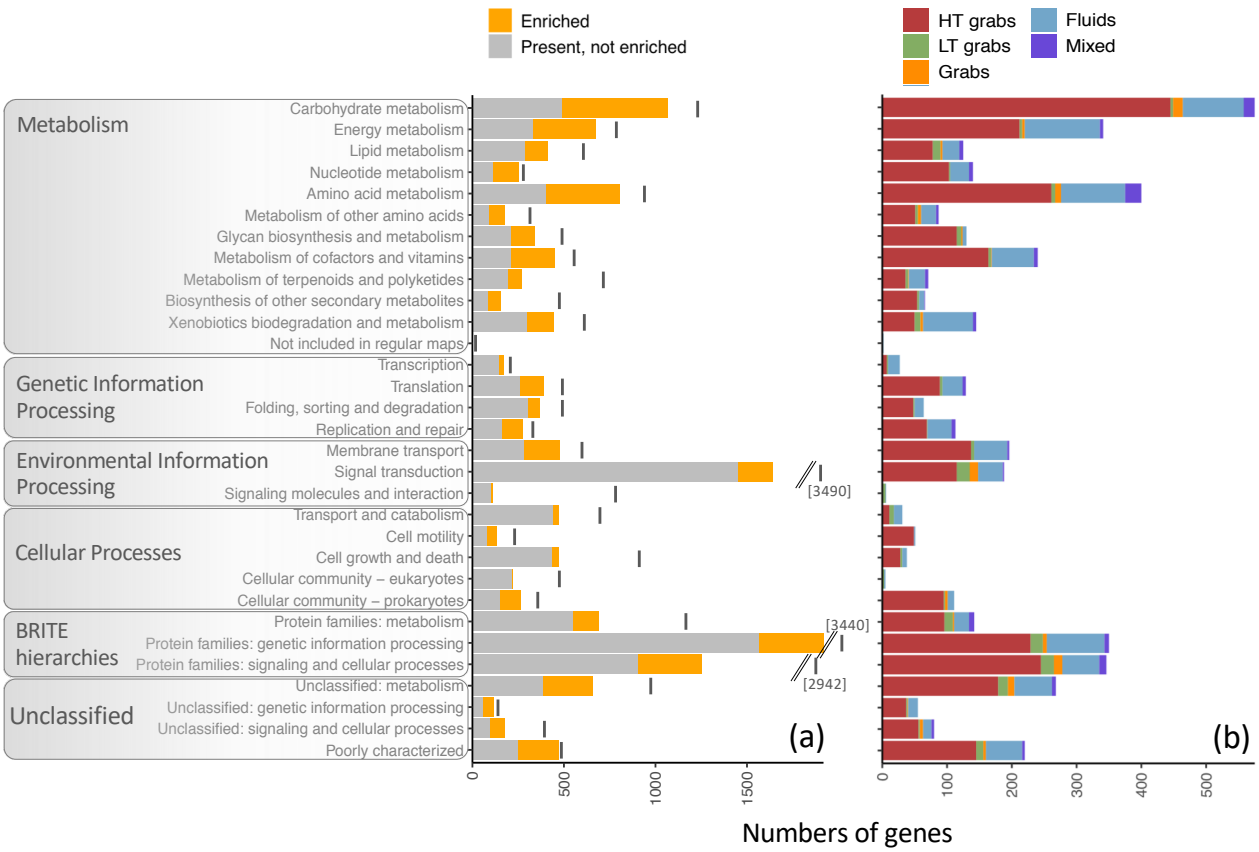


Figure 4.5 Genes present and enriched

Breakdown of genes belonging to KEGG categories and sub-categories that are both present and enriched in the metagenomes. Black lines in a) indicate the total number of genes in the KEGG database for each sub-category. Bars in b) give details of the yellow bars in a).

higher with systematic removal of one sample to even out the numbers of samples in the comparison. For four of the highT samples, removals each yielded variation similar to that of identical runs (5-7%), whereas removals of EMw1 and MVw12 respectively resulted in differences of 17% and 19% in enriched genes. Because the objective of enrichment tests is to find the common elements that robustly vary between sample types across a range of environmental conditions, inclusion of all samples is preferred. Therefore, every enriched gene reported from tests between groups of unequal size has an associated confidence

score— ***high***: enrichment loss tied to 0-1 samples; ***moderate***: enrichment loss tied to 2-3 of the samples; ***low***: enrichment loss tied to 4 or more samples. These scores will be taken into consideration in pathway-level interpretations. For example, a pathway with several enriched genes of high confidence will be considered more reliable than a pathway with several enriched genes of low confidence.

Three pairwise enrichment tests on the clr-transformed sample-by-gene matrix identified 3,440 genes that were enriched in highT or lowT grab samples or diffuse fluids. Of these enriched genes, 63% were enriched in highT grabs, 6% in lowT grabs and 4% in both. The remaining 27% were enriched in diffuse fluids or in a mix of fluids and grabs among the three pairwise tests. In several KEGG metabolism sub-categories, over 50% of the genes present in the metagenomes were enriched (Figure 4.5a), including nucleotide, carbohydrate, energy, and amino acid metabolisms, as well as metabolism of cofactors and vitamins. Most of these enrichments occurred in highT grabs (Figure 4.5b). Several genes in the Unclassified sub-categories were also enriched, leaning towards highT grabs.

4.3.5 Enrichment leanings by pathway

Table 4.2 lists the KEGG pathways that included genes enriched in any one of the three sample types, along with average enrichment scores and enrichment leanings towards each sample type. The total number of genes in each pathway is given for the four major categories (Metabolism, Genetic and Environmental Information Processing, and Cellular Processes), and the percent of genes present gives a measure of community-level pathway completeness. The 3,440 enriched genes were involved in 225 described pathways (many were involved in multiple pathways), 42 BRITE hierarchy protein families, and 23 sub-categories of

Table 4.2 Gene abundances and enrichments by pathway

(First page only. Full table can be found after the Discussion section.)

	KEGG B-LEVEL TERMS	KEGG PATHWAYS	Total gene counts	Total genes in pathway	Percent (No.) genes present	Percent (No.) genes enriched	Average enrichment scores: HTw | LTw	HTw | HTf	LTw | LTf	Enrichment leanings: Diffuse fluids	HT worms	LT worms
METABOLISM	Carbohydrate metabolism	**Citrate cycle (TCA cycle) ***	1304381	62	92% (57)	54% (34)					2.53	1.61
		Amino sugar and nucleotide sugar metabolism	1421712	147	86% (126)	59% (87)				0.26	3.25	
		Galactose metabolism	391228	77	74% (57)	38% (30)				0.53	3.41	
		Pentose and glucuronate interconversions	214225	81	64% (52)	23% (19)				0.18	2.66	
		Starch and sucrose metabolism	501147	105	79% (83)	35% (37)				0.81	2.99	
		Pentose phosphate pathway	723240	84	70% (59)	39% (33)				0.78	2.81	
		Pyruvate metabolism	1459723	110	84% (92)	47% (52)				0.53	2.52	
		Glycolysis / Gluconeogenesis	1402133	106	79% (84)	55% (59)				0.60	2.39	
		Butanoate metabolism	884725	99	84% (83)	47% (47)				0.66	2.30	
		Fructose and mannose metabolism	526093	111	79% (88)	43% (48)				1.86	2.64	
		Glyoxylate and dicarboxylate metabolism	1236293	100	87% (87)	44% (44)				1.57	2.17	
		Propanoate metabolism	865610	110	77% (85)	39% (43)				1.24	1.78	
		C5-Branched dibasic acid metabolism	429585	29	83% (24)	55% (16)				2.22	2.15	
		Ascorbate and aldarate metabolism	125976	54	76% (41)	29% (16)				1.39	1.06	
		Inositol phosphate metabolism	151643	76	67% (51)	17% (13)				2.49	1.15	
	Energy metabolism	**Nitrogen metabolism ***	783538	65	77% (50)	30% (20)					2.68	
		Methane metabolism	1150633	190	83% (158)	37% (71)				0.34	2.52	
		Carbon fixation pathways in prokaryotes	1621315	106	84% (89)	59% (63)				1.28	1.82	
		Oxidative phosphorylation	2002936	218	78% (170)	33% (72)				2.94	2.18	
		Carbon fixation in photosynthetic organisms	722617	36	94% (34)	52% (19)				2.40	1.37	
		Sulfur metabolism	825827	109	78% (85)	35% (39)				2.16	0.90	
		Photosynthesis	449898	63	95% (60)	66% (42)				3.62	0.96	
		Photosynthesis - antenna proteins	5261	42	71% (30)	45% (19)				2.70		
	Lipid metabolism	**Steroid biosynthesis**	6980	34	62% (21)	23% (8)					1.60	2.53
		Secondary bile acid biosynthesis	100314	15	33% (5)	20% (3)					3.69	
		Fatty acid biosynthesis *	739203	41	83% (34)	46% (19)					2.63	
		Fatty acid elongation	676	24	67% (16)	8% (2)					1.15	1.34
		Glycerophospholipid metabolism	556570	111	84% (93)	25% (28)				0.55	2.70	
		Synthesis and degradation of ketone bodies	75587	8	100% (8)	62% (5)				1.78	3.56	
		Primary bile acid biosynthesis	7650	18	33% (6)	11% (2)				0.25	2.01	
		Sphingolipid metabolism	93711	45	69% (31)	20% (9)				2.08	3.07	
		Fatty acid degradation	412492	57	74% (42)	33% (19)				1.21	1.39	
		Glycerolipid metabolism	290189	90	61% (55)	20% (18)				1.82	1.96	
		Ether lipid metabolism	13014	31	77% (24)	6% (2)				2.05	1.41	
		Linoleic acid metabolism	11383	25	40% (10)	8% (2)				2.05	1.41	
		alpha-Linolenic acid metabolism	36481	25	48% (12)	12% (3)				2.05	1.23	
		Steroid hormone biosynthesis	1434	46	28% (13)	8% (4)				1.07		
		Biosynthesis of unsaturated fatty acids	22009	32	59% (19)	6% (2)				2.03		
		Arachidonic acid metabolism	28235	45	47% (21)	8% (4)				2.38	0.05	
	Amino acid metabolism	**Histidine metabolism ***	562354	47	83% (39)	44% (21)					2.98	0.73
		Alanine, aspartate and glutamate metabolism *	1312196	70	84% (59)	48% (34)					2.46	
		Valine, leucine and isoleucine biosynthesis *	636108	19	89% (17)	63% (12)					1.51	
		Lysine biosynthesis	662146	48	92% (44)	54% (26)				1.94	3.12	
		Tyrosine metabolism	237031	81	63% (51)	23% (19)				1.51	2.59	
		Valine, leucine and isoleucine degradation	539806	70	87% (61)	50% (35)				1.14	1.83	
		Cysteine and methionine metabolism	1344755	117	74% (87)	40% (47)				1.55	1.36	
		Phenylalanine metabolism	327237	77	77% (59)	24% (19)				1.76	1.49	
		Tryptophan metabolism	325573	80	56% (45)	15% (12)				0.34	1.18	
		Phenylalanine, tyrosine and tryptophan biosynthesis	942565	71	83% (59)	52% (37)				2.30	1.91	
		Arginine biosynthesis	656669	60	82% (49)	40% (24)				1.92	1.50	
		Lysine degradation	324605	89	67% (60)	24% (22)				2.03	1.46	
		Arginine and proline metabolism	638266	108	77% (83)	40% (44)				1.96	0.85	
		Glycine, serine and threonine metabolism	1284322	100	90% (90)	54% (54)				2.04	0.93	
	Metabolism of other amino acids	**D-Glutamine and D-glutamate metabolism ***	166413	10	60% (6)	40% (4)					3.41	
		D-Arginine and D-ornithine metabolism	14144	9	67% (6)	33% (3)					3.20	
		Cyanoamino acid metabolism	202802	36	39% (14)	11% (4)					2.92	
		Selenocompound metabolism	456423	33	88% (29)	63% (21)				0.52	2.70	
		Phosphonate and phosphinate metabolism	45915	43	51% (22)	13% (6)					2.15	
		Taurine and hypotaurine metabolism	166936	24	96% (23)	54% (13)				0.37	1.14	
		D-Alanine metabolism	116846	6	83% (5)	50% (3)				1.38	1.38	
		beta-Alanine metabolism	222157	44	89% (39)	36% (16)				0.98	1.20	
		Glutathione metabolism	448404	56	61% (34)	21% (12)				2.27	1.33	

Pathways in green: strength of enrichment leanings towards highT and/or lowT grabs is at least one full point above leanings towards diffuse fluids. Pathways in **bold**: enrichments leaned exclusively to worms.
Pathways marked with * were chosen for in-depth taxonomic analysis.

unclassified genes.

Enrichment leanings were used to prioritize pathways for in-depth taxonomic analysis among the three habitat types. Pathways with enrichment leanings strictly towards grabs samples (**bold** text in Table 4.2) were considered to have the greatest potential for describing the distinguishing functional features of grab sample microbes from those in diffuse fluids. Further criteria applied to selection of pathways for in-depth analysis considered the percentages of pathway genes both present and enriched in the metagenomes. Initial in-depth taxonomic analyses described here involved thirteen grab-sample-leaning pathways (indicated with asterisks in Table 4.2) for which a minimum 50% of genes in the pathway were present in the metagenomes and a minimum 30% of genes were enriched. Also included were genes for carbon fixation and protease, lipase, chitinase, and glucosidase genes.

Phase 2 enrichment tests to identify enriched taxa within the selected pathways were simplified from three down to two pairwise tests (i.e., comparing highT and lowT assemblages to all diffuse fluids rather than only those in the corresponding temperature regime) for two reasons. First was that there were far more enrichments in highT grabs versus highT fluids than lowT grabs versus lowT fluids. Secondly, for efficiency—as each ALDEx2 test takes several hours to complete, it made sense to compare all grabs to all fluids in a single test and rely on the grabs versus grabs test to distinguish between highT and lowT.

4.3.6 Relevance of core and non-core community taxa to selected pathways

Core community taxa identified in Chapter 3 generally separated into highT and lowT in Principal Components Analysis (PCA; Figure 4.6) but there was considerable overlap with

the abundant non-core taxa. The selected pathways and genes clearly differentiated into distinct zones linked to the three sample types. Only highT core taxa were closely aligned with pathways in highT grabs, however, some highT core taxa aligned closer to lowT and diffuse fluid pathways. The few lowT core taxa included were not closely aligned with pathways in lowT grabs, with the exception of *Nannocystis* (Deltaproteobacteria). Some of the abundant non-core taxa aligned with pathways in diffuse fluids but many were positioned between fluids and grabs indicating even distribution of genes across both sample types.

Figure 4.6 Principal Components Analysis

PCA of taxon-specific gene abundances in each of the three sample types. Arrows denote pathways. Circles denote the included core and abundant non-core taxa. Colored wedges were added manually to highlight the three habitat types.

For a deeper exploration of the links between specific taxa and pathways, the PCA was divided into 6 regions. Taxonomic clusters from the hierarchical clustering step of heatmap creation were linked to the positions of the taxa in the cluster to the 6 regions of the PCA. Taxa formed three major clusters (labeled Tax1, Tax2 and Tax3; Figure 4.7), driven by the abundance of genes in the three sample types. Tax1 was a cluster of non-core taxa from PCA regions 5 and 6 with high abundance of pathway genes in diffuse fluid samples. Cluster Tax2 contained mostly highT core taxa belonging to Bacteroidetes and Deltaproteobacteria from PCA regions 2-4 that had elevated gene abundances across highT and lowT grabs. Cluster Tax3 contained taxa that were broadly distributed across all regions of the PCA except region 6. Within Tax3, lowT core taxa from PCA region 5 were distinct from other taxa, and the rest of Tax3 was a mix of highT core and abundant non-core taxa. Taxa from PCA region 1 formed two distinct sub-clusters of Tax3, both with elevated gene abundances in highT grabs and diffuse fluids but one contained only core taxa and the other had a mix of core and non-core. Taxa in the latter sub-cluster included abundant members of hydrothermal vent microbial communities (e.g., *Sulfurimonas*) as well as two related highT core community Epsilonbacteraeota (*Arcobacter* and *Hydrogenomonas*).

