The Pennsylvania State University

The Graduate School

Department of Geosciences

A MOLECULAR ANALYSIS OF SUBSURFACE MICROBIAL COMMUNITIES ACROSS A HYDROTHERMAL GRADIENT IN OKINAWA TROUGH SEDIMENTS

A Dissertation in

Geosciences and Biogeochemistry

by

Leah D. Brandt

Submitted in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

December 2016

The dissertation of Leah D. Brandt was reviewed and approved* by the following:

Christopher H. House
Professor of Geosciences
Dissertation Advisor
Co-Chair of Committee

Katherine H. Freeman
Evan Pugh Professor of Geosciences
Biogeochemistry Program Co-Advisor
Co-Chair of Committee

Jennifer H. Macalady
Associate Professor of Geosciences

Ming Tien
Professor of Biochemistry
Department of Biochemistry and Molecular Biology

Demian Saffer
Professor of Geosciences
Associate Head for Graduate Programs and Research in Geosciences

*Signatures are on file in the Graduate School

ABSTRACT

Decades ago, life in the deep subseafloor was assumed to be non-existent. At thousands of meters under the surface and decoupled from the photic zone, the deep marine world is a hostile environment. However, the discovery of life thriving around deep-sea hydrothermal vent systems revolutionized our perception of the extent and tenacity of life on Earth and set in motion a movement to understand life at the seafloor and its significant in global biogeochemical and nutrient cycling. Seafloor sediments host an incredible diversity of microbial life, and much interest has sought to understand the spatial and stratigraphic extent of the biosphere, and also taxonomic and functional capacities of such resilient organisms. This dissertation represents a series of studies centered around these concepts – I analyze samples from a sediment profile exposed to a hydrothermal gradient as a scaled-down proxy for how microbial life may exist and/or adapt to conditions as they become buried deeper into the subsurface.

I have focused on the application of culture-independent, molecular methods to understand whether the taxonomic and functional data reflect changes through this temperature gradient in support of a more temperature-adapted microbial community. In chapter 2, I examined the microbial community composition at approximately meter intervals by analyzing the taxonomically specific 16S rRNA marker gene. We presumed the biosphere to be restricted to only the upper 15 m, based on phylotype vetting and decreased sequencing recovery below. However, we observed a significant proportion of archaeal sequences throughout the 15 m, with a particularly high abundance in the deepest 15 m horizon. An in-depth look at the taxa at 15 m indicates an appearance of an uncultured, high-temperature taxon here, and an abundance of a thermophilic, methane-oxidizing archaeon, which suggests thermophilic niche at this particular temperature/depth regime. The exciting results from Chapter 2 were the motivation for a continued metagenomics analysis of select samples in the same sediment profile for Chapter 3.

Probing through total genomic DNA from six samples, I found evidence for temperature-dependent trends through the detection of genes of specific proteins involved in thermal processes, and attempted to correlate these genes with a taxonomic identity. I found that the deepest, hottest sample encompassed organisms from both thermophilic and hyperthermophilic temperature regimes. The dichotomy reflected between the existence of two temperature-specific niches implies that, due to the dynamic nature of the hydrothermal vent system, the deepest horizon may be undergoing a transition in temperature, thus, microbial community. Lastly, Chapter 4 was intended to provide a dataset from extractable RNA in order to distinguish representatives of the active microbial population from those represented from extant DNA; however, many challenges encountered in extraction and sequencing yield have restricted the dataset and ability to make reliable interpretations.

Considering the current state of knowledge in the marine subsurface due to its challenges in sampling, low biomass yield, and diversity distantly related to what is known from surface life, this work herein contributes greatly to our understanding of microbial biogeography in terms of temperature constraints in marine subsurface sediments. The challenges in phylotype vetting and need for quality controls speak to the degree of complexities in performing and interpreting molecular analyses from subseafloor sediments. Moreover, I have also produced significant datasets, both 16S rRNA gene and metagenomics, that can continue to be used for future investigations and comparisons of microbial life in the marine seafloor.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

This dissertation would not have been possible without the constant support and encouragement of my dissertation advisor, Christopher House, who always saw the silver lining in extraction failures, contamination issues, and puzzling results. Thank you, Chris, for teaching me to grow as an independent scientist – to think in a publishable-quality mindset – and to stay patient, positive, determined and excited despite all laboratory frustrations. It can only get easier from here! Also, to my other committee members Kate Freeman, Jenn Macalady, and Ming Tien – you have reminded to stay big-picture focused, which I have found crucial in communicating as a scientist. I have appreciated your honest feedback and suggestions. Kate and Jenn, your classes and teaching have been inspirational, and have formed my foundation in biogeochemistry. Additional "thank-yous" to Jen Biddle, Mandi Martino, and Trinity Hamilton for their expertise and advice whenever I needed a sounding board. Much of my laboratory work could not have been completed without the technical support of Zhidan Zhang, Irene Schaperdoth, and Denny Walitzer.

Of course, I need to thank my fellow graduate student friends – Stamatina, Ashlee, Rosie, Emily, Elizabeth, and Paul – for understanding the ebbs, flows, ups, and downs of graduate school, but also maintaining a healthy, fit and fun atmosphere with me. But most importantly, I need to thank my husband, Joe Zinsky, and my parents, Christine and Martin Brandt, for their constant love, encouragement, and humor that have kept me motivated, excited, and proud through this journey. You have been wonderful examples of sacrifice, hard work, dedication, and family for me to live by.

Chapter 1

Introduction

The Subsurface Beneath the Seafloor

The Earth's deep marine biosphere represents a frontier for investigating the extent, distribution, and perseverance of life. Buried beneath meters to hundreds of meters of sediments, microbial life is far removed from sunlight, strong oxidants, abundant organic matter, and other nutrients circulating in seawater (Orcutt *et al.,* 2011). Yet, there is evidence of microbial life taking advantage of seemingly limiting and unfavorable conditions. Although the sedimentary subsurface microbial abundance is estimated to represent approximately 1-10% of Earth's total living biomass (Kallmeyer *et al.*, 2012), the importance of this ecosystem bridges the biological and geological element cycles, where microbial life at this interface is directly related to whether elements are sequestered over geologic time, or returned to the ocean as active components of biological, chemical, and climatic cycling (Hinrichs and Inagaki, 2012). The discovery of a vast subsurface community has channeled research efforts regarding energetic requirements and limits for life, the contribution of the subsurface on global biogeochemical cycling, and the potential for life in subsurface environments on other planetary bodies.