Overlaying the results of the phase two enrichment tests (pathway specific genes by sample) onto the hierarchical clusters of taxa and functional pathways shows the strength of the various features in defining taxonomic clusters (Figure 4.8). For example, *Sulfurimonas*, which is known to be a dominant organism in many hydrothermal vent habitats, is functionally very similar to highT core *Arcobacter* in the selected pathways, and both are highly enriched in the associated genes. But they differ in enrichment of genes for fatty acid biosynthesis, O-antigen nucleotide sugar biosynthesis, and some indicators of heterotrophy

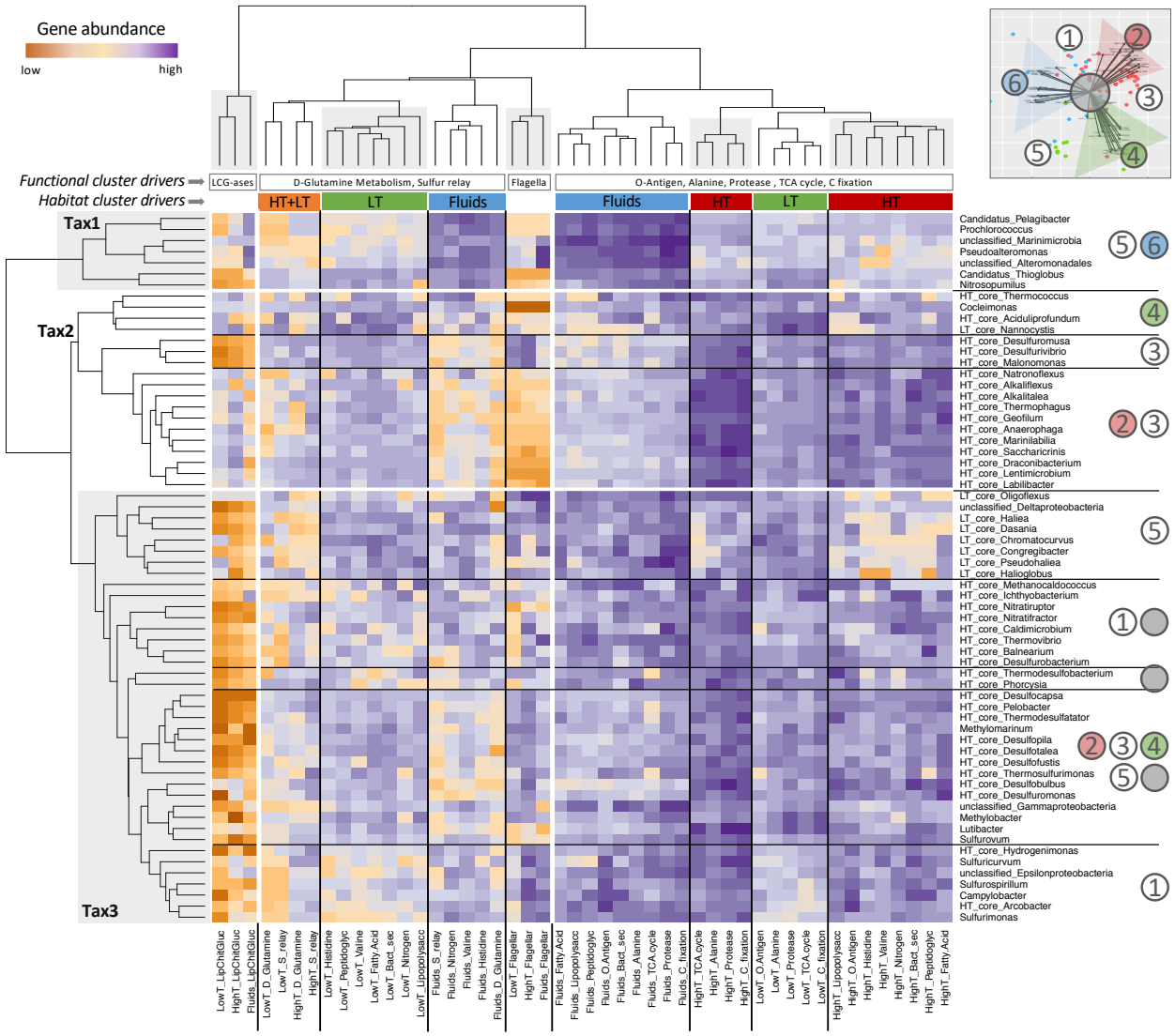


Figure 4.7 Gene abundance heatmap

Centered log-ratio transformed gene abundances involved in functional pathways in each of three sample types. Taxa listed on the vertical axis are core community (HT_ or LT_core) and abundant non-core community taxa. Numbered circles next to taxa reference the 6 regions of the PCA plot (Figure 4.6 and small inset) and indicate where that cluster of taxa was distributed in the PCA plot. Grey circles with no number indicate taxa that fell close to center of the PCA plot. Along the top, between the heatmap and dendrogram, are two rows that indicate either functional or habitat/sample type drivers of clusters in the top dendrogram. Sample types: diffuse fluids=Fluids; highT grabs=HT or HighT; lowT grabs=LT or LowT. Pathway and gene abbreviations: Alanine=alanine, aspartate and glutamate metabolism; Bact sec=bacterial secretion system; C fixation=carbon fixation; D-glutamine=D-glutamine and D-glutamate metabolism; Fatty Acid=fatty acid biosynthesis; Flagellar=flagellar assembly; Histidine=histidine metabolism; LipChitGluc=lipases, chitinases, glucosidases; Lipopolysacc=lipopolysaccharide biosynthesis; Nitrogen=nitrogen metabolism; O-Antigen=O-antigen nucleotide sugar biosynthesis; Peptidoglyc=peptidoglycan biosynthesis; Protease=proteases; S relay=sulfur relay system; TCA cycle=tricarboxylic acid cycle; Valine=valine, leucine and isoleucine biosynthesis.

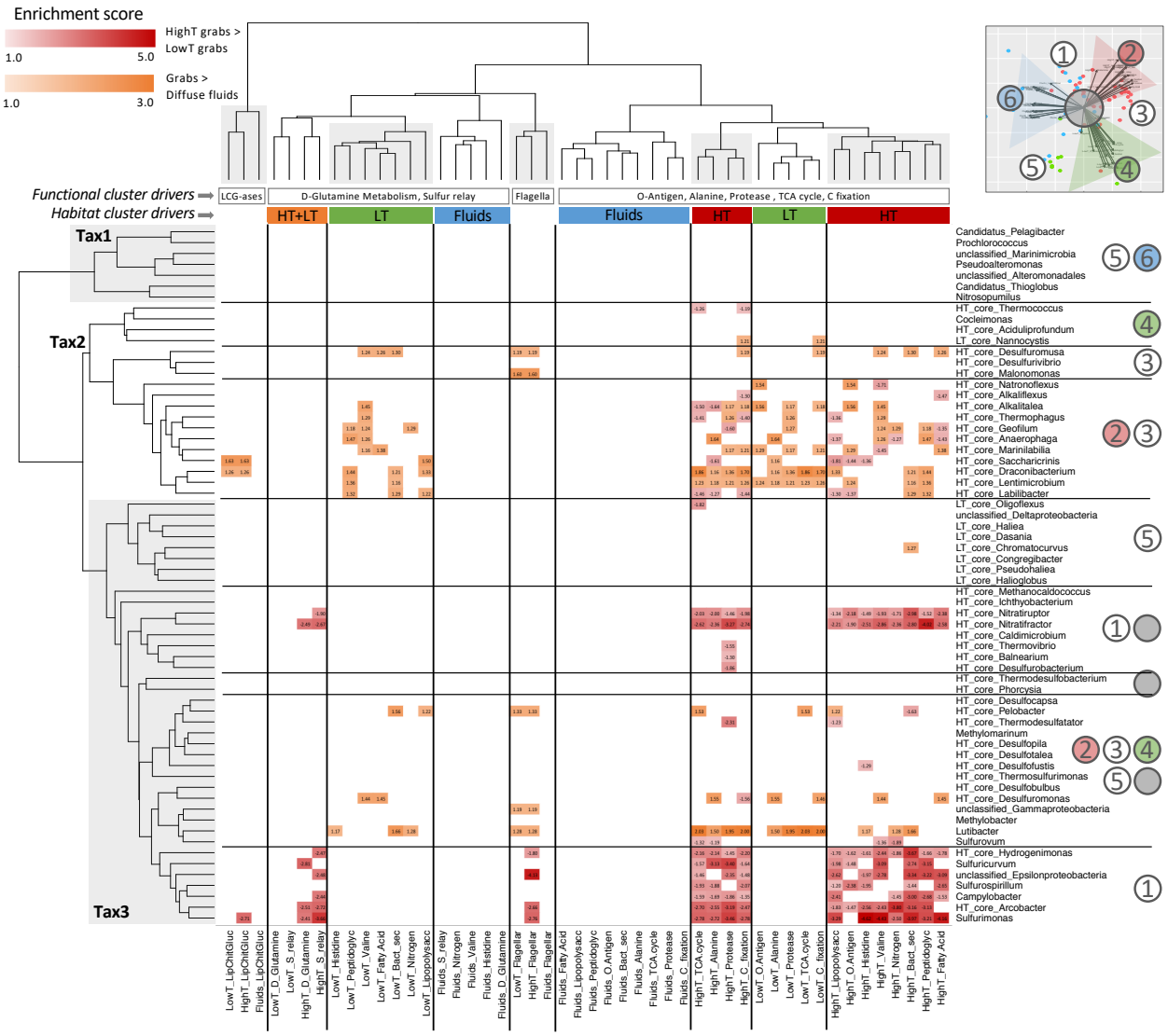


Figure 4.8 Taxonomic enrichments heatmap

Same as Figure 4.7 but with heatmap representing enrichment scores from ALDEx2 tests on pathway-specific genes. Only genes enriched in tubeworm habitats are shown. Enrichment scores in red hues are enrichments in highT grabs relative to lowT grabs while scores in orange hues are enrichments in grab samples relative to diffuse fluids.

in the form of lipase, chitinase, and glucosidase genes. In the other Tax3 sub-cluster from PCA region 1, enriched genes belong to highT core *Nitratifractor* and *Nitratiruptor* but differ from the *Sulfuimonas*/*Arcobacter* enrichments by the absence of enriched flagellar assembly genes. Taxa in the Tax2 cluster have similar gene abundances across highT and lowT grabs and are primarily characterized by enrichments that distinguish grab samples from fluid samples. HighT core *Draconibacterium* and *Lentimicrobium* have the greatest enrichments in this cluster and are distinguished from highT core taxa in Tax3 by the absence of enrichments in metabolisms of some amino acids and nitrogen. These examples demonstrate the usefulness of combined visualizations in identifying functional differences between core community taxa.

4.3.7 Community metrics

Genes involved in functional modules within pathways for carbon fixation and nitrogen and sulfur metabolism did not cluster strictly along sample type or temperature categories (Figure 4.9). In carbon fixation modules, highT grabs were primarily dominated by genes involved in the Reductive acetyl-CoA pathway, although samples EMw1 and MVw12 clustered apart from the rest and MVw12 had more genes for the 3-hydroxypropionate bi-cycle. The fluid samples associated with these two grabs (EMf1 and MVf12) were the only two fluids with high abundance of genes linked to the reductive tricarboxylic acid (rTCA) cycle, while others were dominated by genes involved in the reductive pentose phosphate pathway (Calvin-Bensen-Bassham, CBB) and 3-hydroxypropionate bi-cycle.

Among nitrogen (N) cycling modules, samples from Middle Valley had the greatest abundances of N cycling genes, specifically for nitrification and denitrification in both the

highT grab and associated fluids, and N fixation and nitrification in the lowT grab. Samples from Clam Bed also had an abundance of genes for N fixation and nitrification. Grab sample EMw1 had moderate abundance of genes involved in all N cycling processes except N-fixation. Sulfur cycling genes were generally most abundant in fluids with a mix of dissimilatory sulfate reduction and thiosulfate oxidation, except fluid sample EMf2 and EMf3 that were dominated by genes for assimilatory sulfate reduction.

The balance of autotrophy and heterotrophy genes was fairly even in diffuse fluid and highT grab samples, with a near one-to-one ratio (Figure 4.10). The balance leaned more towards heterotrophy genes in lowT grab samples, but the difference was not significant (Mann-Whitney test, $p=0.1$).

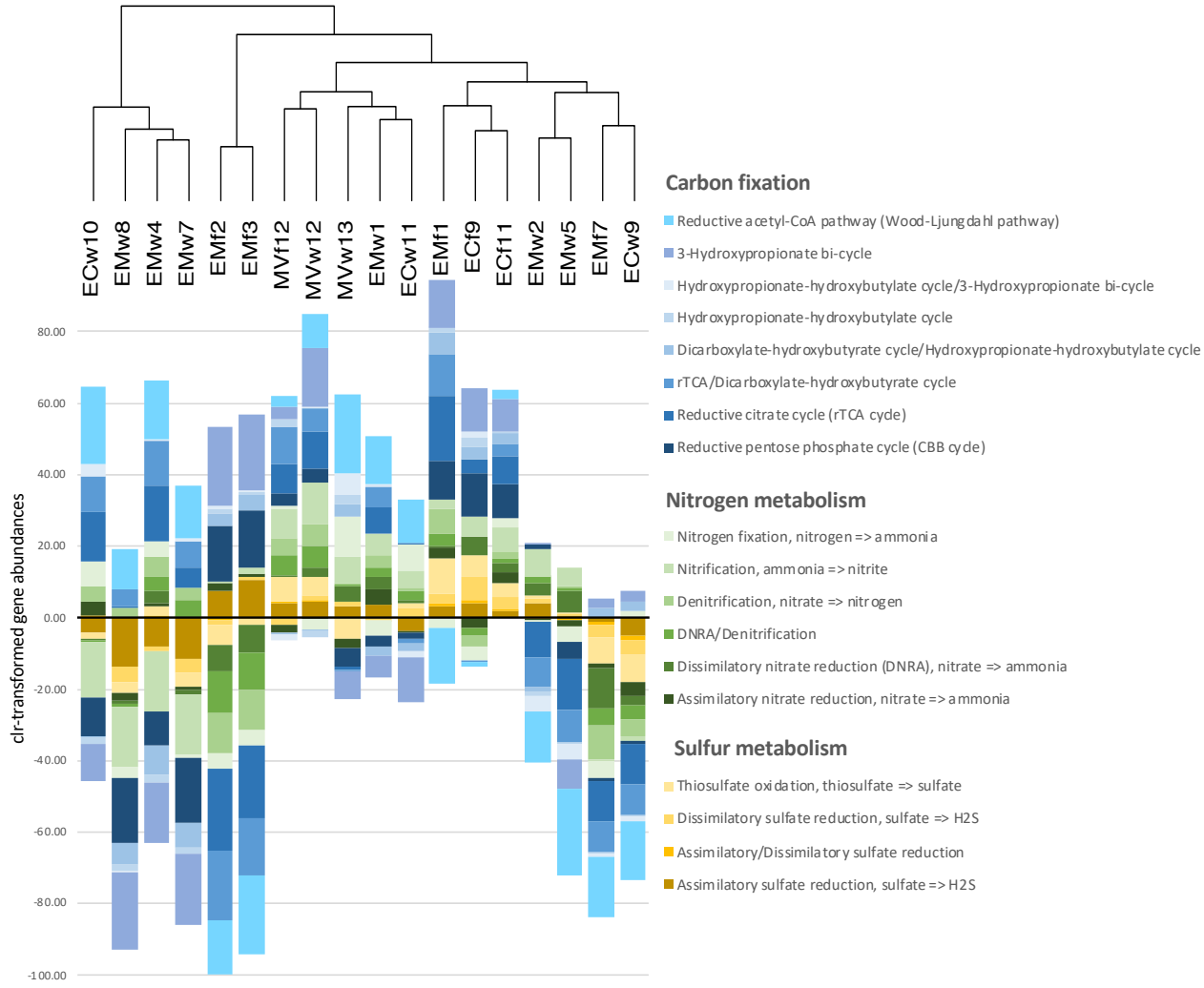


Figure 4.9 Functional metrics of C, N, and S-cycling processes

The balance of centered log-ratio transformed gene abundances for functional modules within pathways for carbon fixation and nitrogen and sulfur metabolism. Hierarchical clustering shown in the dendrogram is calculated based on the genes involved in these processes. Bars above the zero line denote genes with above average abundance in the sample, and below the line have below average abundances. For genes shared by two modules, both are indicated.

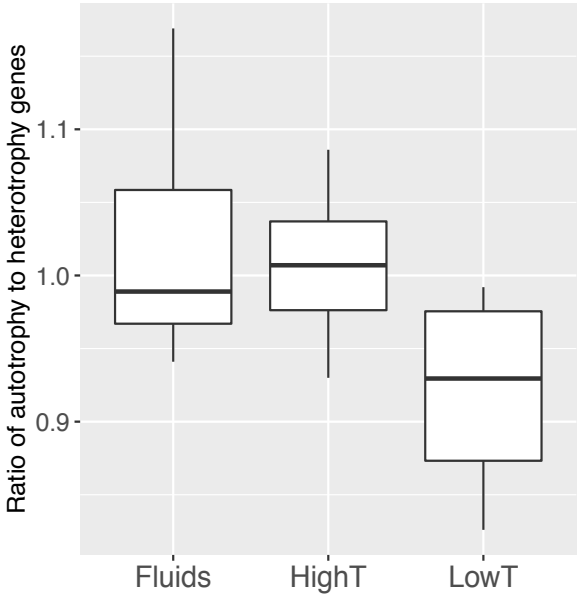


Figure 4.10 Balance of autotrophy-heterotrophy

The ratio of autotrophy to heterotrophy genes normalized to the total of genes with an identified product. Ratios >1 have more autotrophy genes. Ratios <1 have more heterotrophy genes. Sample ECw9 was excluded from the LowT group as it had a ratio several standard deviations above the mean for all the samples. Mann-Whitney tests indicated no significant differences between any of the three sample types.

4.4 Discussion

Analysis of metagenomic sequence data offers tremendous potential insight into the functional capacities of microbial communities while also bringing unique challenges in synthesizing copious amounts of genetic information into meaningful community functional descriptors. Although there are many examples of comparative metagenomic studies, most offered limited guidance for two reasons: 1) Studies almost universally report relative abundances of normalized data, but the approaches to normalization vary widely and can affect the end results of a study (Pereira et al., 2018), making comparisons between studies

difficult. 2) The recent shift in the handling of data from high-throughput sequencing to reflect the compositional nature of these data further complicates these comparisons and even calls into question earlier findings that used inappropriate data handling (Gloor and Reid, 2016). As compositional data handling of metagenomes has not been widely adopted, the comparative analysis of diffuse fluid and tubeworm-associated metagenomes required a considerable amount of method development.

4.4.1 Challenges and method development

One challenge was the frequent involvement of genes in multiple processes. Enriched genes were involved in as many as 15 different pathways, complicating gene-level descriptions of enrichment. Gene-level metagenome studies from hydrothermal vents have relied on relative abundances of "key genes" for carbon, nitrogen, sulfur, hydrogen and methane cycling as evidence of potential (Reveillaud et al., 2016; Galambos et al., 2019), but offered no defining criteria justifying their use as proxies. Whether the presence of one or two genes in a pathway sufficiently represents functional potential is unclear, although later work combining stable isotope incubations and transcriptomics has provided practical evidence in support of the practice within those particular cycles (Fortunato and Huber, 2016). However, in an exploratory phase where the goal is to identify defining functional features distinguishing habitats, it was more practical to work at the pathway level.

In prioritization and selection of pathways for in-depth analysis, the completion of pathways (percent of genes present in the metagenomes) had to be considered in order to address the broad distribution of genes across pathways. Frequently, in pathways for which only a few genes were present, enriched genes were those widely involved in many

processes. By targeting pathways with at least 50% of genes present, the likelihood of pathway relevance was higher, albeit still only suggestive of community functional potential. More definitive assessments of functional potential will come from species-level investigations, which can be accomplished through generation of MAGs and calculations of pathway completion scores (Galambos et al., 2019). Generation of MAGs is currently underway.

Another challenge with these metagenomic data stems from the considerable taxonomic novelty in these vent samples, which was brought to light in Chapter 3. The lack of close relatives in protein databases likely contributed to the low percentages of genes with an identified product (avg. 59%). Furthermore, the anomalously low percentage of genes identified in sample ECw9 (34%), which also had a high abundance of Archaea, underscores how far the field of proteomics lags behind for this domain. Representation of nearly all KEGG database genes in the Poorly Characterized category and a high percentage of genes characterized only by BRITE hierarchies in the metagenomes suggests considerable potential for functional novelty in these samples and highlights the need for continued culturing efforts to characterize novel hydrothermal vent microbes.

The development of pathway leanings was essential to the exploratory phase of metagenomic analysis and offered an objective prioritization of pathway selection for in-depth taxonomic investigations. Leanings also provide opportunities for general observations and development of hypotheses about dynamics within the sampled habitats. For example, leanings towards highT grabs were far more numerous than leanings towards lowT grabs or fluids, which may be related to aspects of taxonomic composition in highT habitats discussed in Chapter 3. HighT grabs had lower taxonomic diversity than lowT grabs

and the covariance network of microbes and fauna (Figure 3.3) showed tighter associations between taxa in highT grabs than in lowT grabs. It is possible that enrichments are a more useful tool for finding distinguishing features in low diversity habitats where environmental selective pressure is high. Alternatively, more enrichments in highT grabs could be evidence of increased horizontal gene transfer (HGT) reflecting useful genes passed among a few taxa. Evidence of abundant transposes in hydrothermal chimney metagenomes (Brazelton and Baross, 2009; Xie et al., 2011) and of potential co-evolution of bacteria and viruses through gene sharing from diffuse fluid metagenomes (Anderson et al., 2014) suggest that HGT may be common in hydrothermal systems. Fortunately, tools to identify evidence of HGT in metagenomes are becoming available (Douglas and Langille, 2019; Song et al., 2019) with the growing awareness of the potential impact of HGT on functional inferences.

An unfortunate outcome of the metagenomic analysis was the low recovery of genes from microeukaryotes. While this is frequently the case in hydrothermal vent metagenomes (C. Fortunato, personal communication), it highlights the need for modified metagenomic approaches to gain genomic information on vent microeukaryotes. Recent tools developed to identify, annotate, and bin eukaryotic reads from complex metagenomic data (West et al., 2018; Humann et al., 2019) offer promise for further analyses of these data. Eukaryotic reads have been identified from all metagenomes and await annotation.

4.4.2 Community-level functional potential

Although the results presented here describe only a small fraction of the metagenomic data, they demonstrate a path towards characterization of microbial functional potential in tubeworm bushes and other vent faunal aggregations that can serve

as a baseline for future 'omics research. The 13 pathways selected for in-depth taxonomic analysis, which make up less than 6% of the characterized functional pathways with enriched genes, were chosen for their potential to describe unique functional attributes of tubeworm assemblage microbes. Separation of the PCA pathway loadings into three zones, with PC1 distinguishing worms and fluids, indicates that enrichment scores and leanings appear to be a useful tool for that selection process. The combined approach to visualizations that linked PCA with taxonomic and functional clustering in heatmaps and enrichment scores provided insight into the relative importance of both core community and abundant non-core taxa to the selected pathways. Inclusion of genes for autotrophy (carbon fixation) and heterotrophy (protease, lipase, chitinase and glucosidase) formed a bridge between this exploratory visual approach and the targeted functional module analysis of community metrics.

Combined visualizations revealed enrichments in members of the Tax2 cluster that generally distinguished grabs from fluids while enrichments in members of the Tax3 cluster distinguished highT from lowT grabs. While Tax2 contained mostly highT core taxa, cluster Tax3 included most of the abundant non-core taxa along with highT and lowT core taxa. Enrichments in lowT grabs were few and offered little information on distinguishing functional features of this habitat, likely due to the elevated taxonomic diversity and smaller number of core community taxa. In contrast, enrichment patterns in clusters Tax2 and Tax3 revealed functional differences among highT core taxa. For example, enrichments within the Tax3 cluster show the primary difference between the highT core taxa *Arcobacter*, *Hydrogenimonas*, *Nitratrifractor* and *Nitratiruptor* lies in the potential for motility, with *Arcobacter* and *Hydrogenimonas* enriched in genes for flagellar assembly. *Arcobacter* clusters

tightly with *Sulfurimonas*, a frequently abundant yet non-core community member of vent microbial assemblages, but the two differ in enrichments linked to outer membrane biogenesis (biosynthesis of fatty acids for phospholipids and O-antigen nucleotide sugars for lipopolysaccharides) suggesting different strategies related to cell wall construction between the abundant and core community taxa. Further distinguishing functional features among highT core taxa involve the Tax2 cluster members *Draconibacterium* and *Lentimicrobium,* which have numerous enrichments in grabs relative to diffuse fluids—they are functionally distinguished from highT core taxa in the Tax3 cluster by the absence of enrichments tied to amino acid and nitrogen metabolisms. This Tax2 cluster contains Bacteroidetes that are a mix of highT core taxa and close relatives of *Carboxylicivirga*, which had the strongest association with highT grabs of all core community taxa in Chapter 3 but could not be included in comparisons here as it does not have its genome sequenced.

It is worth mentioning the bias that may have come from selecting pathways for their potential to describe unique functional attributes of grab samples (i.e., that only had leanings towards grab samples). There may also be unique functional process differences between sample types in pathways with enrichment leanings towards both fluids and grabs, but these may only be apparent at a finer scale (i.e., between functional modules within pathways). Identifying these distinctions will rely more heavily on the community metrics approach discussed below. Microbial community metrics may also provide a useful tool for understanding the normal range of ecosystem states and recognizing excursions from normal. Such ecosystem state changes, especially when they involve core community organisms, could help explain observed shifts in faunal assemblages or act as indicators of environmental stress.

The energy metabolism pathway *Carbon fixation pathways in prokaryotes,* which had enrichment leanings towards both highT grabs and diffuse fluids in Table 4.2, includes six major carbon fixation modules. The community metrics approach revealed variations among the abundances of genes in those six modules, although not strictly aligned with sample types. Interestingly, the rTCA and CBB cycles, typically reported as having the highest abundance of genes for carbon fixation in vent metagenomic studies (Xie et al., 2011; Reveillaud et al., 2016), were frequently outnumbered by abundant genes for the reductive Acetyl-CoA cycle and 3-hydroxypropionate bi-cycle in grab and fluid metagenomes. This suggests a more diverse suite of primary producers associated with faunal assemblages. Alternatively, this may accurately reflect the balance of carbon fixation cycle genes in hydrothermal systems, and challenge previous studies that did not treat data as compositional.

Community metrics related to nitrogen and sulfur cycling were also not strictly variable along temperature or sample type delineations. Nitrogen cycling genes had the greatest representation in samples from Middle Valley and Clam Bed suggesting more active cycling of nitrogen at those sites. Genes for nitrification, rarely reported in vent metagenome studies, had elevated abundance in Middle Valley samples, which is consistent with the trend of increasing ammonium concentrations from south to north along the Juan de Fuca ridge (Bourbonnais et al., 2012). Sulfur cycling genes mirrored previous metagenome studies with elevated abundance in fluids (Fortunato and Huber, 2016; Reveillaud et al., 2016), but appear less important in faunal assemblages.

The community metric indicating the autotrophy-heterotrophy balance was not particularly informative as the ratio did not deviate far from one-to-one and was not

significantly different among the three sample types. The ratio did lean more towards heterotrophy in lowT grabs, albeit insignificantly so, as was expected by the shift to more heterotrophic taxa demonstrated in Chapter 3. This indicates that even a weak shift in the auto-heterotrophy balance can be indicative of compositional shifts between sample types. Interestingly, similarity in the ratios in diffuse fluids and highT grabs suggests consistent potential for autotrophy between the two, but the variation in process modules discussed above suggests there is some functional redundancy that may ensure consistent biomass production throughout changing environmental conditions. Future work incorporating targeted incubations and metatranscriptomics may reveal clearer sample type distinctions in expression levels of genes in carbon, nitrogen and sulfur cycling genes.

4.4.3 Conclusion

Although the results discussed here represent only a small fraction of the information available in the metagenomes, there is enormous potential in the combined approaches described. Many of the analyses presented were exploratory and experimental but performed reasonably well in identifying functional potential that distinguished between highT and lowT grabs as well as between faunal grab samples and diffuse fluids. The combined visuals of Principal Components Analysis and clustered heatmaps based on gene abundances revealed common functional potential within groups of abundant non-core and core community taxa among the three sample types. Overlaying enrichment scores added additional information on the relative potential of taxa across the different pathways analyzed. The clear distinction of pathway loadings in the Principal Components Analysis

supports the utility of enrichment scores and enrichment leanings for prioritizing the potential importance of pathways to distinguishing particular habitat types.

Information provided by community-level functional metrics was somewhat limited but may be more informative when combined with levels of gene expression. Combining metagenomic data with transcriptomics and targeted incubations offers great promise for delving deeper into the inferences of functional potential. Application of these additional methods is greatly needed to characterize the functional contributions of these long overlooked microbial habitats, but continuing analysis of these metagenomic data will form the baseline characterization from which to move forward.

Table 4.3 (continued) Gene abundances and enrichments by pathway

	KEGG B-LEVEL TERMS	KEGG PATHWAYS	Total gene counts	Total genes in pathway	Percent (No.) genes present	Percent (No.) genes enriched	Average enrichment scores: HTw | LTw	HTw | HTf	LTw | LTf	Enrichment leanings: Diffuse fluids	HT worms	LT worms
METABOLISM (cont.)	Nucleotide metabolism	Pyrimidine metabolism	1294766	100	83% (83)	53% (53)				2.10	2.56	
		Purine metabolism	2098591	210	80% (169)	41% (87)				1.42	1.52	
	Glycan biosynthesis and metabolism	**Glycosphingolipid biosynthesis - globo and isoglobo series**	45107	15	47% (7)	26% (4)					3.17	2.10
		Lipopolysaccharide biosynthesis *	800320	59	68% (40)	50% (30)					3.24	1.56
		Glycosphingolipid biosynthesis - ganglio series	29810	14	36% (5)	7% (1)					4.01	
		O-Antigen nucleotide sugar biosynthesis *	863252	96	53% (51)	42% (41)					3.31	
		Peptidoglycan biosynthesis *	873721	51	63% (32)	35% (18)					3.12	
		Glycosylphosphatidylinositol (GPI)-anchor biosynthesis	641	26	81% (21)	3% (1)					2.83	
		Various types of N-glycan biosynthesis	60287	52	60% (31)	11% (6)					2.56	
		Glycosphingolipid biosynthesis - lacto and neolacto series	757	27	37% (10)	3% (1)						2.10
		Glycosaminoglycan biosynthesis - chondroitin sulfate / dermatan sulfate	605	18	72% (13)	5% (1)					1.76	
		Other glycan degradation	129776	22	77% (17)	50% (11)				2.08	3.80	
		Glycosaminoglycan biosynthesis - heparan sulfate / heparin	189	22	64% (14)	4% (1)						1.47
		Glycosaminoglycan degradation	91392	15	87% (13)	40% (6)				1.52	2.79	
		Arabinogalactan biosynthesis - Mycobacterium	18712	12	50% (6)	16% (2)				1.47	1.19	
		N-Glycan biosynthesis	60666	45	80% (36)	13% (6)				2.01	1.03	
		Lipoarabinomannan (LAM) biosynthesis	5822	14	57% (8)	14% (2)				1.43	1.28	
	Metabolism of cofactors and vitamins	**Lipoic acid metabolism**	62208	11	36% (4)	9% (1)					3.65	
		Thiamine metabolism	473894	36	69% (25)	36% (13)					3.28	
		Retinol metabolism	34870	48	33% (16)	10% (5)					2.82	
		Ubiquinone and other terpenoid-quinone biosynthesis	430553	59	83% (49)	37% (22)					2.66	
		One carbon pool by folate	675500	37	81% (30)	45% (17)				1.99	2.20	
		Biotin metabolism	516809	23	91% (21)	65% (15)				2.90	3.01	
		Folate biosynthesis	852931	82	73% (60)	36% (30)				1.44	1.30	
		Nicotinate and nicotinamide metabolism	575773	97	65% (63)	29% (29)				2.08	1.36	
		Porphyrin and chlorophyll metabolism	1030581	129	75% (97)	48% (62)				2.27	1.44	
		Pantothenate and CoA biosynthesis	753035	44	91% (40)	52% (23)				2.01	1.33	
		Riboflavin metabolism	270992	49	57% (28)	30% (15)				1.79	1.08	
		Vitamin B6 metabolism	228106	24	75% (18)	25% (6)				3.42	1.22	
	Metabolism of terpenoids and polyketides	**Biosynthesis of enediyne antibiotics**	461	74	5% (4)	1% (1)						4.28
		Sesquiterpenoid and triterpenoid biosynthesis	9846	73	10% (7)	4% (3)					1.85	2.28
		Zeatin biosynthesis	55630	10	20% (2)	10% (1)					1.58	
		Nonribosomal peptide structures	12380	54	61% (33)	2% (1)					1.49	
		Terpenoid backbone biosynthesis	591055	58	76% (44)	41% (24)				1.54	2.72	
		Type I polyketide structures	1573	115	23% (26)	1% (1)						1.11
		Insect hormone biosynthesis	48190	18	22% (4)	6% (1)				1.65	1.22	
		Polyketide sugar unit biosynthesis	164302	59	34% (20)	10% (6)				2.01	1.32	
		Carotenoid biosynthesis	26530	49	59% (29)	20% (10)				1.84	1.23	
		Biosynthesis of type II polyketide products	1059	72	22% (16)	1% (1)				1.90		
		Biosynthesis of vancomycin group antibiotics	52511	29	48% (14)	10% (3)				1.81	1.45	
		Biosynthesis of siderophore group nonribosomal peptides	19550	38	53% (20)	13% (5)				3.24	1.83	
		Limonene and pinene degradation	75497	12	50% (6)	33% (4)				1.78	1.22	
		Biosynthesis of ansamycins	98786	32	41% (13)	6% (2)				1.72	1.31	
		Geraniol degradation	75395	16	94% (15)	63% (10)				2.32	0.11	
	Biosynthesis of other secondary metabolites	**Biosynthesis of various secondary metabolites - part 2**	75690	55	16% (9)	5% (3)					3.27	
		Novobiocin biosynthesis	170644	30	37% (11)	17% (5)					3.01	
		Neomycin, kanamycin and gentamicin biosynthesis	31621	58	5% (3)	5% (3)					2.78	
		Isoquinoline alkaloid biosynthesis	77468	62	23% (14)	8% (5)					2.39	
		Acarbose and validamycin biosynthesis	94112	27	7% (2)	7% (2)					1.38	
		Prodigiosin biosynthesis	184746	20	15% (3)	15% (3)					1.28	
		Glucosinolate biosynthesis	132841	21	10% (2)	10% (2)					1.26	
		Carbapenem biosynthesis	96973	19	26% (5)	11% (2)					1.25	
		Flavone and flavonol biosynthesis	1121	15	7% (1)	7% (1)					1.15	
		Phenazine biosynthesis	104158	16	38% (6)	13% (2)					1.13	
		Monobactam biosynthesis	300831	28	64% (18)	43% (12)				1.44	1.99	
		Tropane, piperidine and pyridine alkaloid biosynthesis	141370	26	65% (17)	23% (6)				1.83	2.18	
		Streptomycin biosynthesis	316642	20	65% (13)	50% (10)				2.20	2.31	
		Phenylpropanoid biosynthesis	91697	38	32% (12)	13% (5)				1.78	1.22	
		Penicillin and cephalosporin biosynthesis	26837	16	56% (9)	19% (3)				1.59		
		Betalain biosynthesis	5650	7	57% (4)	29% (2)				3.58		1.09