Marine sediments are the result of accumulating particles that sink to the seafloor from the overlying water column (Orcutt *et al.*, 2011). As in the terrestrial realm, not all sediments are compositionally the same or are affected by the same geological and geochemical parameters. Thickness of sediments varies from relatively thin near newly formed crust at mid-ocean ridges and beneath low productivity zones (*e.g.* South Pacific Gyre), to kilometers-thick at trenches and highly productive continental margins (Orcutt *et al.*, 2011). The chemical composition of

sediments is also variable, depending on the origin of the deposited material, such as "rain" down of biological oozes, terrestrial matter delivered by rivers, or deposits from vent-derived minerals (Orcutt *et al.*, 2011). This spatial geographic and geochemical diversity of the seafloor is responsible for creating a range of habitats and niches.

One of the more enigmatic relationships in marine subsurface research is the stratigraphic relationship of bioavailable organic matter and inorganic compounds with community composition through a sediment profile. In an early study looking at suboxic diagenesis of organic matter in the top 0-80 cm of pelagic sediments, pore water profiles indicated that oxidants were generally consumed in order of decreasing energy production per mole of organic carbon oxidized (Froelich *et al.*, 1978), suggesting that these distinct redox zones were shaped by organisms with specific metabolic traits (Jorgensen *et al.*, 2012). While this hypothesis has been supported by down-core stratification of specific microbial groups, such as anaerobic methane oxidizers (ANME) (Boetius *et al.*, 2000) or anaerobic ammonium oxidizers (anammox) (Strous *et al.*, 1999), predictions about microbial communities based on geochemical conditions have not been consistent across sedimentary marine environments (Jorgensen *et al.*, 2012). Understanding the stratigraphic variability in geochemical and lithological properties with concomitant changes in the total microbial community structure and the relative abundance of individual taxa is an ongoing area of investigation and remains to be considered a complex relationship (Jorgensen *et al.*, 2012).

Challenges in Studying the Marine Subsurface

There are many factors that make the exploration of subseafloor microbial communities particularly challenging. First, the opportunities to obtain samples from the ocean seafloor are limited by equipped drilling vessels or submersibles. Both sampling options are very expensive

and require an experienced crew to operate and engineer the ship and equipment. For these reasons, there are many regions of the seafloor that remain completely unexplored, and our concept of the global marine subsurface relies on predicting or modeling those regions that are not sampled (Colwell and D'Hondt, 2013, Martino, 2014). In order to penetrate compacted sediments and rock layers within the ocean crust, scientists have relied on the drilling technology, which was not designed for sterile microbiological sampling. Thus, scientists have had to devise ways to ensure that the subsampling of cores has not been subject to external seawater or drilling mud contamination.

Aside from sampling logistics and cost, extremely low rates of cellular activity make laboratory cultivation very difficult (D'Hondt *et al.*, 2002; 2009). Scientists have, therefore, been reliant on culture-independent analysis of nucleic acids to better understand community composition and functional potential within marine sediments. Often, the preliminary extractions and amplifications of environmental DNA and RNA from marine sediments are especially challenging in sediments where biomass densities are extremely low or in sediments with complex or altered clay compositions. Downstream taxonomic and phylogenetic datasets, in particular, rely on amplification of DNA and RNA marker genes via oligonucleotide primersets, which are designed from databases of known sequenced organisms. This can introduce biases in amplification by excluding "novel" lineages or intensifying signals of sequences that are compatible with the primer design (Colwell and D'Hondt, 2013). Within sequence data produced from extractable and amplifiable DNA, many taxa appear to be only distantly related to known representatives from pure cultures or surface environments (Sørensen *et al.*, 2004; Inagaki *et al.*, 2006; Lipp *et al.*, 2008; Fry *et al.*, 2008; Teske and Sørensen, 2008). Hence, when sequences *are* obtained, there are uncertainties and limitations in classifying and inferring metabolic capabilities within the subsurface communities.

Progress in Marine Subsurface Research

Advances in drilling recovery, molecular extraction methodologies, and sequencing technology over the past several decades have yielded high through-put datasets and more streamline analyses to better understand the micro- to macroscale relationships of the marine subsurface ecosystem. Investigative approaches in studying microbial diversity have advanced since the 1980s from laboratory isolations of organisms to analyses of macromolecules directly from environmental samples, as previously mentioned (Olsen *et al*., 1986 and Rappé and Giovannoni, 2003). This progression towards a cultivation-independent methodology has dramatically increased our knowledge of the unculturable community, estimated to represent approximately 99% of all microorganisms, due to the expansion of standard reference libraries (Rappé and Giovannoni, 2003). The transition of DNA sequencing chemistries and technologies from Sanger-based capillary sequencing to massively parallel, high-throughput "next-generation" sequencing has revolutionized ecological science. Large amounts of data, on the order of millions of sequence reads, at a relatively low cost per sequence yield have lead to a better representation of sample diversity (Shokralla *et al*., 2012). With current piqued interests in deep sea drilling, scientists have been able to apply these molecular tools directly to a suite of recovered, deep-sea sediment cores to answer questions regarding microbial populations and processes within the extent of the deep biosphere.

The Ocean Drilling Program (ODP) expedition to the Peru Margin (ODP Leg 201) was a pivotal breakthrough for marine subsurface research that involved concurrent molecular and geochemical analyses along meter-scale vertical sediment profiles (Jørgensen *et al.*, 2006). The comprehensive dataset produced from ODP Leg 201 represents the first synthesis of correlating taxonomic and functional data from DNA and RNA with geochemical interpretations from approximately 87 m of recovered core. Sediment cores recovered from the Peru Margin were

from underlying, highly productive surface waters off the Peru shelf and slope. The profiled microbial ecosystems appeared to be stratified with respect to geochemistry and stimulated at interfaces between seawater sulfate and methane (Jørgensen *et al.*, 2006, Schrenk *et al.*, 2010). Based on porewater chemistry interpretations, several horizons were predicted to be dominant biological methanogenic or methanotrophic zones, yet, known lineages of Archaea involved in these processes were sparse in the taxonomic analyses (Jørgensen *et al.*, 2006). Instead, extractable 16S ribosomal RNA (rRNA) was classified within uncultured archaeal lineages that have cosmopolitan distributions across other marine environments (Biddle *et al.*, 2006). Furthermore, the carbon isotopic composition of both archaeal cells and lipids from these sulfate-methane transition zones (SMTZs) suggested that bulk assimilation of carbon was derived from fossil organic matter, rather than methane. Overall, the results from multiple analytical approaches did not accurately reflect what had been initially hypothesized from initial porewater geochemistry profile.