	KEGG B-LEVEL TERMS	KEGG PATHWAYS	Total gene counts	Total genes in pathway	Percent (No.) genes present	Percent (No.) genes enriched	Average enrichment scores			Enrichment leanings		
							HTw | LTw	HTw | HTf	LTw | LTf	Diffuse fluids	HT worms	LT worms
METABOLISM (cont.)	Xenobiotics biodegradation and metabolism	**Nitrotoluene degradation**	170075	22	82% (18)	45% (10)					3.00	
		Drug metabolism - other enzymes	325405	32	78% (25)	50% (16)				0.44	2.28	
		Benzoate degradation	235569	106	81% (86)	30% (32)				0.45	1.57	
		Ethylbenzene degradation	27615	16	81% (13)	6% (1)					1.04	
		Chloroalkane and chloroalkene degradation	104120	43	44% (19)	23% (10)				0.63	1.40	
		Styrene degradation	47754	25	64% (16)	12% (3)					0.75	
		Naphthalene degradation	39425	29	69% (20)	17% (5)				1.74	1.65	
		Drug metabolism - cytochrome P450	63012	25	40% (10)	16% (4)				0.80	1.32	
		Metabolism of xenobiotics by cytochrome P450	61668	34	38% (13)	15% (5)				1.08	1.32	
		Toluene degradation	23825	46	65% (30)	13% (6)				2.09		0.75
		Aminobenzoate degradation	102686	81	62% (50)	17% (14)				2.33	1.31	
		Dioxin degradation	25962	31	45% (14)	10% (3)				3.67	2.06	
		Caprolactam degradation	47678	22	73% (16)	36% (8)				2.02		0.43
		Steroid degradation	10260	15	93% (14)	60% (9)				1.27		
		Xylene degradation	29122	34	76% (26)	12% (4)				3.85	2.06	
		Polycyclic aromatic hydrocarbon degradation	9253	43	42% (18)	2% (1)				1.86		
		Chlorocyclohexane and chlorobenzene degradation	38578	41	63% (26)	15% (6)				1.24		
		Atrazine degradation	10978	16	56% (9)	25% (4)				1.90		
		Furfural degradation	1815	8	100% (8)	63% (5)				2.14		
		Fluorobenzoate degradation	12916	18	61% (11)	17% (3)				1.97		
	Not in regular maps	Global maps only	19208	7	57% (4)	29% (2)				3.86	1.47	
GENETIC INFORMATION PROCESSING	Transcription	Spliceosome	6010	123	83% (102)	4% (5)				0.32		2.78
		Basal transcription factors	14040	36	78% (28)	6% (2)					0.47	
		RNA polymerase	304888	51	86% (44)	37% (19)				1.65	1.37	
	Translation	Ribosome	711	143	94% (134)	49% (70)				1.71	1.26	
		mRNA surveillance pathway	7230	63	83% (52)	6% (4)				1.88		1.16
		Aminoacyl-tRNA biosynthesis	1763782	66	61% (40)	55% (36)				4.38	1.97	
		RNA transport	97509	142	73% (103)	6% (9)				1.65	0.05	
		Ribosome biogenesis in eukaryotes	78782	82	73% (60)	11% (9)				1.97		0.40
	Folding, sorting and degradation	**Sulfur relay system ***	469372	29	72% (21)	38% (11)					3.15	
		Protein export	795202	39	82% (32)	46% (18)				0.42	1.31	
		RNA degradation	904771	79	84% (66)	29% (23)				1.63	1.42	
		Protein processing in endoplasmic reticulum	167185	144	66% (95)	4% (6)				1.44	1.64	
		Ubiquitin mediated proteolysis	4559	131	74% (97)	2% (3)				2.78		1.75
		Proteasome	12264	49	82% (40)	6% (3)				1.61		
	Replication and repair	Homologous recombination	1168246	72	78% (56)	38% (27)				1.86	1.26	
		Mismatch repair	1006450	44	91% (40)	52% (23)				1.85	1.35	
		DNA replication	914704	60	90% (54)	43% (26)				1.69	1.18	
		Nucleotide excision repair	667483	47	91% (43)	26% (12)				1.85	1.03	
		Non-homologous end-joining	3659	19	68% (13)	11% (2)				1.80		
		Base excision repair	522903	43	84% (36)	35% (15)				1.85	1.45	
		Fanconi anemia pathway	9485	54	57% (31)	9% (5)				1.49	0.20	
ENVIRONMENTAL INFORMATION PROCESSING	Signal transduction	**Phospholipase D signaling pathway**	27500	108	41% (44)	3% (3)					3.61	
		Sphingolipid signaling pathway	1792	80	54% (43)	3% (2)						3.49
		Calcium signaling pathway	2790	158	39% (62)	2% (3)					1.45	1.63
		cAMP signaling pathway	2301	163	44% (72)	3% (5)					1.45	1.58
		Phosphatidylinositol signaling system	122017	55	75% (41)	7% (4)					1.46	1.54
		AMPK signaling pathway	96105	84	63% (53)	8% (7)					2.87	0.00
		cGMP-PKG signaling pathway	2105	121	43% (52)	3% (4)						2.87
		FoxO signaling pathway	26898	102	46% (47)	5% (5)					1.66	1.11
		MAPK signaling pathway - plant	108513	54	20% (11)	4% (2)					1.59	1.15
		Two-component system	280834	495	72% (357)	29% (144)				0.11	2.63	
		MAPK signaling pathway	1019	235	31% (73)	0% (1)						1.74
		MAPK signaling pathway - yeast	7249	91	31% (28)	1% (1)						1.71
		Ras signaling pathway	1782	180	33% (60)	1% (1)						1.71
		TGF-beta signaling pathway	644	83	34% (28)	1% (1)						1.71
		Apelin signaling pathway	3531	104	43% (45)	3% (3)				0.11		1.68
		Wnt signaling pathway	1272	119	53% (63)	3% (3)						1.55
		Hippo signaling pathway	940	115	50% (58)	1% (1)						1.42
		PI3K-Akt signaling pathway	74489	263	27% (71)	3% (8)				0.16	1.43	
		mTOR signaling pathway	3014	117	56% (65)	3% (4)				1.67		2.90
		Plant hormone signal transduction	6587	45	18% (8)	2% (1)						1.15
		Rap1 signaling pathway	2361	162	40% (64)	1% (2)				1.58		1.71
		HIF-1 signaling pathway	288399	76	50% (38)	14% (11)				2.14	1.35	

	KEGG B-LEVEL TERMS	KEGG PATHWAYS	Total gene counts	Total genes in pathway	Percent (No.) genes present	Percent (No.) genes enriched	Average enrichment scores: HTw | LTw	HTw | HTf	LTw | LTf	Enrichment leanings: Diffuse fluids	HT worms	LT worms
EIP (cont.)	Membrane transport	**Bacterial secretion system ***	1093646	74	88% (65)	39% (29)					3.23	
		Phosphotransferase system (PTS)	68747	71	63% (45)	38% (27)				1.98	3.43	
		ABC transporters	1692809	515	71% (367)	28% (142)				1.63	1.50	
	Signaling molecules and interaction	**Neuroactive ligand-receptor interaction**	2160	310	22% (67)	2% (6)					1.43	1.57
		Cell adhesion molecules	365	101	21% (21)	1% (1)						2.50
		ECM-receptor interaction	1531	65	22% (14)	3% (2)				1.58		1.58
CELLULAR PROCESSES	Transport and catabolism	**Endocytosis**	2561	168	65% (109)	1% (2)						3.03
		Lysosome	97012	105	76% (80)	11% (12)				1.27	3.03	
		Autophagy - yeast	18659	82	56% (46)	4% (3)						1.39
		Peroxisome	261354	73	78% (57)	14% (10)				0.86	2.21	
		Autophagy - animal	3671	109	67% (73)	3% (3)						1.29
		Mitophagy - yeast	683	34	53% (18)	9% (3)				1.92		1.68
		Autophagy - other	693	23	87% (20)	9% (2)				1.53		1.28
		Mitophagy - animal	1958	60	48% (29)	2% (1)				1.16		
		Phagosome	4930	95	43% (41)	2% (2)				1.95		
	Cell motility	**Flagellar assembly ***	572406	51	88% (45)	59% (30)					2.86	
		Bacterial chemotaxis	468114	26	92% (24)	77% (20)				2.24	3.47	
		Regulation of actin cytoskeleton	2516	163	39% (63)	2% (3)				1.58		1.57
	Cell growth and death	**Necroptosis**	212556	96	34% (33)	7% (7)					2.23	1.68
		Cell cycle - Caulobacter	661084	31	100% (31)	39% (12)					1.34	1.97
		Apoptosis - fly	97365	56	52% (29)	7% (4)					2.45	
		Apoptosis	60271	110	35% (39)	5% (5)					1.41	0.54
		Meiosis - yeast	58068	103	48% (49)	1% (1)					1.54	
		Apoptosis - multiple species	43203	40	28% (11)	3% (1)					1.35	
		Cellular senescence	1646	117	47% (55)	3% (3)				0.17	1.41	
		Ferroptosis	104072	33	45% (15)	9% (3)				2.17	3.00	
		Oocyte meiosis	1713	79	67% (53)	3% (2)				0.17		
		p53 signaling pathway	43818	67	33% (22)	4% (3)				1.89	1.38	
		Cell cycle - yeast	2529	126	51% (64)	1% (1)				1.28		
		Cell cycle	5010	110	64% (70)	3% (3)				1.85	1.41	
	Cellular community - eukaryotes	**Gap junction**	1703	60	52% (31)	3% (2)						2.84
		Adherens junction	912	63	48% (30)	2% (1)						1.71
		Signaling pathways regulating pluripotency of stem cells	552	114	44% (50)	1% (1)						1.42
		Focal adhesion	2581	149	35% (52)	3% (4)				1.58		1.68
		Tight junction	4535	106	55% (58)	2% (2)				0.05		
	Cellular community - prokaryotes	**Biofilm formation - Pseudomonas aeruginosa**	293200	90	51% (46)	22% (20)					3.00	1.79
		Biofilm formation - Vibrio cholerae	594794	106	55% (58)	20% (21)					2.69	1.45
		Quorum sensing	1410095	282	43% (121)	19% (54)				0.09	2.34	
		Biofilm formation - Escherichia coli	274417	61	64% (39)	33% (20)				0.68	2.52	

Less characterized genes

	KEGG B-LEVEL TERMS	KEGG C-LEVEL CATEGORIES	Total gene counts	No. genes enriched	Average enrichment scores: HTw | LTw	HTw | HTf	LTw | LTf	Enrichment leanings: Diffuse fluids	HT worms	LT worms
BRITE HIERARCHIES ONLY	Protein families: metabolism	**Peptidoglycan biosynthesis and degradation proteins**	259319	13					2.85	1.95
		Lipopolysaccharide biosynthesis proteins	54262	11					3.41	1.36
		Protein phosphatases and associated proteins	5921	7				0.13	2.56	
		Peptidases and inhibitors	986745	67					2.17	
		Protein kinases	237512	14					1.47	
		Prenyltransferases	24879	2				0.18	1.56	
		Polyketide biosynthesis proteins	3462	1						1.31
		Amino acid related enzymes	55415	5				1.26	1.64	
		Lipid biosynthesis proteins	34977	3				0.26	1.21	
		Cytochrome P450	5143	1				1.34		
		Glycosyltransferases	57716	17				4.22	2.86	
		Photosynthesis proteins	13793	4				3.09		

Less characterized genes (cont.)

	KEGG B-LEVEL TERMS	KEGG C-LEVEL CATEGORIES	Total gene counts	No. genes enriched	Average enrichment scores: HTw | LTw	HTw | HTf	LTw | LTf	Enrichment leanings: Diffuse fluids	HT worms	LT worms
BRITE HIERARCHIES ONLY	Protein families: genetic information processing	**Ubiquitin system**	3469	3					1.57	2.49
		Membrane trafficking	14328	5					0.57	1.61
		Messenger RNA biogenesis	52382	6				0.64	2.34	
		Transcription factors	603832	64				0.53	2.01	
		Spliceosome	6010	1						1.33
		DNA repair and recombination proteins	798067	25				1.28	1.91	
		Chaperones and folding catalysts	870899	35				0.88	1.17	
		Transcription machinery	525808	15				1.89	1.96	
		Transfer RNA biogenesis	1396229	56				2.09	1.76	
		Chromosome and associated proteins	1108479	37				0.68	0.97	
		Mitochondrial biogenesis	724217	20				1.63	1.32	
		Ribosome biogenesis	1653872	45				2.81	1.72	
		Ribosome	711	2				3.79	1.98	
		DNA replication proteins	461408	13				2.20	1.30	
		Translation factors	763057	23				2.75	1.40	
	Protein families: signaling and cellular processes	**Prokaryotic defense system**	1426544	59					2.85	1.45
		Two-component system	280834	13					2.75	1.48
		Cytoskeleton proteins	259115	7					2.39	1.66
		Secretion system	543269	44					3.00	0.98
		Bacterial toxins	16161	2					3.57	
		Ion channels	2037	3						3.49
		Bacterial motility proteins	362722	15					3.00	
		Domain-containing proteins not elsewhere classified	802	1						1.77
		Transporters	3735222	189				0.49	2.15	
		G protein-coupled receptors	381	1						1.50
		GTP-binding proteins	326	1						1.40
		Proteoglycans	53	1						1.40
		CD molecules	3411	3				0.05	0.97	0.30
		Exosome	97092	8				1.78	2.41	
		Antimicrobial resistance genes	22104	15				0.84	1.32	
UNCLASSIFIED	Unclassified: metabolism	**Glycan metabolism**	39317	3					2.96	1.62
		Secondary metabolism	11649	2					3.91	
		Carbohydrate metabolism	50258	8				0.77	3.96	
		Energy metabolism	423588	29				0.75	2.66	
		Enzymes with EC numbers	2415634	217				0.75	2.40	
		Amino acid metabolism	66015	10				1.84	3.48	
		Others	13098	2				0.01	1.04	
		Lipid metabolism	774	1				1.59		
		Cofactor metabolism	77937	3				2.05	1.59	
	Unclassified: genetic information processing	**Replication and repair**	926108	28					2.65	0.80
		Others	13098	1					2.24	
		Transcription	115852	3				0.00	1.22	
		Protein processing	415908	18				1.59	1.96	
		Translation	28487	6				1.39	0.96	
	Unclassified: signaling and cellular processes	**Unclassified viral proteins**	99531	3					3.11	
		Cell growth	103130	14					2.51	
		Structural proteins	303326	18				0.30	2.66	
		Transport	405761	19				0.55	2.31	
		Others	13098	11					1.73	
		Cell motility	21125	2					0.16	
		Signaling proteins	203477	16				1.65	1.20	
	Poorly characterized	Function unknown	2309658	166				0.82	2.27	
		General function prediction only	731776	50				1.29	1.45	

Chapter 5. General conclusion

Hydrothermal vent ecosystems have captivated both researchers and casual science-minded enthusiasts for over four decades, yet the research presented here highlights the continued potential for meaningful discoveries in these enigmatic systems. This research addressed critical gaps in understanding of hydrothermal vent ecosystem function, including the diversity, distribution, and potential roles of microeukaryotes and the functional contributions of non-endosymbiotic microbes in faunal assemblages. In this chapter, I summarize the scientific advancements that resulted from this research, discuss the limitations of the results and methods used, and describe future research directions.

5.1 Scientific Advancements

This research set out with three objectives. The first was to uncover the habitat preferences of hydrothermal vent microeukaryotes and identify potential endemics, which was addressed in Chapters 2 and 3. In chapter 2, research that was published in *Environmental Microbiology* in 2019, I addressed these goals using 18S rRNA gene sequencing of a large collection of samples from diffuse, near-field, plume and background fluids along the Mariana Arc. I characterized the distribution and relative abundance of OTUs to distinguish vent-only and cosmopolitan OTUs and used the frequency of abundance among sequence libraries to reveal that vent-only OTUs were rare and infrequently abundant. Furthermore, many 18S rRNA gene sequences reported in previous hydrothermal vent studies likely belong to organisms with a cosmopolitan distribution across vent and background habitats and not vent specialists. Using pairwise co-occurrences between prokaryotes with known

preferences for extreme habitat conditions and microeukaryotes, I identified microeukaryotic groups that may also tolerate acidic, high temperature, or low oxygen conditions. I suggest tolerance rather than preference as the infrequency with which these microeukaryotes reached the 1% relative abundance threshold did not suggest thriving populations similar to those of vent endemic Epsilonbacteraeota.

The use of network co-occurrences presented a complicated challenge in deciphering meaningful community patterns from the tangled web of pairwise interactions. There are varied approaches to network analysis and strong opinions surrounding their usefulness for inferring real biological associations (Röttjers and Faust, 2018; Blanchet et al., 2020). Therefore, increased attempts to confirm inferred species tolerances will only strengthen this field. Clustering of prokaryotes with known crustal habitat preferences and tolerance to extremes in the co-occurrence network offered a solid basis for inferring similar preferences and tolerances for microeukaryotes in close association with the cluster. Inference of potential specialization for hydrothermal habitats from combined co-occurrence and distribution and abundance data included apusomonads, excavates, labyrinthulids, and two clades within the Syndiniales. These inferences provide direction for future targeted investigations and culturing efforts.

The second objective was to characterize the microbial diversity living within the relative stability of faunal habitats compared to microbes in the associated diffuse fluids. Exploration of microbes among faunal assemblages initially stemmed from the goal of characterizing microeukaryotic diversity, thereby further contributing to the first research objective though expanded information on vent endemism in hydrothermal habitats other than fluids. The only study to assess the diversity of microeukaryotes in close association with

hydrothermal vent fauna was the first ever vent microeukaryote paper by Small and Gross in 1985, and none have continued that work until now. However, given the likely dependence of microeukaryotes on prokaryotic food sources, objective two was expanded to an assessment of the full microbial assemblage across the range of habitat conditions of the foundation tubeworm species *Ridgeia piscesae*.

In Chapter 3, I used 16S and 18S rRNA gene sequences and quantitative PCR from samples of faunal assemblages and diffuse fluids from three locations along the Juan de Fuca Ridge to assess differences in microbial composition and diversity between diffuse fluid and non-endosymbiotic, tubeworm-associated assemblages. I found increased microbial diversity in tubeworm assemblages relative to fluids, particularly in the dominant bacterial domain, and identified microbial groups with significant increases in OTU richness, including several microeukarya. The larger project that evolved with these samples, incorporated macro- and meiofaunal species abundances, provided by members of the Tunnicliffe lab, to identify covarying microbial and faunal taxa that may be linked through interspecies interactions. Faunal and microbial taxa that were enriched in tubeworm grabs and covaried across the range of tubeworm habitat conditions were termed 'core community' members. The intention of identifying a core community was to provide a way to focus the functional metagenomic research in Chapter 4 on microbial taxa with the greatest potential importance in faunal habitats.

Faunal and microbial taxonomic composition indicated habitat delineation into highT and lowT assemblages with base temperatures around 25°C, and core communities were described for both habitats. HighT core communities included Epsilonbacteraeota and Bacteroidetes along with faunal taxa known to inhabit the higher temperature bushes,

including paralvinellid worms and dirivultid copepods. LowT core communities involved Alpha- and Gammaproteobacteria and several meiofauna, including juvenile macrofauna. This research is presented in a manuscript that has now been accepted for publication in *ISME Communications*.

Use of co-occurrences in Chapter 3 was modified to reflect the compositional nature of high-throughput sequence data, and utilized the metric of proportionality, or covariance across samples. Prior to that modification, samples had been analyzed using the same Spearman correlation method as in Chapter 2. Resulting covariance of faunal taxa and partitioning in network analyses was more congruent with the observed seafloor distributions of species using proportionality than using Spearman correlations (V. Tunnicliffe, personal communication), confirming proportionality as the preferred method. Given the contention surrounding the use of co-occurrences for inferring real biological associations, as mentioned above, this shift in methodologies may help move the practice towards greater acceptance and utility.

The third objective was to evaluate the potential functional contributions of microbes in tubeworm habitats. Using metagenomic sequencing of DNA from the same Juan de Fuca Ridge samples as in Chapter 3, I assessed the relative contributions of core community and abundant non-core community taxa to selected microbial functions in Chapter 4. I developed methods to foster exploratory investigations using highly complex metagenome data and identified functional differences between core taxa that distinguished highT from lowT grabs and tubeworm grabs from diffuse fluids. This work is still in progress, but the methods and preliminary results presented in Chapter 4 will be expanded into a more comprehensive analysis for publication in the coming year.

5.2 Limitations of this research

Limitations and difficulties were primarily encountered during the functional characterization in Chapter 4, but some limitations also apply to interpreting 16S/18S rRNA results. High microbial diversity necessitated binning OTUs by taxonomic identity for co-occurrence/covariance analyses to avoid a network that was too dense to interpret. Analyzing combined OTUs obscures some of the diversity and may have affected the strength of associations. In addition, large numbers of 16S/18S rRNA genes with no close database relatives offer little room for interpretation. While these novel groups of organisms illicit some excitement for the implied discovery potential, including them in any sort of co-occurrence/covariance analysis was not particularly meaningful, particularly when the group of "Unclassified (order/class/phylum/domain)" contained several hundred species-level OTUs.

One obvious caveat of the findings in Chapter 2, that unfortunately came to light after it was published, is that those data were not treated as compositional. At the time these methods came to my attention I had just completed co-occurrence analysis of the Juan de Fuca Ridge microbes, macro- and meiofauna using the same methods as in Chapter 2. After compositional reanalysis of the data, the results more closely agreed with seafloor observations and previous research (Verena Tunnicliffe, personal communication). Compositional data methods are still not widely applied in microbial ecology although the reasoning for using them seems sound and the reanalysis of the Chapter 3 data supports it as well.

A few lessons were learned during the metagenomic analysis. 1) Metagenomic analysis is an inherently complicated tool for functional exploration, but with thoughtfully designed

approaches to guide explorations, such as the development of enrichment leanings and prior identification of core community organisms, it is possible to glean meaningful information on the differences between communities. 2) It is important to appreciate the usefulness and recognize the limits of metagenomic data. DNA evidence of genes does not equate to a functional contribution, but rather the potential for it. However, combined approaches that include metagenomics, metatranscriptomics, MAGs, and targeted incubations, can offer more informative insights into community functions. 3) Inferences from metagenomic data are only as good as the reference databases and tools. Some identified core community taxa did not have a sequenced genome so several genomes from the same family were included instead. This problem increases exponentially with taxa that were unclassified at the level of family or higher. The low percentage of genes with an identified product speaks to major gaps in reference databases, as well as failings of the IMG annotator itself, the current version of which misses nearly 20% of the genes in the KEGG database.

A major limitation of metagenomic data comes from the propensity of microbes to share genes through horizontal gene transfer (HGT). Does an abundance of rTCA cycle genes identified as belonging to *Sulfurovum* relate directly to the numbers of cells of *Sulfurovum*, or have many species incorporated genes for this carbon fixation pathway from *Sulfurovum* because it is a useful process in these habitats? Fortunately, tools are now being developed to identify evidence of HGT in metagenomes.

5.3 Future research directions

The primary focus of continued research will be on expanding exploration of the Juan de Fuca Ridge metagenomes to produce a more complete picture of functional potential. Among

the goals will be to identify pathways that exhibit signs of functional redundancy across the sample suite, as was suggested by the variety of processes involved in carbon fixation. These functions will provide targets for combined approaches described below (4). Continued evaluation of functional metrics will include the balance of aerobic-anaerobic processes to evaluate the level of adaptation to low oxygen conditions among microbes and additional cycling genes related to methane and hydrogen. Genes used for indicating heterotrophy will also be expanded to include sugar, fatty acid and amino acid transporters.

The results presented so far have suggested several directions for additional future research into hydrothermal vent microeukaryotes and non-endosymbiotic microbes in faunal assemblages.

1) The high degree of novelty among potential vent-endemic microeukaryotes indicates the need for increased efforts to characterize these organisms. Attempting to culture them is an obvious option, with which I have had some success over the last decade from research expeditions in the Mariana Arc, Tonga Arc, and Juan de Fuca Ridge. Cultures have been maintained in the lab and identified by 18S rRNA gene sequencing including two species of fungi, two stramenopiles, one excavate cultured from multiple locations, one amoeba, two euglenozoa, and a ciliate. These have been the result of methods trials to develop useful culture media and in-situ colonization devices, but future attempts will include more specific testing of growth conditions and genome sequencing. Alternatives to culturing include single-cell genomics, which has been used to define several new branches in the tree of life in recent years (Rinke et al., 2013; Rinke et al., 2014).

2) In addition to culturing, identification of microeukaryote genes in metagenomes may be useful in gaining some information about these organisms. Most annotation pipelines, including the IMG pipeline used in Chapter 4, are not geared towards identifying genes from eukaryotes, which have a more complicated structure of introns and exons than prokaryotes. This may explain, in part, the low percent of genes with an identified product in the Juan de Fuca Ridge metagenomes. Eukaryotic reads have already been identified from these metagenomes using tools designed for identifying and annotating eukaryotic genes (West et al., 2018; Humann et al., 2019), and work will continue towards creating taxonomic bins that may provide some species-level information. Practically speaking, obtaining sequence information from microeukaryotes will be simpler from future metatranscriptomes, as the introns are removed during transcription to mRNA.

3) Analysis of metagenome assembled genomes (MAGs) can provide species and strain-level functional potential from both known and novel taxa. MAGs have already been created from the Juan de Fuca Ridge metagenomes (Table 5.3) including 96 of high quality (>90% completion, <5% contamination) and 26 of moderate quality (>80% completion, <5% contamination). They belong to 14 phyla and include several that are unclassified at the levels of genera to phyla, and even two that were only classified as Bacteria. Among the high-quality MAGs are seven belonging to the *Ridgeia piscesae* symbiont *Endoriftia persephone*. These MAGs, assembled from seven different highT and lowT grab samples, will provide opportunities to explore genomic variations, similar to what has been done for *Sulfurovum* (Anderson et al., 2017), in the free-living population of symbionts. Analysis of MAGs will

likely lead to two manuscripts, one characterizing the novel species and one in collaboration with Maëva Perez on *Endoriftia*, in addition to the primary metagenomic data manuscript.

4) Improved understanding of the contributions of microbes in faunal assemblages will come from combining metagenomics with process-specific incubation experiments, metatranscriptomics, and generation of MAGs. Excellent examples undertaken using diffuse fluid samples will provide guidance (Fortunato and Huber, 2016; Fortunato et al., 2018; Galambos et al., 2019). One limitation that will need to be overcome to pursue RNA-based work from faunal assemblages is the urgency with which samples must be preserved after removal from the natural setting to avoid RNA degradation. RNA is a transient molecule, so the time between sampling and storage in an RNA preservative must be minutes, not hours. This limitation made such sampling implausible for the current research, but plans are being discussed for a method of preservation on the seafloor within minutes after sampling to allow future progress.

5) A major goal and challenge for future work will be to try to confirm interspecies interactions between microbes and fauna inferred within core communities. Some may be difficult if not impossible to confirm as covariance patterns may simply indicate shared environmental preferences. While inferred interactions between microbial taxa may be supported by metagenomic data, supporting evidence of interactions between microbes and fauna may require some creatively designed experiments.

These future research directions will improve our understanding of these captivating ecosystems and hopefully bring many more discoveries.

Table 5.1 List of Metagenome Assembled Genomes

Phylum/Class	Classification	No. High-quality MAGs	No. Medium-quality MAGs
Acidobacteria	Class Aminicenantia	3	
	Class Thermoanaerobaculia	1	
	Unclassified	1	1
Bacteroidetes	Order Bacteroidales	11	3
	Order Flavobacteriales	1	
	Family Marinifilaceae	2	
	Family Salinivirgaceae	1	
	Family Saprospiraceae		1
	Family Crocinitomicaceae	1	
	Family Flavobacteriaceae	1	2
	Family Melioribacteraceae	2	1
	Draconibacterium	1	
Calditrichaeota	Order Calditrichales	1	
	Unclassified		1
Epsilonbacteraeota	Order Campylobacterales	2	1
	Family Arcobacteraceae	5	
	Family Sulfurimonadaceae	2	1
	Family Sulfurospirillaceae	2	
	Family Sulfurovaceae	8	1
	Sulfurimonas	2	
	Sulfurovum	2	
	Nitratifractor	2	1
Chloroflexi	Family Caldilineaceae	1	1
Deferribacteraeota	Family Deferribacterales	1	1
Desulfobacteraeota	Order Desulfatiglandales	1	
	Order Desulfobacterales	2	
	Order Dissulfuribacterales	1	
	Family Desulfarculaceae	1	
	Family Desulfobacteraceae	2	
	Family Desulfobulbaceae	6	
	Family Desulfomonilaceae	1	
	Unclassified	2	2
Fermentibacteraeota	Family Fermentibacteraceae	1	
Fusobacteria	*Psychrilyobacter*		1
Gemmatimonadota	Unclassified	1	
Myxococcota	Unclassified		1
Planctomycetes	Unclassified		1
Alphaproteobacteria	Family Anderseniellaceae	1	
	Unclassified	1	
Gammaproteobacteria	Order Thiohalomonadales	1	
	Family Thioglobaceae	1	
	Family Thiotrichaceae	3	1
	Family Marinicellaceae		1
	Marithrix		1
	Endoriftia persephone	7	
	Thiolapillus	1	
	Methylomarinum	2	
	Methyloprofundus	2	
	Thioglobus		1
Verrucomicrobia	Order Victivallales	4	1
	Family Akkermansiaceae		1
Unclassified Bacteria		2	
	TOTAL MAGs	**96**	**26**

References

Adl, S.M., Bass, D., Lane, C.E., Lukes, J., Schoch, C.L., Smirnov, A. et al. (2019) Revisions to the Classification, Nomenclature, and Diversity of Eukaryotes. *Journal of Eukaryotic Microbiology* **66**: 4-119.

Aitchison, J. (1986) *The Statistical Analysis of Compositional Data*. London: Chapman & Hall.

Aitchison, J., Barcelo-Vidal, C., Martin-Fernandez, J.A., and Pawlowsky-Glahn, V. (2000) Logratio analysis and compositional distance. *Mathematical Geology* **32**: 271-275.

Akerman, N.H., Butterfield, D.A., and Huber, J.A. (2013) Phylogenetic diversity and functional gene patterns of sulfur-oxidizing subseafloor Epsilonproteobacteria in diffuse hydrothermal vent fluids. *Frontiers in Microbiology* **4**.

Alain, K., Olagnon, M., Desbruyères, D., Pagé, A., Barbier, G., Juniper, S.K. et al. (2002) Phylogenetic characterization of the bacterial assemblage associated with mucous secretions of the hydrothermal vent polychaete Paralvinella palmiformis. *FEMS Microbiology Ecology* **42**: 463-476.

Amann, R.I., Ludwig, W., and Schleifer, K.H. (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol Rev* **59**: 143-169.

Amend, J.P., McCollom, T.M., Hentscher, M., and Bach, W. (2011) Catabolic and anabolic energy for chemolithoautotrophs in deep-sea hydrothermal systems hosted in different rock types. *Geochimica Et Cosmochimica Acta* **75**: 5736-5748.

Anderson, R.E., Sogin, M.L., and Baross, J.A. (2014) Evolutionary strategies of viruses, bacteria and archaea in hydrothermal vent ecosystems revealed through metagenomics. *PLoS One* **9**: e109696.

Anderson, R.E., Reveillaud, J., Reddington, E., Delmont, T.O., Eren, A.M., McDermott, J.M. et al. (2017) Genomic variation in microbial populations inhabiting the marine subseafloor at deep-sea hydrothermal vents. *Nat Commun* **8**: 1114.

Arístegui, J., Gasol, J.M., Duarte, C.M., and Herndl, G.J. (2009) Microbial oceanography of the dark ocean's pelagic realm. *Limnology and Oceanography* **54**: 1501-1529.

Arrigo, K.R. (2005) Marine microorganisms and global nutrient cycles. *Nature* **437**: 349-355.

Atkins, M.S., Teske, A.P., and Anderson, O.R. (2000) A survey of flagellate diversity at four deep-sea hydrothermal vents in the eastern Pacific Ocean using structural and molecular approaches. *Journal of Eukaryotic Microbiology* **47**: 400-411.

Atkins, M.S., Hanna, M.A., Kupetsky, E.A., Saito, M.A., Taylor, C.D., and Wirsen, C.O. (2002) Tolerance of flagellated protists to high sulfide and metal concentrations potentially encountered at deep-sea hydrothermal vents. *Marine Ecology Progress Series* **226**: 63-75.

Azam, F., Fenchel, T., Field, J.G., Gray, J.S., Meyerreil, L.A., and Thingstad, F. (1983) The Ecological Role of Water-Column Microbes in the Sea. *Marine Ecology Progress Series* **10**: 257-263.

Baldrighi, E., Aliani, S., Conversi, A., Lavaleye, M., Borghini, M., and Manini, E. (2013) From microbes to macrofauna: an integrated study of deep benthic communities and their response to environmental variables along the Malta Escarpment (Ionian Sea). *Scientia Marina* **77**: 625-639.

Bass, D., and Cavalier-Smith, T. (2009). Cercozoa. Version 22 March 2009 (under construction). URL http://tolweb.org/Cercozoa/121187/2009.03.22

Bates, A.E., Harmer, T.L., Roeselers, G., and Cavanaugh, C.M. (2011) Phylogenetic Characterization of Episymbiotic Bacteria Hosted by a Hydrothermal Vent Limpet (Lepetodrilidae, Vetigastropoda). *Biological Bulletin* **220**: 118-127.

Behnke, A., Engel, M., Christen, R., Nebel, M., Klein, R.R., and Stoeck, T. (2011) Depicting more accurate pictures of protistan community complexity using pyrosequencing of hypervariable SSU rRNA gene regions. *Environ Microbiol* **13**: 340-349.

Bergquist, D.C., Eckner, J.T., Urcuyo, I.A., Cordes, E.E., Hourdez, S., Macko, S.A., and Fisher, C.R. (2007) Using stable isotopes and quantitative community characteristics to determine a local hydrothermal vent food web. *Marine Ecology Progress Series* **330**: 49-65.

Biard, T., Bigeard, E., Audic, S., Poulain, J., Gutierrez-Rodriguez, A., Pesant, S. et al. (2017) Biogeography and diversity of Collodaria (Radiolaria) in the global ocean. *Isme Journal* **11**: 1331-1344.

Bjorbaekmo, M.F.M., Evenstad, A., Rosaeg, L.L., Krabberod, A.K., and Logares, R. (2020) The planktonic protist interactome: where do we stand after a century of research? *ISME J* **14**: 544-559.

Blanchet, F.G., Cazelles, K., and Gravel, D. (2020) Co-occurrence is not evidence of ecological interactions. *Ecol Lett* **23**: 1050-1063.

Bourbonnais, A., Lehmann, M.F., Butterfield, D.A., and Juniper, S.K. (2012) Subseafloor nitrogen transformations in diffuse hydrothermal vent fluids of the Juan de Fuca Ridge evidenced by the isotopic composition of nitrate and ammonium. *Geochem Geophys Geosyst* **13**.

Bower, S.M., Carnegie, R.B., Goh, B., Jones, S.R., Lowe, G.J., and Mak, M.W. (2004) Preferential PCR amplification of parasitic protistan small subunit rDNA from metazoan tissues. *J Eukaryot Microbiol* **51**: 325-332.

Bråte, J., Krabberød, A.K., Dolven, J.K., Ose, R.F., Kristensen, T., Bjørklund, K.R., and Shalchian-Tabrizi, K. (2012) Radiolaria associated with large diversity of marine alveolates. *Protist* **163**: 767-777.

Brazelton, W.J., and Baross, J.A. (2009) Abundant transposases encoded by the metagenome of a hydrothermal chimney biofilm. *ISME J* **3**: 1420-1424.