The results from ODP Leg 201 demonstrate that the relationships between phylogenetic diversity, functional diversity, and geochemical interpretations in subsurface sediments are still poorly constrained. However, this study has brought to light the significance of Archaea in the marine subsurface as well as the potential eco-physiological flexibility of certain widespread, uncultured archaeal lineage. Since ODP Leg 201, deep biosphere research has continued to gain momentum across a range of marine environments, from the ultraoligotrophic sediments in the South Pacific Gyre (IODP 329), to a deep coal bed formation off the Shimokita Peninsula (IODP 337). Many ongoing studies are even focusing on biogeographical trends in specific uncultured archaeal lineages (*e.g.* representatives from the Miscellaneous Crenarchaeotic Group in Kubo *et al.*, 2012) and their potential functional roles in the subsurface (*e.g.* Lloyd *et al.*, 2013). In most cases, the utilization of next-generation DNA sequencing platforms has been a valuable and integral part of advancing our knowledge in marine subsurface ecology.

Temperature and the Extent of the Deep Biosphere

Within sediment profiles, cell concentrations generally decrease with depth, which is attributed to the decreasing porosity of sediments and bioavailability of organic carbon needed to fuel metabolic reactions and cell growth (Parkes *et al.*, 2000). But, the extent and density of the biosphere vary across marine environments (Figure 1-1). One of the major factors limiting microbial distributions to great depths within the subsurface is the increasing temperature during burial (Parkes *et al*., 2000). The upper temperature limits of life for organisms (up to 122°C) have been studied in great detail from energy-rich hydrothermal vent fluids. However, temperature estimates associated with the more energy-limited sedimentary subsurface are derived from studies that indicate biodegradation of petroleum is inactivated at temperatures above 80°C (Wilhelms *et al*., 2001, Head *et al.*, 2003). It has been suggested that this lower temperature fringe, relative to hyperthermophilic isolates, is linked to inadequate requirements to support rapid re-synthesis of essential biomolecules (Head *et al.*, 2003). The study of hyperthermophilic communities in sediments, though, has been limited to the very surface and has not involved extensive taxonomic and metagenomic analyses.

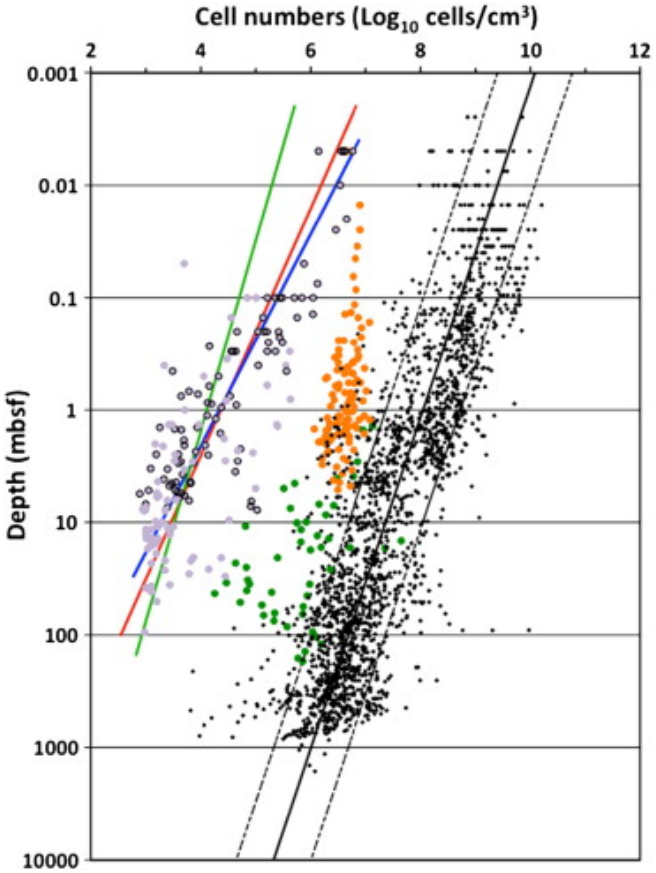

Figure **1-1**: Source: Parkes *et al.*, 2014. Depth (mbsf) distribution of prokaryotic cells in sub-surface sediments at 106 locations, including 17 ODP/IODP Legs (black dots). Orange circles are mud volcano breccia samples, green circles are hydrothermal samples, purples circles are South Pacific Gyre sediments (Kallmeyer *et al.*, 2012).

One of the first studies to investigate a possible hyperthermophilic community in marine sediments comes from the Parkes *et al*, 2000 compilation of ODP Juan de Fuca Ridge sediment cores from 1990s expeditions. Access to sediments that reach temperatures above 100°C would require up to 4 km of recovered sediment core, based on a geothermal gradient of 25°C/km. Juan de Fuca Ridge sediments were appealing in such an investigation because of their large thermal gradients, thus, high temperatures at relatively shallow sediment depth. Parkes *et al.*, 2000 found that cell enumerations declined with depth, but persisted, or even increased, in certain high temperature intervals, which suggested that certain microbial life might occupy a unique high-

temperature subsurface habitat resulting from mixing of recharging seawater and hydrothermal fluid. In an even broader sense, it is quite possible that certain groups of microbes are capable of persisting much deeper into the surface that what has been sampled. Thus, modeling exercises that incorporate microbial extent into the ocean crust should consider the possibility of these extreme communities. While the Parkes *et al.*, 2000 study has extended the concept of the sub-seafloor biosphere to new depth, the techniques used would, today, be considered rudimentary and lacking comprehensive support of a hyperthermophilic subsurface biosphere. An expansion of the Parkes *et al*., 2000 using present-day molecular tools would be a pivotal step in understanding the relationship between 1) the putative temperature limit of life and life in subsurface sediments and 2) taxonomic, functional, and activity changes as a function of subsurface temperature limits.

Much like the Middle Valley sediments of Juan de Fuca Ridge, hydrothermal sediments represent few places in the crust where a similar physical environment (*i.e.* sediments) is in contact with multiple geothermal regimes (Biddle *et al.*, 2012). At the sediment-water interface of most hydrothermal vents, microbial communities change rapidly across intense gradients over a much shorter spatial scale. But, the ability to analyze taxonomically similar microbial populations across a gentle temperature gradient has not been thoroughly studied in microbial ecology. Hydrothermal vent systems are formed at tectonic boundaries and mid-plate hot spots of volcanic activity. Fluids circulating through the crust experience extreme heating and water-rock interactions, resulting in a net gain of thermally energy and gaseous compounds (Orcutt *et al.*, 2011). These evolved hydrothermal fluids, which are chemically reduced in comparison to source seawater, are transported to the surface, where they interact with seawater to often produce precipitated mineral deposits, or “chimneys” (Orcutt *et al*., 2011). Generally, most hydrothermal venting occurs in areas with little sediment cover on young oceanic crust, so sedimented systems, such as Guaymas Basin, Juan de Fuca Ridge, and the Okinawa Backarc basin are more of an

anomaly than the norm. The Okinawa Backarc Basin hydrothermal system, in particular, is one of the few environments sampled that formed as a result of plate subduction, rather than a spreading center. In this PhD dissertation, I communicate the results of a comprehensive study that implements a suite of molecular analyses to answer profound questions about the sedimentary subsurface in the first survey of a hydrothermal vent system in a continental margin setting.