Brown, M.W., Sharpe, S.C., Silberman, J.D., Heiss, A.A., Lang, B.F., Simpson, A.G., and Roger, A.J. (2013) Phylogenomics demonstrates that breviate flagellates are related to opisthokonts and apusomonads. *Proc Biol Sci* **280**: 20131755.

Bruno, J.F., and Bertness, M.D. (2001) Habitat modification and facilitation in benthic marine communities. In *Marine Community Ecology*. Bertness, M.D., Gaines, S.D., and Hay, M.E. (eds). Sunderland, MA: Sinauer Associates, pp. 201-218.

Burgaud, G., Le Calvez, T., Arzur, D., Vandenkoornhuyse, P., and Barbier, G. (2009) Diversity of culturable marine filamentous fungi from deep-sea hydrothermal vents. *Environmental Microbiology* **11**: 1588-1600.

Burki, F. (2014) The eukaryotic tree of life from a global phylogenomic perspective. *Cold Spring Harb Perspect Biol* **6**: a016147.

Burki, F., and Keeling, P.J. (2014) Rhizaria. *Current Biology* **24**: R103-R107.

Butterfield, D.A., Massoth, G.J., McDuff, R.E., Lupton, J.E., and Lilley, M.D. (1990) Geochemistry of hydrothermal fluids from Axial Seamount Hydrothermal Emissions Study vent field, Juan de Fuca Ridge: subseafloor boiling and subsequent fluid-rock interaction. *Journal of Geophysical Research* **95**: 12895-12921.

Butterfield, D.A., Roe, K.K., Lilley, M.D., Huber, J.A., Baross, J.A., Embley, R.W., and Massoth, G.J. (2004) Mixing, reaction and microbial activity in the sub-seafloor revealed by temporal and spatial variation in diffuse flow vents at Axial Volcano. *Subseafloor Biosphere at Mid-Ocean Ranges* **144**: 269-289.

Campbell, B.J., Engel, A.S., Porter, M.L., and Takai, K. (2006) The versatile epsilon-proteobacteria: key players in sulphidic habitats. *Nat Rev Microbiol* **4**: 458-468.

Campbell, B.J., Jeanthon, C., Kostka, J.E., Luther, G.W., 3rd, and Cary, S.C. (2001) Growth and phylogenetic properties of novel bacteria belonging to the epsilon subdivision of the Proteobacteria enriched from Alvinella pompejana and deep-sea hydrothermal vents. *Appl Environ Microbiol* **67**: 4566-4572.

Carignan, V., and Villard, M.A. (2002) Selecting indicator species to monitor ecological integrity: a review. *Environ Monit Assess* **78**: 45-61.

Caron, D.A. (2009) New Accomplishments and Approaches for Assessing Protistan Diversity and Ecology in Natural Ecosystems. *Bioscience* **59**: 287-299.

Caron, D.A., and Countway, P.D. (2009) Hypotheses on the role of the protistan rare biosphere in a changing world. *Aquatic Microbial Ecology* **57**: 227-238.

Caron, D.A., Gast, R., Countway, P.D., and Heidelberg, K.B. (2009a) Microbial eukaryote diversity and biogeography. In *Microbe*, pp. 71-77.

Caron, D.A., Countway, P.D., Savai, P., Gast, R.J., Schnetzer, A., Moorthi, S.D. et al. (2009b) Defining DNA-based operational taxonomic units for microbial-eukaryote ecology. *Appl Environ Microbiol* **75**: 5797-5808.

Cavalier-Smith, T., and Chao, E.E. (2010) Phylogeny and Evolution of Apusomonadida (Protozoa: Apusozoa): New Genera and Species. *Protist* **161**: 549-576.

Chadwick, W.W., Cashman, K.V., Embley, R.W., Matsumoto, H., Dziak, R.P., de Ronde, C.E.J. et al. (2008) Direct video and hydrophone observations of submarine explosive eruptions at NW Rota-1 volcano, Mariana arc. *Journal of Geophysical Research-Solid Earth* **113**.

Chen, I.A., Markowitz, V.M., Chu, K., Palaniappan, K., Szeto, E., Pillay, M. et al. (2017) IMG/M: integrated genome and metagenome comparative data analysis system. *Nucleic Acids Res* **45**: D507-D516.

Cho, B.C., and Azam, F. (1988) Major Role of Bacteria in Biogeochemical Fluxes in the Oceans Interior. *Nature* **332**: 441-443.

Colaço, A., Dehairs, F., and Desbruyères, D. (2002) Nutritional relations of deep-sea hydrothermal fields at the Mid-Atlantic Ridge: a stable isotope approach. *Deep-Sea Research Part I-Oceanographic Research Papers* **49**: 395-412.

Colaço, A., Desbruyères, D., and Guezennec, J. (2007) Polar lipid fatty acids as indicators of trophic associations in a deep-sea vent system community. *Marine Ecology-an Evolutionary Perspective* **28**: 15-24.

Comas-Cufí, M. (2019). coda.base: A Basic Set of Functions for Compositional Data Analysis. R package version 0.2.1. URL https://CRAN.R-project.org/package=coda.base

Cooney, E.W., Barr, D.J.S., and Barstow, W.E. (1985) The Ultrastructure of the Zoospore of Hyphochytrium-Catenoides. *Canadian Journal of Botany-Revue Canadienne De Botanique* **63**: 497-505.

Coyne, K.J., Countway, P.D., Pilditch, C.A., Lee, C.K., Caron, D.A., and Cary, S.C. (2013) Diversity and Distributional Patterns of Ciliates in Guaymas Basin Hydrothermal Vent Sediments. *Journal of Eukaryotic Microbiology* **60**: 433-447.

Cronin-O'Reilly, S., Taylor, J.D., Jermyn, I., Allcock, A.L., Cunliffe, M., and Johnson, M.P. (2018) Limited congruence exhibited across microbial, meiofaunal and macrofaunal benthic assemblages in a heterogeneous coastal environment. *Scientific Reports* **8**.

Danovaro, R., Gambi, C., Dell'Anno, A., Corinaldesi, C., Fraschetti, S., Vanreusel, A. et al. (2008) Exponential decline of deep-sea ecosystem functioning linked to benthic biodiversity loss. *Curr Biol* **18**: 1-8.

Dayton, P.K. (1972) Toward an Understanding of Community Resilience and the Potential Effects of Enrichments to the Benthos at McMurdo Sound, Antarctica. Pages 81-95. In *Proceedings of the Colloquium on Conservation Problems*. Parker, B.C. (ed). Lawrence, Kansas, USA.: Allen Press.

Del Campo, J., Pons, M.J., Herranz, M., Wakeman, K.C., Del Valle, J., Vermeij, M.J.A. et al. (2019) Validation of a universal set of primers to study animal-associated microeukaryotic communities. *Environ Microbiol* **21**: 3855-3861.

Dick, G.J. (2019) The microbiomes of deep-sea hydrothermal vents: distributed globally, shaped locally. *Nature Reviews Microbiology* **17**: 271-283.

Díez, B., Pedrós-Alió, C., and Massana, R. (2001) Study of genetic diversity of eukaryotic picoplankton in different ocean regions by small-subunit rRNA gene cloning and sequencing. *Applied and Environmental Microbiology* **67**: 2932-2941.

Djurhuus, A., Mikalsen, S.O., Giebel, H.A., and Rogers, A.D. (2017) Cutting through the smoke: the diversity of microorganisms in deep-sea hydrothermal plumes. *R Soc Open Sci* **4**: 160829.

Douglas, G.M., and Langille, M.G.I. (2019) Current and Promising Approaches to Identify Horizontal Gene Transfer Events in Metagenomes. *Genome Biol Evol* **11**: 2750-2766.

Du Preez, C., and Fisher, C.P. (2018) Long-Term Stability of Back-Arc Basin Hydrothermal Vents. *Frontiers in Marine Science* **5**.

Duarte, C.M. (2015) Seafaring in the 21St Century: The Malaspina 2010 Circumnavigation Expedition. *Limnology and Oceanography Bulletin* **24**: 11-14.

Dubilier, N., Bergin, C., and Lott, C. (2008) Symbiotic diversity in marine animals: the art of harnessing chemosynthesis. *Nat Rev Microbiol* **6**: 725-740.

Edgcomb, V., Orsi, W., Bunge, J., Jeon, S., Christen, R., Leslin, C. et al. (2011) Protistan microbial observatory in the Cariaco Basin, Caribbean. I. Pyrosequencing vs Sanger insights into species richness. *ISME J* **5**: 1344-1356.

Edgcomb, V.P., Kysela, D.T., Teske, A., de Vera Gomez, A., and Sogin, M.L. (2002) Benthic eukaryotic diversity in the Guaymas Basin hydrothermal vent environment. *Proceedings of the National Academy of Sciences* **99**: 7658-7662.

Embley, R.W., Baker, E.T., Butterfield, D.A., Chadwick, W.W., Lupton, J.E., Resing, J.A. et al. (2007) Exploring the Submarine Ring of Fire Mariana Arc - Western Pacific. *Oceanography* **20**: 68-79.

Embley, R.W., Chadwick, W.W., Baker, E.T., Butterfield, D.A., Resing, J.A., de Ronde, C.E.J. et al. (2006) Long-term eruptive activity at a submarine arc volcano. *Nature* **441**: 494-497.

Eren, A.M., Vineis, J.H., Morrison, H.G., and Sogin, M.L. (2013) A filtering method to generate high quality short reads using illumina paired-end technology. *PLoS One* **8**: e66643.

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

Fernandes, A.D., Macklaim, J.M., Linn, T.G., Reid, G., and Gloor, G.B. (2013) ANOVA-like differential expression (ALDEx) analysis for mixed population RNA-Seq. *PLoS One* **8**: e67019.

Forget, N.L., and Juniper, S.K. (2013) Free-living bacterial communities associated with tubeworm (Ridgeia piscesae) aggregations in contrasting diffuse flow hydrothermal vent habitats at the Main Endeavour Field, Juan de Fuca Ridge. *Microbiologyopen* **2**: 259-275.

Fortunato, C.S., and Huber, J.A. (2016) Coupled RNA-SIP and metatranscriptomics of active chemolithoautotrophic communities at a deep-sea hydrothermal vent. *ISME J*.

Fortunato, C.S., Larson, B., Butterfield, D.A., and Huber, J.A. (2018) Spatially distinct, temporally stable microbial populations mediate biogeochemical cycling at and below the seafloor in hydrothermal vent fluids. *Environ Microbiol* **20**: 769-784.

Foshtomi, M.Y., Braeckman, U., Derycke, S., Sapp, M., Van Gansbeke, D., Sabbe, K. et al. (2015) The Link between Microbial Diversity and Nitrogen Cycling in Marine Sediments Is Modulated by Macrofaunal Bioturbation. *Plos One* **10**.

Fryer, P., Gill, J.B., and Jackson, M.C. (1997) Volcanologic and tectonic evolution of the Kasuga seamounts, northern Mariana Trough: Alvin submersible investigations. *Journal of Volcanology and Geothermal Research* **79**: 277-+.

Galambos, D., Anderson, R.E., Reveillaud, J., and Huber, J.A. (2019) Genome-resolved metagenomics and metatranscriptomics reveal niche differentiation in functionally redundant microbial communities at deep-sea hydrothermal vents. *Environ Microbiol* **21**: 4395-4410.

Giner, C.R., Pernice, M.C., Balague, V., Duarte, C.M., Gasol, J.M., Logares, R., and Massana, R. (2020) Marked changes in diversity and relative activity of picoeukaryotes with depth in the world ocean. *ISME J* **14**: 437-449.

Girguis, P.R., and Lee, R.W. (2006) Thermal preference and tolerance of alvinellids. *Science* **312**: 231.

Gloor, G.B., and Reid, G. (2016) Compositional analysis: a valid approach to analyze microbiome high-throughput sequencing data. *Can J Microbiol* **62**: 692-703.

Gloor, G.B., Macklaim, J.M., Pawlowsky-Glahn, V., and Egozcue, J.J. (2017) Microbiome Datasets Are Compositional: And This Is Not Optional. *Front Microbiol* **8**: 2224.

Goffredi, S.K. (2010) Indigenous ectosymbiotic bacteria associated with diverse hydrothermal vent invertebrates. *Environmental Microbiology Reports* **2**: 479-488.

Gollner, S., Ivanenko, V.N., Arbizu, P.M., and Bright, M. (2010a) Advances in taxonomy, ecology, and biogeography of Dirivultidae (copepoda) associated with chemosynthetic environments in the deep sea. *PLoS One* **5**: e9801.

Gollner, S., Riemer, B., Arbizu, P.M., Le Bris, N., and Bright, M. (2010b) Diversity of Meiofauna from the 9 degrees 50 ' N East Pacific Rise across a Gradient of Hydrothermal Fluid Emissions. *Plos One* **5**.

Gong, J., Dong, J., Liu, X.H., and Massana, R. (2013) Extremely High Copy Numbers and Polymorphisms of the rDNA Operon Estimated from Single Cell Analysis of Oligotrich and Peritrich Ciliates. *Protist* **164**: 369-379.

Gooday, A.J., Schoenle, A., Dolan, J.R., and Arndt, H. (2020) Protist diversity and function in the dark ocean - Challenging the paradigms of deep-sea ecology with special emphasis on foraminiferans and naked protists. *Eur J Protistol* **75**: 125721.

Govenar, B. (2010) Shaping Vent and Seep Communities: Habitat Provision and Modification by Foundation Species. In *The vent and seep biota: aspects from microbes to ecosystems*. Kiel, S. (ed). Dordrecht: Springer, pp. 403-432.

Govenar, B. (2012) Energy Transfer Through Food Webs at Hydrothermal Vents: Linking the Lithosphere to the Biosphere. *Oceanography* **25**: 246-255.

Govenar, B.W., Bergquist, D.C., Urcuyo, I.A., Eckner, J.T., and Fisher, C.R. (2002) Three Ridgeia piscesae assemblages from a single Juan de Fuca Ridge sulphide edifice: structurally different and functionally similar. *Cahiers De Biologie Marine* **43**: 247-252.

Groisillier, A., Massana, R., Valentin, K., Vaulotl, D., and Guilloul, L. (2006) Genetic diversity and habitats of two enigmatic marine alveolate lineages. *Aquatic Microbial Ecology* **42**: 277-291.

Grossart, H.-P., Wurzbacher, C., James, T.Y., and Kagami, M. (2016) Discovery of dark matter fungi in aquatic ecosystems demands a reappraisal of the phylogeny and ecology of zoosporic fungi. *Fungal Ecology* **19**: 28-38.

Guilhon, M., Montserrat, F., and Turra, A. (2020) Recognition of ecosystem-based management principles in key documents of the seabed mining regime: implications and further recommendations. *ICES Journal of Marine Science*.

Guillou, L., Viprey, M., Chambouvet, A., Welsh, R., Kirkham, A., Massana, R. et al. (2008) Widespread occurrence and genetic diversity of marine parasitoids belonging to *Syndiniales (Alveolata*). *Environmental Microbiology* **10**: 3349-3365.

Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L. et al. (2013) The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. *Nucleic Acids Research* **41**: D597-D604.

Hamann, E., Gruber-Vodicka, H., Kleiner, M., Tegetmeyer, H.E., Riedel, D., Littmann, S. et al. (2016) Environmental Breviatea harbour mutualistic Arcobacter epibionts. *Nature* **534**: 254-258.

Hammer, Ø., Harper, D.A.T., and Ryan, P.D. (2001) PAST: Paleontological statistics software package for education and data analysis. *Palaeontologia Electronica* **4**: 1-9.

Hampl, V., Silberman, J.D., Stechmann, A., Diaz-Trivino, S., Johnson, P.J., and Roger, A.J. (2008) Genetic Evidence for a Mitochondriate Ancestry in the 'Amitochondriate' Flagellate Trimastix pyriformis. *Plos One* **3**.

Handley, K.M. (2019) Determining Microbial Roles in Ecosystem Function: Redefining Microbial Food Webs and Transcending Kingdom Barriers. *mSystems* **4**.

Herndl, G.J., Agogué, H., Baltar, F., Reinthaler, T., Sintes, E., and Varela, M.M. (2008) Regulation of aquatic microbial processes: the A 'microbial loop' of the sunlit surface waters and the dark ocean dissected. *Aquatic Microbial Ecology* **53**: 59-68.

Herndl, G.J., Reinthaler, T., Teira, E., van Aken, H., Veth, C., Pernthaler, A., and Pernthaler, J. (2005) Contribution of Archaea to total prokaryotic production in the deep Atlantic Ocean. *Applied and Environmental Microbiology* **71**: 2303-2309.

Hope, J.A., Paterson, D.M., and Thrush, S.F. (2020) The role of microphytobenthos in soft-sediment ecological networks and their contribution to the delivery of multiple ecosystem services. *Journal of Ecology* **108**: 815-830.

Horn, H.S. (1966) Measurement of Overlap in Comparative Ecological Studies. *American Naturalist* **100**: 419-&.

Hu, S.R.K., Herrera, E.L., Smith, A.R., Pachiadaki, M.G., Edgcomb, V.P., Sylva, S.P. et al. (2021) Protistan grazing impacts microbial communities and carbon cycling at deep-sea hydrothermal vents. *Proceedings of the National Academy of Sciences of the United States of America* **118**.

Huber, J.A., Butterfield, D.A., and Baross, J.A. (2002) Temporal changes in archaeal diversity and chemistry in a mid-ocean ridge subseafloor habitat. *Applied and Environmental Microbiology* **68**: 1585-1594.

Huber, J.A., Butterfield, D.A., and Baross, J.A. (2003) Bacterial diversity in a subseafloor habitat following a deep-sea volcanic eruption. *FEMS Microbiology Ecology* **43**: 393-409.