IODP Expedition 331: Okinawa Backarc Basin and the Deep, Hot Biosphere

The Integrated Ocean Drilling Program (IODP) Expedition 331 to the Okinawa backarc basin provided an opportunity to study the biosphere within the sedimentary subsurface of a deep, hot "subvent" system. The Okinawa Backarc Basin was formed as a result of the subduction of the Philippine Sea plate beneath the Eurasian plate. Generally, backarc basins are characteristic features of oceanic convergent plate boundaries. Unsurprisingly, hydrothermal vent networks have been documented here, and several studies have investigated surface microbiology and geochemistry immediately surrounding a few specific vents. In contrast to other backarc basins, the Okinawa Backarc Basin is the only example of a young backarc basin that has developed along a continental margin (Letouzey and Kimura, 1985).

The motivation for sampling the Okinawa Backarc Basin hydrothermal systems is two-fold. Firstly, backarc basins inherently produce a significantly greater abundance of volatile species that are derived from devolatilization of the subducting slab than spreading center hydrothermal systems. Thus, backarc basins have a different composition of hydrothermal fluid and porewater in sediments. For example, the chemistry of hydrothermal fluids collected from active sulfide chimneys in the Okinawa Trough is characterized by higher concentrations of CO2, CH4, NH4, I, and K, and higher alkalinity than those in sediment-free, mid-ocean ridge hydrothermal fluids (Sakai *et al.*, 1990a, 1990b; Gamo *et al.*, 1991; Konno *et al.*, 2006; Takai and

Nakamura, 2001; Kawagucci *et al.*, 2001; Takai *et al.*, 2011). Secondly, IODP Expedition 331 is the first to core an active hydrothermal system within a sediment-filled backarc basin in a continental margin type setting. Most other sedimented hydrothermal vent systems (*e.g.* Guaymas Basin and Juan de Fuca Ridge) are associated with spreading centers. The sediment profiles sampled from the Okinawa Backarc Basin represent an environment that can be compared to sediments from both other margins and hydrothermal vent systems. Furthermore, this is an opportunity to better understand the changes in microbial life associated with sediments along a temperature gradient and also the extent or limits of the deep biosphere. Collectively, these factors suggest a new environment, in terms of sediment matrix, geochemistry, and potentially microbial community composition, which has not yet been considered in deep biosphere or global biogeochemical research.

Iheya North Hydrothermal Vent: Site Descriptions

The IODP Expedition 331 drilling efforts were focused around the active hydrothermal vent site within the Iheya North hydrothermal system (Figure 1-2). Three sites ~100, 450, and 1500 m east and one site ~600 m northwest of the active vent all showed significant differences of hydrothermal inputs from one another. No temperature data were collected from the closest site to the active vent, Site C0013. But, based on the recovered melted acrylide core liner that cored down to 12 mbsf, the temperature, likely, was well over 82°C (onboard testing later found that these core liners began to deform above 82°C) within that core section. Site C0014, approximately 450 m away from the active vent, had an interpolated temperature gradient of 3°C/m (Figure 1-3). The furthest site, C0017, had a temperature gradient of approximately 0.5 °C/m (Figure 1-3). Both temperature profiles from Sites C0014 and C0017 display irregularities suggestive of lateral flow. The concave upward profile at Site C0017 is consistent with recharge

of cold seawater into the system at this site. On the contrary, the temperature and geochemistry profiles indicate that Site C0014 represents a moderate hydrothermal mixing zone in the subsurface. Site C0015 was upslope 600 m from the active vent on a mound, where many colonies of dead deep-sea mussels and pavements of carbonate and/or sulfate crusts were found. The crest of this hill has been hypothesized to represent an enormous methane seepage field in the past.

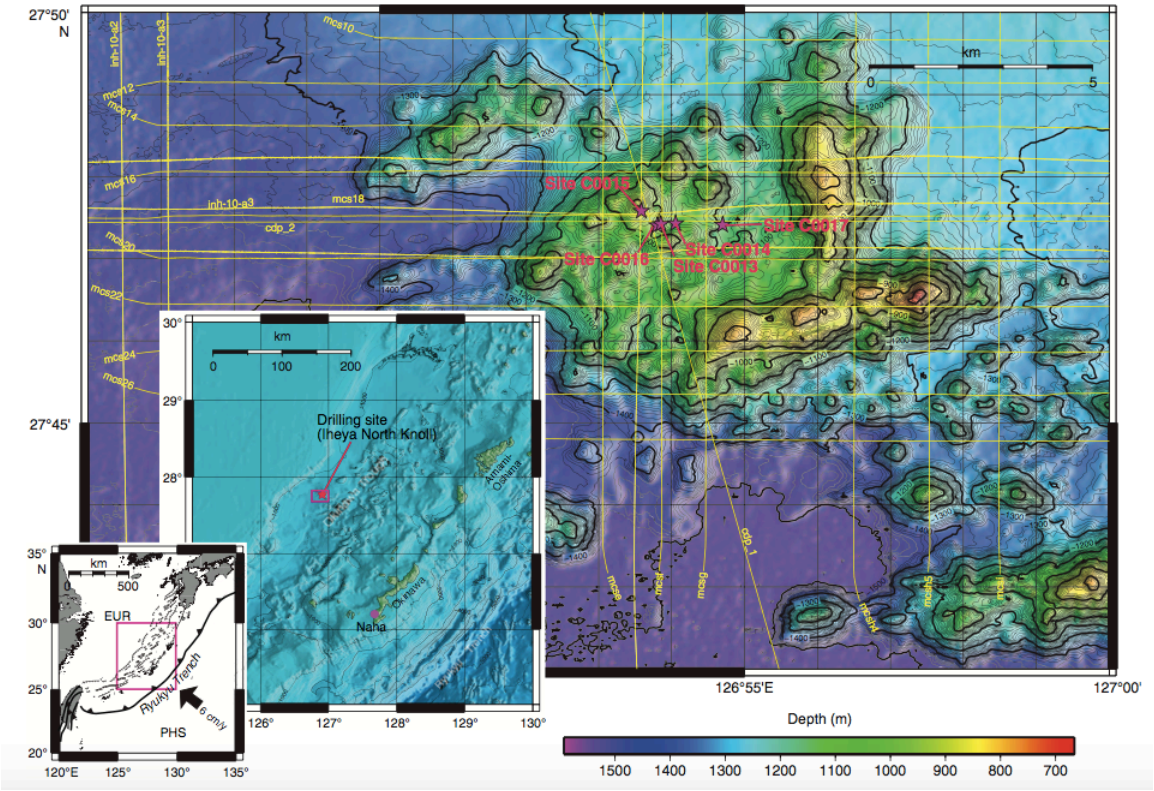

Figure **1-2**: Source: Takai *et al.*, 2011. Area map of Iheya North Knoll showing Sites C0013-C0017 (stars) drilling during Expedition 331. Inserts show the Iheya North Knoll in relation to Okinawa and Okinawa in relation to major tectonic components. EUR = Eurasian plate, PHS = Philippine Sea plate.

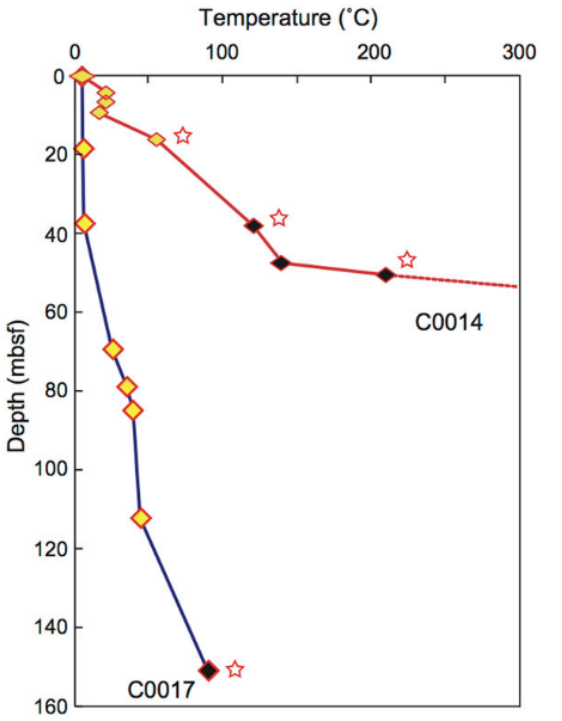

Figure **1-3**: Source: Yanagawa *et al.*, 2013. Depth profile of temperature at Sites C0014 and C0017, as measured by the APCT-3 (yellow) temperature shoe and thermoseal strip taped to the outer surface of the core liner. The stars represent the error associated with the thermoseal strips. Site C0017 is representative of a non-hydrothermal continental margin sediment profile.

Dissertation Outline

In the following sections I discuss a taxonomic and functional profiling of the continental margin-type sediments impacted by the Iheya North Field subsurface hydrothermal system in the Okinawa backarc basin. I first use a commonly practiced approach of comparing the 16S rRNA taxonomic marker gene among Site C0014 with one from Site C0015 to determine general trends in microbial community composition with depth. I also incorporate porewater geochemical data and published IODP Expedition 331 Proceedings results to make interpretations whether the community composition reflects changes along the temperature profile, or the temperature

gradient is too strong or quickly progressing for an adapted biosphere to be established. The subsequent chapter builds off of these results in an analysis of metagenomic data from several depth horizons with the intent to capture relevant metabolic and functional changes through this hydrothermal gradient. I use several data mining approaches to better understand how microbial processes and adaptations are different among several horizons along this transect. The last section was initially intended to use taxonomic information from extractable RNA as a proxy for active microbial populations through the Sites C0014 and C0017 sediment profiles; however, laboratory and sequencing challenges have restricted the dataset and interpretations comparing the indigenous, active microbial community from RNA to that from extant DNA.

References

Biddle J.F., Lipp J.S., Lever M.A., Lloyd K.G., Sørensen K.B., Anderson R., Fredricks H.F., Elvert M., Kelly T.J., Schrag D.P., Sogin M.L., Brenchley J.E., Teske A., House C.H. and Hinrichs K.-U. (2006). Heterotrophic archaea dominate sedimentary subsurface ecosystems off Peru. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 3846–3851.

Biddle J.F., Sylvan J.B., Brazelton W.J., Tully B.J., Edwards K.J., Moyer C.L., Heidelberg J.F. and Nelson W.C. (2012) Prospects for the study of evolution in the deep biosphere. *Front. Microbiol.* **2**, 285-292.

Boetius A., Ravenschlag K., Schubert C.J., Rickert D., Widdel F., Gleseke A., Amann R., Jørgensen B.B., Witte U. and Pfannkuche O. (2000) A marine microbial consortium apparently mediating anaerobic oxidation of methane. *Nature* **407**, 623-626.

Colwell and D'Hondt (2013) Nature and Extent of the Deep Biosphere. *Rev. Mineral. Geochem.* **75**, 547-574.

D'Hondt S., Rutherford S. and Spivack A.J. (2002) Metabolic Activity of Subsurface Life in Deep-Sea Sediments. *Science* **295**, 2067-2070.

D'Hondt S., Spivack A.J., Pockalny R., Ferdelman T.G., Fischer J.P., Kallmeyer J., Abrams L.J., Smith D.C., Graham D., Hasiuk F. and Schrum H. (2009) Subseafloor sedimentary life in the South Pacific Gyre. *Proc. Natl. Acad. Sci. USA* **106**, 11651-11656.

Froelich P.N., Klinkhammer G.P., Bender M.L., Luedtke N.A., Heath G.R., Cullen D., Dauphin P., Hammond D., Hartman B. and Maynard V. (1978) Early oxidation of organic matter in pelagic sediments of the eastern equatorial Atlantic: suboxic diagenesis. *Geochim. Cosmochim. Acta* **43**, 1075-1090.

Gamo T., Ishibashi J., Tsunogai U., Okamura K. and Chiba H. (2006) Unique geochemistry of submarine hydrothermal fluids from arc-back-arc settings of the Western Pacific. In Christie D.M., Fisher C.R., Lee S.M., Givens S. (Eds.), Back-arc Spreading Systems: Geological, Biological, Chemical, and Physical interactions. *Geophys*. *Monogr*. **166**, 147–161.

Head I.M., Jones M. and Larter S.R. (2003) Biological activity in the deep subsurface and the origin of heavy oil. *Nature* **426**, 344-352.

Hinrichs K.-U. and Inagaki F. (2012) Downsizing the Deep Biosphere. *Science* **338**, 204-205.