Huber, J.A., Cantin, H.V., Huse, S.M., Welch, D.B., Sogin, M.L., and Butterfield, D.A. (2010) Isolated communities of Epsilonproteobacteria in hydrothermal vent fluids of the Mariana Arc seamounts. *FEMS Microbiol Ecol* **73**: 538-549.

Huber, J.A., Mark Welch, D.B., Morrison, H.G., Huse, S.M., Neal, P.R., Butterfield, D.A., and Sogin, M.L. (2007) Microbial population structures in the deep marine biosphere. *Science* **318**: 97-100.

Hugerth, L.W., Larsson, J., Alneberg, J., Lindh, M.V., Legrand, C., Pinhassi, J., and Andersson, A.F. (2015) Metagenome-assembled genomes uncover a global brackish microbiome. *Genome Biol* **16**: 279.

Humann, J.L., Lee, T., Ficklin, S., and Main, D. (2019) Structural and Functional Annotation of Eukaryotic Genomes with GenSAS. *Methods Mol Biol* **1962**: 29-51.

Huse, S.M., Huber, J.A., Morrison, H.G., Sogin, M.L., and Welch, D.M. (2007) Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biol* **8**: R143.

Huse, S.M., Welch, D.B.M., Voorhis, A., Shipunova, A., Morrison, H.G., Eren, A.M., and Sogin, M.L. (2014) VAMPS: a website for visualization and analysis of microbial population structures. *Bmc Bioinformatics* **15**.

Ingels, J., Vanreusel, A., Pape, E., Pasotti, F., Macheriotou, L., Arbizu, P.M. et al. (2020) Ecological variables for deep-ocean monitoring must include microbiota and meiofauna for effective conservation. *Nat Ecol Evol*: doi: 10.1038/s41559-41020-01335-41556.

Ishibashi, J., and Urabe, T. (1995) Hydrothermal activity related to arc-backarc magmatism in the western Pacific. In *Backarc Basins: Techtonics and Magmatism*. Taylor, B. (ed). New York: Plenum Press, pp. 451-495.

Jackson, D.A. (1997) Compositional data in community ecology: The paradigm or peril of proportions? *Ecology* **78**: 929-940.

Jannasch, H.W., and Wirsen, C.O. (1979) Chemo-Synthetic Primary Production at East Pacific Sea-Floor Spreading Centers. *Bioscience* **29**: 592-598.

Johnson, K.S., Beehler, C.L., Sakamotoarnold, C.M., and Childress, J.J. (1986) Insitu Measurements of Chemical-Distributions in a Deep-Sea Hydrothermal Vent Field. *Science* **231**: 1139-1141.

Jones, E.B.G., and Fell, J.W. (2012) 4 Basidiomycota
Marine Fungi. In. Jones, E.B.G., and Pang, K.-L. (eds): De Gruyter, pp. 49-64.

Jungbluth, S.P., Amend, J.P., and Rappe, M.S. (2017) Metagenome sequencing and 98 microbial genomes from Juan de Fuca Ridge flank subsurface fluids. *Sci Data* **4**: 170037.

Kalanetra, K.M., and Nelson, D.C. (2010) Vacuolate-attached filaments: highly productive Ridgeia piscesae epibionts at the Juan de Fuca hydrothermal vents. *Mar Biol* **157**: 791-800.

Karl, D.M., Wirsen, C.O., and Jannasch, H.W. (1980) Deep-sea primary production at the Galapagos hydrothermal vents. *Science* **207**: 1345-1347.

Karsenti, E., Acinas, S.G., Bork, P., Bowler, C., De Vargas, C., Raes, J. et al. (2011) A Holistic Approach to Marine Eco-Systems Biology. *Plos Biology* **9**.

Kouris, A., Juniper, S.K., Frebourg, G., and Gaill, F. (2007) Protozoan-bacterial symbiosis in a deep-sea hydrothermal vent folliculinid ciliate (Folliculinopsis sp.) from the Juan de Fuca Ridge. *Marine Ecology-an Evolutionary Perspective* **28**: 63-71.

Kouris, A., Limen, H., Stevens, C.J., and Juniper, S.K. (2010) Blue mats: faunal composition and food web structure in colonial ciliate (Folliculinopsis sp.) mats at Northeast Pacific hydrothermal vents. *Marine Ecology Progress Series* **412**: 93-101.

Kühn, S., Medlin, L., and Eller, G. (2004) Phylogenetic position of the parasitoid nanoflagellate Pirsonia inferred from nuclear-encoded small subunit ribosomal DNA and a description of Pseudopirsonia n. gen. and Pseudopirsonia mucosa (Drebes) comb. nov. *Protist* **155**: 143-156.

Lamy, T., Koenigs, C., Holbrook, S.J., Miller, R.J., Stier, A.C., and Reed, D.C. (2020) Foundation species promote community stability by increasing diversity in a giant kelp forest. *Ecology* **101**: e02987.

Langmead, B., and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**: 357-359.

Larsen, J., and Patterson, D.J. (1990) Some Flagellates (Protista) from Tropical Marine-Sediments. *Journal of Natural History* **24**: 801-937.

Lee, W.-K., Juniper, S.K., Perez, M., Ju, S.-J., and Kim, S.-J. (In press) Diversity and characterization of bacterial communities of five co-occurring species at a hydrothermal vent on the Tonga Arc. *Ecology and Evolution*.

Leipe, D.D., Tong, S.M., Goggin, C.L., Slemenda, S.B., Pieniazek, N.J., and Sogin, M.L. (1996) 16S-like rDNA sequences from Developayella elegans, Labyrinthuloides haliotidis, and Proteromonas lacertae confirm that the stramenopiles are a primarily heterotrophic group. *European Journal of Protistology* **32**: 449-458.

Léveillé, R.J., Levesque, C., and Juniper, S.K. (2005) Biotic interactions and feedback processes in deep-sea hydrothermal vent ecosystems. In *Interactions between macro- and microorganisms in marine sediments*. Kristensen, E., Haese, R.R., and Kostka, J.E. (eds). Washington, DC: American Geophysical Union, pp. 299-321.

Levin, L.A., and Le Bris, N. (2015) The deep ocean under climate change. *Science* **350**: 766-768.

Levin, L.A., Baco, A.R., Bowden, D.A., Colaco, A., Cordes, E.E., Cunha, M.R. et al. (2016) Hydrothermal Vents and Methane Seeps: Rethinking the Sphere of Influence. *Frontiers in Marine Science* **3**.

Li, D., Liu, C.M., Luo, R., Sadakane, K., and Lam, T.W. (2015) MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**: 1674-1676.

Lie, A.A.Y., Liu, Z.F., Hu, S.K., Jones, A.C., Kim, D.Y., Countway, P.D. et al. (2014) Investigating Microbial Eukaryotic Diversity from a Global Census: Insights from a Comparison of Pyrotag and Full-Length Sequences of 18S rRNA Genes. *Applied and Environmental Microbiology* **80**: 4363-4373.

Lima-Mendez, G., Faust, K., Henry, N., Decelle, J., Colin, S., Carcillo, F. et al. (2015) Ocean plankton. Determinants of community structure in the global plankton interactome. *Science* **348**: 1262073.

Limen, H., Stevens, C.J., Bourass, Z., and Juniper, S.K. (2008) Trophic ecology of siphonostomatoid copepods at deep-sea hydrothermal vents in the northeast Pacific. *Marine Ecology Progress Series* **359**: 161-170.

Logares, R., Audic, S., Bass, D., Bittner, L., Boutte, C., Christen, R. et al. (2014) Patterns of rare and abundant marine microbial eukaryotes. *Curr Biol* **24**: 813-821.

Lopez-Garcia, P., Gaill, F., and Moreira, D. (2002) Wide bacterial diversity associated with tubes of the vent worm Riftia pachyptila. *Environ Microbiol* **4**: 204-215.

López-García, P., Gaill, F., and Moreira, D. (2002) Wide bacterial diversity associated with tubes of the vent worm Riftia pachyptila. *Environmental Microbiology* **4**: 204-215.

López-García, P., Vereshchaka, A., and Moreira, D. (2007) Eukaryotic diversity associated with carbonates and fluid-seawater interface in Lost City hydrothermal field. *Environ Microbiol* **9**: 546-554.

López-García, P., Rodríguez-Valera, F., Pedrós-Alió, C., and Moreira, D. (2001) Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton. *Nature* **409**: 603-607.

López-García, P., Philippe, H., Gail, F., and Moreira, D. (2003) Autochthonous eukaryotic diversity in hydrothermal sediment and experimental microcolonizers at the Mid-Atlantic Ridge. *Proceedings of the National Academy of Sciences* **100**: 697-702.

Lovell, D., Pawlowsky-Glahn, V., Egozcue, J.J., Marguerat, S., and Bahler, J. (2015) Proportionality: a valid alternative to correlation for relative data. *PLoS Comput Biol* **11**: e1004075.

Lupton, J., Butterfield, D., Lilley, M., Evans, L., Nakamura, K.I., Chadwick, W. et al. (2006) Submarine venting of liquid carbon dioxide on a Mariana Arc volcano. *Geochemistry Geophysics Geosystems* **7**: -.

Martín-Fernández, J.A., Palarea-Albaladejo, J., and Olea, R.A. (2011) Dealing with Zeros. In *Compositional Data Analysis*. Pawlowsky-Glahn, V., and Buccianti, A. (eds), pp. 43-58.

Martinez, F., Fryer, P., and Becker, N. (2000) Geophysical characteristics of the southern Mariana Trough, 11 degrees 50 ' N-13 degrees 40 ' N. *Journal of Geophysical Research-Solid Earth* **105**: 16591-16607.

Massana, R., Guillou, L., Díez, B., and Pedrós-Alió, C. (2002) Unveiling the organisms behind novel eukaryotic ribosomal DNA sequences from the ocean. *Applied and Environmental Microbiology* **68**: 4554-4558.

Massana, R., Castresana, J., Balague, V., Guillou, L., Romari, K., Groisillier, A. et al. (2004) Phylogenetic and ecological analysis of novel marine stramenopiles. *Appl Environ Microbiol* **70**: 3528-3534.

McMurtry, G.M., Sedwick, P.N., Fryer, P., VonderHaar, D.L., and Yeh, H.W. (1993) Unusual geochemistry of hydrothermal vents on submarine arc volcanoes: Kasuga Seamounts, Northern Mariana Arc. *Earth and Planetary Science Letters* **114**: 517-528.

McNichol, J., Stryhanyuk, H., Sylva, S.P., Thomas, F., Musat, N., Seewald, J.S., and Sievert, S.M. (2018) Primary productivity below the seafloor at deep-sea hot springs. *Proc Natl Acad Sci U S A* **115**: 6756-6761.

Medlin, L., Elwood, H.J., Stickel, S., and Sogin, M.L. (1988) The characterization of enzymatically amplified eukaryotic 16S-like rRNA-coding regions. *Gene* **71**: 491-499.

Meier, D.V., Bach, W., Girguis, P.R., Gruber-Vodicka, H.R., Reeves, E.P., Richter, M. et al. (2016) Heterotrophic Proteobacteria in the vicinity of diffuse hydrothermal venting. *Environ Microbiol* **18**: 4348-4368.

Meng, J., Xu, J., Qin, D., He, Y., Xiao, X., and Wang, F. (2014) Genetic and functional properties of uncultivated MCG archaea assessed by metagenome and gene expression analyses. *ISME J* **8**: 650-659.

Meyer, J.L., Akerman, N.H., Proskurowski, G., and Huber, J.A. (2013) Microbiological characterization of post-eruption "snowblower" vents at Axial Seamount Juan de Fuca Ridge. *Frontiers in Microbiology* **4**.

Moon-van der Staay, S.Y., De Wachter, R., and Vaulot, D. (2001) Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity. *Nature* **409**: 607-610.

Moreira, D., and López-García, P. (2003) Are hydrothermal vents oases for parasitic protists? *Trends in Parasitology* **19**: 556-558.

Moriya, M., Nakayama, T., and Inouye, I. (2000) Ultrastructure and 18S rDNA sequence analysis of Wobblia lunata gen. et an. nov., a new heterotrophic flagellate (Stramenopiles, Incertae sedis). *Protist* **151**: 41-55.

Murdock, S.A., and Juniper, S.K. (2019) Hydrothermal vent protistan distribution along the Mariana arc suggests vent endemics may be rare and novel. *Environ Microbiol* **21**: 3796-3815.

Murdock, S.A., Tunnicliffe, V., Boschen-Rose, R.E., and Juniper, S.K. (2021) Emergent “core communities” of microbes, meiofauna and macrofauna at hydrothermal vents. *ISME Communications* **1**: 27.

Nadkarni, M.A., Martin, F.E., Jacques, N.A., and Hunter, N. (2002) Determination of bacterial load by real-time PCR using a broad-range (universal) probe and primers set. *Microbiology* **148**: 257-266.

Nagahama, T., Hamamoto, M., Nakase, T., Takami, H., and Horikoshi, K. (2001) Distribution and identification of red yeasts in deep-sea environments around the northwest Pacific Ocean. *Antonie Van Leeuwenhoek International Journal of General and Molecular Microbiology* **80**: 101-110.

Nagano, Y., and Nagahama, T. (2012) Fungal diversity in deep-sea extreme environments. *Fungal Ecology* **5**: 463-471.

Nakagawa, S., Takai, K., Inagaki, F., Hirayama, H., Nunoura, T., Horikoshi, K., and Sako, Y. (2005) Distribution, phylogenetic diversity and physiological characteristics of epsilon-Proteobacteria in a deep-sea hydrothermal field. *Environ Microbiol* **7**: 1619-1632.

Nakamura, K., Toki, T., Mochizuki, N., Asada, M., Ishibashi, J., Nogi, Y. et al. (2013) Discovery of a new hydrothermal vent based on an underwater, high-resolution geophysical survey. *Deep-Sea Research Part I-Oceanographic Research Papers* **74**: 1-10.

Navarri, M., Jegou, C., Meslet-Cladiere, L., Brillet, B., Barbier, G., Burgaud, G., and Fleury, Y. (2016) Deep Subseafloor Fungi as an Untapped Reservoir of Amphipathic Antimicrobial Compounds. *Mar Drugs* **14**.

Oksanen, J., Guillaume Blanchet, F., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B. et al. (2015). vegan: Community Ecology Package. R package version 2.2-1. . URL http://CRAN.R-project.org/package=vegan.

Olins, H.C., Rogers, D.R., Preston, C., Ussler, W., 3rd, Pargett, D., Jensen, S. et al. (2017) Co-registered Geochemistry and Metatranscriptomics Reveal Unexpected Distributions of Microbial Activity within a Hydrothermal Vent Field. *Front Microbiol* **8**: 1042.

Opatkiewicz, A.D., Butterfield, D.A., and Baross, J.A. (2009) Individual hydrothermal vents at Axial Seamount harbor distinct subseafloor microbial communities. *FEMS Microbiol Ecol* **70**: 413-424.

Orcutt, B.N., Sylvan, J.B., Knab, N.J., and Edwards, K.J. (2011) Microbial ecology of the dark ocean above, at, and below the seafloor. *Microbiol Mol Biol Rev* **75**: 361-422.

Orcutt, B.N., Bradley, J.A., Brazelton, W.J., Estes, E.R., Goordial, J.M., Huber, J.A. et al. (2020) Impacts of deep-sea mining on microbial ecosystem services. *Limnology and Oceanography* **n/a**.

Orsi, W., Song, Y.C., Hallam, S., and Edgcomb, V. (2012) Effect of oxygen minimum zone formation on communities of marine protists. *Isme Journal* **6**: 1586-1601.

Pagé, A., Juniper, S.K., Olagnon, M., Alain, K., Desrosiers, G., Quérellou, J., and Cambon-Bonavita, M.A. (2004) Microbial diversity associated with a *Paralvinella sulfincola* tube and the adjacent substratum on an active deep-sea vent chimney *Geobiology* **2**: 225-238.

Palarea-Albaladejo, J., and Martin-Fernandez, J.A. (2015) zCompositions - R Package for multivariate imputation of left-censored data under a compositional approach. *Chemometrics and Intelligent Laboratory Systems* **143**: 85-96.

Passarelli, C., Olivier, F., Paterson, D.M., and Hubas, C. (2012) Impacts of biogenic structures on benthic assemblages: microbes, meiofauna, macrofauna and related ecosystem functions. *Marine Ecology Progress Series* **465**: 85-97.

Pasulka, A., Hu, S.K., Countway, P.D., Coyne, K.J., Cary, S.C., Heidelberg, K.B., and Caron, D.A. (2019) SSU-rRNA Gene Sequencing Survey of Benthic Microbial Eukaryotes from Guaymas Basin Hydrothermal Vent. *Journal of Eukaryotic Microbiology* **66**: 637-653.

Patterson, D.J. (1989) Stramenopiles - Chromophytes from a Protistan Perspective. *Chromophyte Algae* **38**: 357-379.

Patterson, D.J. (1990) Jakoba-Libera (Ruinen, 1938), a Heterotrophic Flagellate from Deep Oceanic Sediments. *Journal of the Marine Biological Association of the United Kingdom* **70**: 381-393.

Pawlowski, J., and Burki, F. (2009) Untangling the phylogeny of amoeboid protists. *J Eukaryot Microbiol* **56**: 16-25.

Pearson, K. (1897) Mathematical contributions to the theory of evolution — on a form of spurious correlation which may arise when indices are used in the measurement of organs. *Proceedings of the Royal Society of London* **60**: 489-498.

Pereira, M.B., Wallroth, M., Jonsson, V., and Kristiansson, E. (2018) Comparison of normalization methods for the analysis of metagenomic gene abundance data. *BMC Genomics* **19**: 274.