Inagaki, F., Nunoura, T., Nakagawa, S., Teske, A., Lever M., Lauer, A., Suzuki, M., Takai, K., Delwiche, M., Colwell, F.S., Nealson, K.H., Horikoshi, K., D'Hondt, S. and Jørgensen, B.B. (2006) Biogeographical distribution and diversity of microbes in methane hydrate-bearing deep

marine sediments on the Pacific Ocean Margin. *Proc. Natl. Acad. Sci. USA* **103**, 2815-2820.

Jørgensen B.B. D'Hondt S.L. and Miller D.J. (2006) Leg 201 Synthesis: Controls on Microbial Communities in Deeply Buried Sediments. *Proc. ODP Sci. Res.*, 201.

Jorgensen S.L., Hannisdal B., Lanzén A., Baumberger T., Flesland K., Fonseca R., Øvreeås L., Steen I.H., Thorseth I.H., Pedersen R.B. and Schleper C. (2012) Correlating microbial community profiles with geochemical data in highly stratified sediments from the Arctic Mid-Ocean Ridge. *Proc. Natl. Acad. Sci. USA* **109**, 846-855.

Kallmeyer J., Pockalny R., Adhikari R.R., Smith D.C. and D'Hondt S. (2012) Global distribution of microbial abundance and biomass in subseafloor sediment. *Proc. Natl Acad. Sci. USA* **109**, 16213-16216.

Kawagucci S., Chiba H., Ishibashi J., Yamanaka T., Toki T., Muramatsu Y., Ueno Y., Makabe A., Inoue K., Yoshida N., Nakagawa S., Nunoura T., Takai K., Takahata N., Sano Y., Narita T., Teranishi G., Obata H. and Gamo T. (2011) Hydrothermal fluid geochemistry at the Iheya north field in the mid-Okinawa Trough: implication for origin of methane in subseafloor fluid circulation systems. *Geochem. J.*, **45**,109-124.

Konno U., Tsunogai U., Nakagawa F., Nakaseama M., Ishibashi J., Nunoura T. and Nakamura K. (2006) Liquid CO_2 venting on the seafloor: Yonaguni Knoll IV hydrothermal system, Okinawa Trough. *Geophys. Res. Lett*. **33**, L16607.

Kubo K., Lloyd K.G., Biddle J.F., Amann R., Teske A. and Knittel K. (2012) Archaea of the Miscellaneous Crenarchaeotal Group are abundant, diverse and widespread in marine sediments. *ISME J.* **6**, 1949–1965.

Letouzey J. and Kimura M. (1985) The Okinawa Trough genesis, structure and evolution of a back-arc basin developed in a continent. *Marine Pet. Geol.* **2**, 111-130.

Lipp J. S., Morono Y., Inagaki F. and Hinrichs K.-U. (2008) Significant contribution of Archaea to extant biomass in marine subsurface sediments. *Nature* **454**, 991–994.

Lloyd K.G., Schreiber L., Petersen D.G., Kjeldsen K.U., Lever M.A., Steen A.D., Stepanauskas R., Richter M., Kleindienst S., Lenk S. Schramm A., and Jørgensen B.B. (2013) Predominant archaea in marine sediments degrade detrital proteins. *Nature* **496**, 215-218.

Martino, A.J. (2014). Assessment and advancement of approaches for the study of subseafloor microbial communities. PhD Thesis, The Pennsylvania State University, University Park, PA.

Olsen G.J., Lane D.J., Giovannoni S.J., Pace N.R. and Stahl, D.A. (1986) Microbial ecology and evolution: A ribosomal RNA approach. *Ann. Rev. Microbiol.* **40**, 337-365.

Orcutt B.B, 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.

Parkes J.R., Cragg B.A. and Wellsbury P. (2000) Recent studies on bacterial populations and processes in subseafoor sediments: A review. *Hydrogeol. J.* **8**, 11-28.

Parkes J.R., Cragg B., Roussel E., Webster G., Weightman A. and Sass H. (2014) A review of prokaryotic populations and processes in sub-seafloor sediments, including biosphere:geosphere interactions. *Marine Geol.* **352**, 409-425.

Rappé M.S. and Giovannoni S.J. (2003) The uncultured microbial majority. *Ann. Rev. Microbiol.* **57**, 369-394.

Sakai H., Gamo T., Kim E.-S., Shitashima K., Yanagisawa F., Tsutsumi M., Ishibashi J., Sano Y., Wakita H., Tanaka T., Matsumoto T., Naganuma T. and Mitsuzawa K. (1990a) Unique chemistry of the hydrothermal solution in the mid-Okinawa Trough backarc basin. *Geophys. Res. Lett.* **17**, 2133–2136.

Sakai H., Gamo T., Kim E.-S., Tsutsumi M., Tanaka T., Ishibashi J., Wakita H., Yamano M. and Oomori T. (1990b). Venting of carbon dioxide–rich fluid and hydrate formation in mid-Okinawa Trough backarc basin. *Science*, **248**, 1093–1096.

Schrenk M.O., Huber J.A. and Edwards K.J. (2010) Microbial Provinces in the Subseafloor. *Ann. Rev. Marine Sci.* **2**, 279-304.

Shokralla S., Spall J.L., Gibson J.F. and Hajibabei M. (2012) Next-generation sequencing technologies for environmental DNA research. *Molec. Ecol.* **21**, 1794-1805.

Sørensen K.B., Lauer A. and Teske A. (2004) Archaeal phylotypes in a metal-rich and low-activity deep subsurface sediment of the Peru Basin, ODP Leg 201, Site 1231. *Geobiol.* **2**, 151-161.

Strous M., Fuerst J.A., Kramer H.M., Logemann S., Muyzer G., van de Pas-Schoonen K.T., Webb R., Kuenen J.G. and Jetten M.S.M. (1999) Missing lithotroph identified as new planctomycete. *Nature* **4000**, 446-449.

Takai K., Moser D.P., DeFlaun M. Onstott T.C., and Fredrickson J.K. (2001) Archaeal Diversity in Water from Deep South African Gold Mines. *Appl. Environ. Microbiol.* **67**, 5750-5760.

Takai K., Mottl M.J. Nielsen S.H. and the Expedition 331 Scientists (2011) Expedition 331 summary. *Proc. IODP*, 311.

Teske A. and Sørensen K.B. (2008) Uncultured archaea in deep marine subsurface sediments: have we caught them all? *ISME J.* **2**, 3-18.

Wilhelms A., Larter S.R., Farrimond P., di-Primio R. and Zwach C. (2001) Biodegradation of oil in uplifted basins prevented by deep-burial sterilization. *Nature* **411**, 1034-1037.

Yanagawa K., Nunoura T., McAllister S.M., Hirai M., Breuker A., Brandt L., House C.H., Moyer C.L., Birrien J.-L., Aoike K., Sunamura M., Urabe T., Mottl M.J., and Takai K. (2013) The first microbiological contamination assessment by deep-sea drilling and coring by the D/V *Chikyu* at the Iheya North hydrothermal field in the Mid-Okinawa Trough (IODP Expedition 331). *Front. Microbiol.* **4**, 327.

Chapter 2

Marine subsurface microbial community shifts across a hydrothermal gradient in Okinawa Trough sediments

Abstract

The sediments within the Okinawa backarc basin overlay a subsurface hydrothermal network, creating intense temperature gradients with sediment depth and potential limits for microbial diversity. We investigated taxonomic changes across 45 m of recovered core with a temperature gradient of 3°C/m from the dynamic Iheya North Hydrothermal system. The interval transitions sharply from low-temperature marine mud to hydrothermally altered clay at 10 meters below seafloor (mbsf). From recoverable DNA, analysis of the 16S rRNA gene shows a shift over 15 meters from a heterogeneous community of cosmopolitan marine subsurface taxa (*e.g.* Chloroflexi, Candidate division JS1, Planctomycetes, Proteobacteria) toward primarily archaeal sequences in the deepest, hydrothermal clay horizons of the predicted biosphere. Notably, the bacterial phylum Chloroflexi accounts for a major proportion of the total microbial community within the upper 10 mbsf, whereas, (hyper)thermophilic, heterotrophic as well as thermophilic anaerobic methanotrophic archaea (ANME) appear in varying local abundances in deeper, hydrothermal clay horizons with higher *in situ* temperatures (approaching 55°C). In addition, geochemical evidence suggests that methanotrophy may be occurring in various horizons. Here, we present taxonomic results that support a conceptual model in which common marine subsurface taxa (*e.g.* Chloroflexi and Bathyarchaeota (formerly Miscellaneous Crenarchaeotic Group)) persist into the subsurface, and archaeal taxa (thermophilic methanotrophs, Terrestrial Hot Spring Crenarchaeotic Group (THSCG)) show localized peaks in abundances below into hydrothermal clay horizons. There is also relict DNA, or DNA preserved through endured burial

after deposition, in horizons where the conditions suitable for communities has ceased (ANME, THSCG, Bathyarchaeota, and plant DNA).

Introduction

The marine subsurface hosts a diverse ecosystem of microbial life that has direct consequences on whether organic carbon or other elements are sequestered over geologic time, or are recycled as active elements back into the ocean-atmosphere system (Hinrichs and Inagaki, 2012). The limits of life that define the microbial extent within marine subsurface sediments have remained unresolved; however, the increased temperature associated with sediment burial is often perceived as one of the major constraints (Hinrichs and Inagaki, 2012). There is evidence of microbial populations existing and thriving around hydrothermal vent emissions (reaching temperatures approaching 400°C) (Fisher *et al.*, 2007), but many recent studies have begun to focus on exploring whether such a hyperthermophilic biosphere exists at higher temperatures within the sedimentary subsurface (*i.e.* Parkes *et al.*, 2000). Rather than exploring the temperature limits at 4 km below seafloor, where temperatures are predicted by the geothermal gradient to approach 100°C, subsurface hydrothermal sediments have become ideal study sites because of their large thermal gradients over a much shorter vertical profile (*e.g.* Guaymas Basin, Juan de Fuca Ridge, Middle Valley, Okinawa Trough). Deep sea sediments exist at predominantly low temperatures (~1-5°C); however, areas of new ocean crust formation (*e.g.* mid-ocean ridges or zones of backarc spreading) create localized hydrothermal vent systems – emissions of high temperature fluids emanating from the subsurface as a result of magmatic degassing and subsurface water-rock reactions under high temperatures and pressures (Orcutt et al., 2011). Fluids migrating through these sediments undergo heating and water-sediment interactions that create distinct subsurface geochemical conditions from cold marine sediments.

Backarc basins are the result of the rifting away of a magmatic arc from a continental margin and can have significant sediment input from continental runoff, surface productivity, and/or the volcaniclastic debris (Miall, 2000). The Okinawa Trough is deposition basin with relatively high rates, 50 cm yr^{-1}, of sedimentation from the continental shelf and island arc (Chang *et al.*, 2008). In the case of the Okinawa backarc basin system, extension and rifting of the overriding plate, which began in the Miocene (Lee *et al.*, 1980), coincides with the continental shelf. This continental margin-like geographic setting, being the transition from continental to oceanic crust, and overlaying a subsurface hydrothermal system makes the Okinawa backarc basin a unique marine environment from other sediment-hosted hydrothermal systems. Sediment profiles within this system are subject to intense temperature and alteration gradients, making it an ideal system to examine how the sedimentary biosphere may be affected through such gradients.

The Integrated Ocean Drilling Program (IODP) Expedition 331 recovered sediments within the Iheya North Hydrothermal Field in the middle Okinawa Trough to explore the extent and diversity of the "subvent" biosphere. Site C0014, located 450 m away from the main hydrothermal vent, was investigated to test for a taxonomically diverse microbial community across a temperature gradient increasing with depth. Recovered core from this site extends to approximately 100 m depth into the seafloor. The recovery of as much as 5 m of sediment containing foraminifers along with the estimated sedimentation rate suggests a maximum of 10 kyr of sediment accumulation here (Takai et al., 2011). The sediment profile at Site C0014 exhibits a transition from hemipelagic ooze with pumiceous volcaniclastic sediments to a hydrothermally altered sequence of clays within the top ~10 mbsf of sediment (Takai et al., 2011). Temperature measurements indicate a gradient of approximately 3°C/m (Table A-1 and Figure A-1(E)) (Takai et al., 2011), which is roughly an order of magnitude greater than continental margin sites (*e.g.* Cascadia Margin, IODP 311 and Costa Rica Margin, IODP 344),

but is more gradual than intense, centimeter-scale gradients from other hot, surface sediments. Expanding upon studies that suggest the presence of microbial life deep into sediments (Parkes *et al.*, 2011; Ciobanu *et al.*, 2014), this study is intended to provide a more comprehensive analysis of the microbial community composition through a temperature gradient. In order to investigate the relationship between microbial communities and the potential affect of increased temperature with burial in the sediment profile, we used high throughput sequencing of the 16S rRNA gene to produce a taxonomic analysis of the bacterial and archaeal communities down core at Site C0014. We use this study as a proxy of the distribution of life at the biotic fringe in deeper subsurface sediments, where we hypothesize that either 1) mesophilic taxa reach a threshold and are replaced by a higher temperature adapted community in deeper, hotter horizons, or 2) there is minimal species change down to some horizon and no (re)establishment of a (hyper)thermophilic community due to the dynamic nature of this subsurface hydrothermal system. Here, we report on the taxonomic distributions at approximately meter interval sampling through the temperature and geochemical gradient and speculate on the extent of the biosphere based on DNA recovery and geochemical measurements.