Perner, M., Bach, W., Hentscher, M., Koschinsky, A., Garbe-Schonberg, D., Streit, W.R., and Strauss, H. (2009) Short-term microbial and physico-chemical variability in low-temperature hydrothermal fluids near 5 degrees S on the Mid-Atlantic Ridge. *Environ Microbiol* **11**: 2526-2541.

Pernice, M.C., Giner, C.R., Logares, R., Perera-Bel, J., Acinas, S.G., Duarte, C.M. et al. (2016) Large variability of bathypelagic microbial eukaryotic communities across the world's oceans. *ISME J* **10**: 945-958.

Pernice, M.C., Forn, I., Gomes, A., Lara, E., Alonso-Saez, L., Arrieta, J.M. et al. (2015) Global abundance of planktonic heterotrophic protists in the deep ocean. *Isme Journal* **9**: 782-792.

Polz, M.F., and Cavanaugh, C.M. (1995) Dominance of One Bacterial Phylotype at a Mid-Atlantic Ridge Hydrothermal Vent Site. *Proceedings of the National Academy of Sciences of the United States of America* **92**: 7232-7236.

Pomeroy, L.R. (1974) The Ocean's Food Web, A Changing Paradigm. *BioScience* **24**: 499-504.

Prieur, D., Chamroux, S., Durand, P., Erauso, G., Fera, P., Jeanthon, C. et al. (1990) Metabolic Diversity in Epibiotic Microflora Associated with the Pompeii Worms Alvinella-Pompejana and Alvinella-Caudata (Polychaetae, Annelida) from Deep-Sea Hydrothermal Vents. *Marine Biology* **106**: 361-367.

Prokopowich, C.D., Gregory, T.R., and Crease, T.J. (2003) The correlation between rDNA copy number and genome size in eukaryotes. *Genome* **46**: 48-50.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P. et al. (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **41**: D590-596.

Quinn, T.P., Richardson, M.F., Lovell, D., and Crowley, T.M. (2017) propr: An R-package for Identifying Proportionally Abundant Features Using Compositional Data Analysis. *Sci Rep* **7**: 16252.

Quinn, T.P., Erb, I., Gloor, G., Notredame, C., Richardson, M.F., and Crowley, T.M. (2019) A field guide for the compositional analysis of any-omics data. *Gigascience* **8**.

Raghukumar, S. (2002) Ecology of the marine protists, the Labyrinthulomycetes (Thraustochytrids and Labyrinthulids). *European Journal of Protistology* **38**: 127-145.

Raghukumar, S. (2017) The Marine Environment and the Role of Fungi. In *Fungi in Coastal and Oceanic Marine Ecosystems: Marine Fungi*. Raghukumar, S. (ed). Cham: Springer International Publishing, pp. 17-38.

Reimann, F., and Schrage, M. (1978) The mucus-trap hypothesis on feeding of aquatic nematodes and implications for biodegradation and sediment texture. *Oecologia* **34**: 75-88.

Reveillaud, J., Reddington, E., McDermott, J., Algar, C., Meyer, J.L., Sylva, S. et al. (2016) Subseafloor microbial communities in hydrogen-rich vent fluids from hydrothermal systems along the Mid-Cayman Rise. *Environ Microbiol* **18**: 1970-1987.

Reysenbach, A.L., and Shock, E. (2002) Merging genomes with geochemistry in hydrothermal ecosystems. *Science* **296**: 1077-1082.

Reysenbach, A.L., Banta, A.B., Boone, D.R., Cary, S.C., and Luther, G.W. (2000) Biogeochemistry - Microbial essentials at hydrothermal vents. *Nature* **404**: 835-835.

Richards, T.A., Jones, M.D., Leonard, G., and Bass, D. (2012) Marine fungi: their ecology and molecular diversity. *Ann Rev Mar Sci* **4**: 495-522.

Riesenfeld, C.S., Schloss, P.D., and Handelsman, J. (2004) Metagenomics: genomic analysis of microbial communities. *Annu Rev Genet* **38**: 525-552.

Rincon-Tomas, B., Gonzalez, F.J., Somoza, L., Sauter, K., Madureira, P., Medialdea, T. et al. (2020) Siboglinidae Tubes as an Additional Niche for Microbial Communities in the Gulf of Cadiz-A Microscopical Appraisal. *Microorganisms* **8**.

Rinke, C., Lee, J., Nath, N., Goudeau, D., Thompson, B., Poulton, N. et al. (2014) Obtaining genomes from uncultivated environmental microorganisms using FACS-based single-cell genomics. *Nat Protoc* **9**: 1038-1048.

Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N.N., Anderson, I.J., Cheng, J.F. et al. (2013) Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499**: 431-437.

Röttjers, L., and Faust, K. (2018) From hairballs to hypotheses-biological insights from microbial networks. *FEMS Microbiol Rev* **42**: 761-780.

Rousk, J., and Bengtson, P. (2014) Microbial regulation of global biogeochemical cycles. *Front Microbiol* **5**: 103.

Sarrazin, J., and Juniper, S.K. (1999) Biological characteristics of a hydrothermal edifice mosaic community. *Marine Ecology Progress Series* **185**: 1-19.

Sarrazin, J., Juniper, S.K., Massoth, G., and Legendre, P. (1999) Physical and chemical factors influencing species distributions on hydrothermal sulfide edifices of the Juan de Fuca Ridge, northeast Pacific. *Marine Ecology Progress Series* **190**: 89-112.

Sarrazin, J., Legendre, P., de Busserolles, F., Fabri, M.C., Guilini, K., Ivanenko, V.N. et al. (2015) Biodiversity patterns, environmental drivers and indicator species on a high-temperature hydrothermal edifice, Mid-Atlantic Ridge. *Deep-Sea Research Part Ii-Topical Studies in Oceanography* **121**: 177-192.

Sauvadet, A.L., Gobet, A., and Guillou, L. (2010) Comparative analysis between protist communities from the deep-sea pelagic ecosystem and specific deep hydrothermal habitats. *Environ Microbiol* **12**: 2946-2964.

Schloss, P.D., and Handelsman, J. (2005) Metagenomics for studying unculturable microorganisms: cutting the Gordian knot. *Genome Biol* **6**: 229.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B. et al. (2009) Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Appl Environ Microbiol* **75**: 7537-7541.

Schnur, S.R., Chadwick, W.W., Embley, R.W., Ferrini, V.L., de Ronde, C.E.J., Cashman, K.V. et al. (2017) A decade of volcanic construction and destruction at the summit of NW Rota-1 seamount: 2004-2014. *Journal of Geophysical Research-Solid Earth* **122**: 1558-1584.

Schoenle, A., Hohlfeld, M., Rosse, M., Filz, P., Wylezich, C., Nitsche, F., and Arndt, H. (2020) Global comparison of bicosoecid Cafeteria-like flagellates from the deep ocean and surface waters, with reorganization of the family Cafeteriaceae. *Eur J Protistol* **73**: 125665.

Schratzberger, M., and Ingels, J. (2018) Meiofauna matters: The roles of meiofauna in benthic ecosystems. *Journal of Experimental Marine Biology and Ecology* **502**: 12-25.

Seenivasan, R., Sausen, N., Medlin, L.K., and Melkonian, M. (2013) Picomonas judraskeda Gen. Et Sp Nov.: The First Identified Member of the Picozoa Phylum Nov., a Widespread Group of Picoeukaryotes, Formerly Known as 'Picobiliphytes'. *Plos One* **8**.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D. et al. (2003) Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research* **13**: 2498-2504.

Sherman, K., Sissenwine, M., Christensen, V., Duda, A., Hempel, G., Ibe, C. et al. (2005) A global movement toward an ecosystem approach to management of marine resources. *Marine Ecology Progress Series* **300**: 275-279.

Shi, Y., Pan, C., Auckloo, B.N., Chen, X., Chen, C.A., Wang, K. et al. (2017) Stress-driven discovery of a cryptic antibiotic produced by Streptomyces sp. WU20 from Kueishantao hydrothermal vent with an integrated metabolomics strategy. *Appl Microbiol Biotechnol* **101**: 1395-1408.

Sievert, S.M., and Vetriani, C. (2012) Chemoautotrophy at deep-sea vents: Past, present, and future. In *Oceanography*, pp. 218-233.

Small, E.B., and Gross, M.E. (1985) Preliminary observations of protistan organisms, especially ciliates, from the 21°N hydrothermal vent site. *Biol Soc Wash Bull* **6**: 401-410.

Sogin, M.L., Morrison, H.G., Huber, J.A., Mark Welch, D., Huse, S.M., Neal, P.R. et al. (2006) Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proc Natl Acad Sci U S A* **103**: 12115-12120.

Song, W., Wemheuer, B., Zhang, S., Steensen, K., and Thomas, T. (2019) MetaCHIP: community-level horizontal gene transfer identification through the combination of best-match and phylogenetic approaches. *Microbiome* **7**: 36.

Spang, A., Saw, J.H., Jorgensen, S.L., Zaremba-Niedzwiedzka, K., Martijn, J., Lind, A.E. et al. (2015) Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* **521**: 173-179.

Steele, J.A., Countway, P.D., Xia, L., Vigil, P.D., Beman, J.M., Kim, D.Y. et al. (2011) Marine bacterial, archaeal and protistan association networks reveal ecological linkages. *Isme Journal* **5**: 1414-1425.

Stoddard, S.F., Smith, B.J., Hein, R., Roller, B.R., and Schmidt, T.M. (2015) rrnDB: improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development. *Nucleic Acids Res* **43**: D593-598.

Stoeck, T., Hayward, B., Taylor, G.T., Varela, R., and Epstein, S.S. (2006) A Multiple PCR-primer Approach to Access the Microeukaryotic Diversity in Environmental Samples. *Protist* **157**: 31-43.

Stoeck, T., Bass, D., Nebel, M., Christen, R., Jones, M.D., Breiner, H.W., and Richards, T.A. (2010) Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Mol Ecol* **19 Suppl 1**: 21-31.

Stoecker, D.K., Johnson, M.D., deVargas, C., and Not, F. (2009) Acquired phototrophy in aquatic protists. *Aquatic Microbial Ecology* **57**: 279-310.

Stokke, R., Dahle, H., Roalkvam, I., Wissuwa, J., Daae, F.L., Tooming-Klunderud, A. et al. (2015) Functional interactions among filamentous Epsilonproteobacteria and Bacteroidetes in a deep-sea hydrothermal vent biofilm. *Environ Microbiol* **17**: 4063-4077.

Takishita, K., Miyake, H., Kawato, M., and Maruyama, T. (2005) Genetic diversity of microbial eukaryotes in anoxic sediment around fumaroles on a submarine caldera floor based on the small-subunit rDNA phylogeny. *Extremophiles* **9**: 185-196.

Takishita, K., Yubuki, N., Kakizoe, N., Inagaki, Y., and Maruyama, T. (2007) Diversity of microbial eukaryotes in sediment at a deep-sea methane cold seep: surveys of ribosomal DNA libraries from raw sediment samples and two enrichment cultures. *Extremophiles* **11**: 563-576.

Thompson, K.F., Miller, K.A., Currie, D., Johnston, P., and Santillo, D. (2018) Seabed Mining and Approaches to Governance of the Deep Seabed. *Frontiers in Marine Science* **5**.

Thompson, L.R., Sanders, J.G., McDonald, D., Amir, A., Ladau, J., Locey, K.J. et al. (2017) A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* **551**: 457-463.

Thornburg, C.C., Zabriskie, T.M., and McPhail, K.L. (2010) Deep-Sea Hydrothermal Vents: Potential Hot Spots for Natural Products Discovery? *Journal of Natural Products* **73**: 489-499.

Trembath-Reichert, E., Butterfield, D.A., and Huber, J.A. (2019) Active subseafloor microbial communities from Mariana back-arc venting fluids share metabolic strategies across different thermal niches and taxa. *ISME J* **13**: 2264-2279.

Tsurumi, M., and Tunnicliffe, V. (2003) Tubeworm-associated communities at hydrothermal vents on the Juan de Fuca Ridge, northeast Pacific. *Deep-Sea Research Part I-Oceanographic Research Papers* **50**: 611-629.

Tunnicliffe, V., and Cordes, E.E. (2020) The tubeworm forests of hydrothermal vents and cold seeps. In *Perspectives on the Marine Animal Forests of the World* Rossi, S., and Bramanti, L. (eds): Springer, pp. 147-192.

Tunnicliffe, V., Juniper, S., and de Burgh, M.E. (1985) The hydrothermal vent community in Axial Seamount, Juan de Fuca Ridge. In *Hydrothermal vents of the eastern Pacific: an overview*. Jones, M.L. (ed): Bulletin of the Biological Society of Washington, pp. 434-464.

Tunnicliffe, V., Germain, C.S., and Hilario, A. (2014) Phenotypic Variation and Fitness in a Metapopulation of Tubeworms (Ridgeia piscesae Jones) at Hydrothermal Vents. *Plos One* **9**.

Urcuyo, I.A., Bergquist, D.C., MacDonald, I.R., VanHorn, M., and Fisher, C.R. (2007) Growth and longevity of the tubeworm Ridgeia piscesae in the variable diffuse flow habitats of the Juan de Fuca Ridge. *Marine Ecology Progress Series* **344**: 143-157.

Urich, T., Lanzen, A., Stokke, R., Pedersen, R.B., Bayer, C., Thorseth, I.H. et al. (2014) Microbial community structure and functioning in marine sediments associated with diffuse hydrothermal venting assessed by integrated meta-omics. *Environmental Microbiology* **16**: 2699-2710.

v. Wintzingerode, F., Göbel, U.B., and Stackebrandt, E. (1997) Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiology Reviews* **21**: 213-229.

van Dongen, S. (2000). Graph clustering by flow simulation. URL http://dspace.library.uu.nl/handle/1874/848

Van Dover, C., and Fry, B. (1989) Stable isotopic compositions of hydrothermal vent organisms. *Marine Biology* **102**: 257-263.

Van Dover, C.L. (2002) Trophic relationships among invertebrates at the Kairei hydrothermal vent field (Central Indian Ridge). *Marine Biology* **141**: 761-772.

Vargas-Gastelum, L., and Riquelme, M. (2020) The Mycobiota of the Deep Sea: What Omics Can Offer. *Life (Basel)* **10**.

Wei, T., and Simko, V. (2017). R package "corrplot": Visualization of a Correlation Matrix (Version 0.84). Available from https://github.com/taiyun/corrplot. URL https://github.com/taiyun/corrplot

West, P.T., Probst, A.J., Grigoriev, I.V., Thomas, B.C., and Banfield, J.F. (2018) Genome-reconstruction for eukaryotes from complex natural microbial communities. *Genome Res* **28**: 569-580.

Williams, R.J., Howe, A., and Hofmockel, K.S. (2014) Demonstrating microbial co-occurrence pattern analyses within and between ecosystems. *Frontiers in Microbiology* **5**: 1-10.

Wisz, M.S., Pottier, J., Kissling, W.D., Pellissier, L., Lenoir, J., Damgaard, C.F. et al. (2013) The role of biotic interactions in shaping distributions and realised assemblages of species: implications for species distribution modelling. *Biol Rev Camb Philos Soc* **88**: 15-30.

Worden, A.Z., Follows, M.J., Giovannoni, S.J., Wilken, S., Zimmerman, A.E., and Keeling, P.J. (2015) Environmental science. Rethinking the marine carbon cycle: factoring in the multifarious lifestyles of microbes. *Science* **347**: 1257594.

Xie, W., Wang, F., Guo, L., Chen, Z., Sievert, S.M., Meng, J. et al. (2011) Comparative metagenomics of microbial communities inhabiting deep-sea hydrothermal vent chimneys with contrasting chemistries. *ISME J* **5**: 414-426.

Xu, C., Sun, X., Jin, M., and Zhang, X. (2017) A Novel Benzoquinone Compound Isolated from Deep-Sea Hydrothermal Vent Triggers Apoptosis of Tumor Cells. *Mar Drugs* **15**.

Yamanaka, T., Shimamura, S., Nagashio, H., Yamagami, S., Onishi, Y., Hyodo, A. et al. (2015) A Compilation of the Stable Isotopic Compositions of Carbon, Nitrogen, and Sulfur in Soft Body Parts of Animals Collected from Deep-Sea Hydrothermal Vent and Methane Seep Fields: Variations in Energy Source and Importance of Subsurface Microbial Processes in the Sediment-Hosted Systems. In *Subseafloor Biosphere Linked to Hydrothermal Systems*. Ishibashi, J., Okino, K., and Sunamura, M. (eds). Tokyo, Japan: Springer Open, pp. 105-129.

Young, V.B. (2017) The role of the microbiome in human health and disease: an introduction for clinicians. *BMJ* **356**: j831.

Zaikova, E., Walsh, D.A., Stilwell, C.P., Mohn, W.W., Tortell, P.D., and Hallam, S.J. (2010) Microbial community dynamics in a seasonally anoxic fjord: Saanich Inlet, British Columbia. *Environ Microbiol* **12**: 172-191.

Zain Ul Arifeen, M., Ma, Y.-N., Xue, Y.-R., and Liu, C.-H. (2019) Deep-Sea Fungi Could Be the New Arsenal for Bioactive Molecules. *Marine drugs* **18**: 9.

Zbinden, M., Shillito, B., Le Bris, N., de Montlaur, C.D., Roussel, E., Guyot, F. et al. (2008) New insigths on the metabolic diversity among the epibiotic microbial communitiy of the hydrothermal shrimp Rimicaris exoculata. *Journal of Experimental Marine Biology and Ecology* **359**: 131-140.

Zhu, F., Massana, R., Not, F., Marie, D., and Vaulot, D. (2005) Mapping of picoeucaryotes in marine ecosystems with quantitative PCR of the 18S rRNA gene. *Fems Microbiology Ecology* **52**: 79-92.