Experimental Procedures

Sample Collection and Extraction

All samples in this study were collected on IODP Expedition 331 at Sites C0014 and C0015. Sediment sections from Holes B, D, and G, cored within approximately 10 m of one another, were used in this study. Aliquots of sediment from all sub-cores were, in a sterile sampling environment, taken from the center of whole rounds stored at -80°C in sterile containers. Samples were shipped on dry ice to the Pennsylvania State University and were kept

at -80°C until analysis. See Table A-1 for a list of samples used in this study. Sample depths are reported in units of depth-below-seafloor, which only takes into account the distance of sampling section from the sediment-water interface and does not take into consideration consolidation of sediment, sediment composition, biostratigraphy, sediment age, or *in situ* temperature. The values reported in this study are the averages of the top and bottom depths of the sub-core (see Table A-1 for depth intervals of each core). DNA was extracted in quadruplicate for a total of 1-2 g sediment using the PowerSoil® DNA Isolation Kit (MoBio Laboratories, Inc.) with modifications as follows: Step 1 – 200 µL sterile TE buffer was added to 0.25-0.5 sediment g in the PowerBead Tubes in addition to the C1 solution; Step 5 – PowerBead Tubes were homogenized for 30 seconds; DNA extract was pooled together at the end. DNA was stored at -20°C until further use. The hydrothermal clay is very difficult from which to extract DNA due to not only the complicated clay matrix that can inhibit PCR reactions or bind to DNA, but also the instability of DNA at lower pH or overall degraded nature of DNA from extreme temperatures (Herrera *et al.*, 2007). For example, the addition of a TE buffer was added to the sediment, as in many cases, the 0.25 g sediment/clay per extraction tube completely absorbed the initial 60 µl lysis solution.

16S rRNA Gene Amplification and Sequencing of DNA

Polymerase chain reactions were executed with pre-dispensed, freeze-dried PCR reagents via the illustra™ puReTaq Read-To-Go PCR Beads (GE Healthcare Life Sciences) to selectively amplify the V6-V9 hypervariable regions of the 16S rRNA gene of archaeal and bacterial species. Selective amplification was performed using the primer pair 906F (5'-AAACTYAAAKGAATTGRCGG-3') and a modified version of 1392R (5'-ACGGGCGGTGTGTRC-3') (Rhodes et al., 2012), which were modified further with the addition of oligonucleotide adapters used in the 454 sequencing protocol, as well as barcodes to

permit numerous samples to be sequenced together and still distinguished in downstream analysis. This modified primerset was initially designed and tested on Dead Sea mesocosm water samples, and has successfully shown phylogenetically diverse amplification of both bacterial and archaeal species (Rhodes *et al.*, 2012). We used the following proportions of reagents with the PCR beads: 1.5 µL of forward primer (10 µM), 1.5 µL of reverse primer (10 µM), 10 µL DNA template, 12 µL sterile water. The mixture was incubated at 94°C for 5 minutes and was followed by 28 cycles of alternating temperatures as follows: 94°C for 1 min, 53°C for 25 s, 72°C for 2 min. For the following samples, 34 cycles of PCR was used: C0014B-1H-5, C0014B-2H-7, C0014B-4H-8, C0014B-5H-15, and C0014G-5H-3. A final elongation step at 72°C was extended for 20 min. PCR products were gel purified on a 1% agarose gel using the PrepEase® Gel Extraction Kit (Affymetrix, Inc.) according to the manufacturer's instructions.

Select samples were also sequenced using Illumina technology (Table A-1). DNA extract was sent to the Marine Biological Laboratory for all preparation and sequencing. Amplification for the Archaeal V6 region used forward primer 958F (AATTGGANTCAACGCCGG) and reverse primer 1048R (CGRCRGCCATGYACCWC). Individual oligos are mixed in equal proportions for a 10 µM working concentration. Polymerase chain reaction mixture conditions for a 100 µL reaction are as follows: 1X HiFi Buffer, 2 mM MgSO4, 0.2 mM dNTPs, 0.3 µM combined primers, 10 units Platinum HiFi, 5-20 ng template. The mixture was incubated at 94°C for 3 minutes and was followed by 30 cycles of alternating temperatures as follows: 94°C for 30 s, 60°C for 45 s, 72°C for 1 min. A final extension was held at 72°C for 2 min. All reactions were done in triplicate. The reactions were cleaned and unwanted small products were removed using Qiagen 96-well MinElute plates. The multiplex pools were size selected by Pippin Prep and quantified using KapaBiosystems qPCR before clustering on the flowcell. Refer to http://vamps.mbl.edu/resources/primers.php for additional information. Data are a part of Projects "DCO_BRA_Bv6" (bacterial samples) and "DCO_BRA_Av6" (archaeal samples).

Analysis of 16S rRNA Gene Amplicons

Sample demultiplexing was performed in Mothur (v.1.30.1), as well as some preliminary quality controls eliminating sequences shorter than 100 basepairs, with more than one mismatch in the barcode sequence, with more than two mismatches in the primer sequence, or with more than eight homopolymers. In addition, sequences were screened by quality score using the "qwindowaverage" function, set at a quality of 35 (Schloss et al., 2009). These files have been made available on the Metagenomics RAST server (metagenomics.anl.gov) under project name "IODP331_amplicons" with MG-RAST ID accession numbers 4633437.3 – 4633472.3 (see Table S1) (Meyer et al., 2008).

The resulting individual fasta files were then processed as a single job with the NGS analysis pipline of the SILVA rRNA gene database project (SILVAngs 1.1) according to the default parameters (Quast et al., 2013). This pipeline included alignment with the SILVA Incremental Aligner (SINA v1.2.10 for ARB SVN (revision 21008)) (Pruesse et al., 2012) against the SILVA SSU rRNA SEED and quality controlled (Quast et al., 2013). Specifically, reads that had fewer than 50 aligned nucleotides and/or more than 2% ambiguities or homopolymers were excluded from further processing, as were putative contaminations, artifacts, and reads with a low alignment quality. The remaining sequences were dereplicated and clustered into operational taxonomic units (OTUs) on a per sample basis, using cd-hit-est (version 3.1.2; http://www.bioinformatics.org/cd-hit) (Li and Godzik, 2006) running in *accurate mode*, ignoring overhangs, and applying identity criteria of 1.00 and 0.98, respectively. Classification was performed by a local nucleotide BLAST (Altschul *et al.*, 1990) search against the non-redundant version of the SILVA SSU Ref dataset (release 115; http://www.arb-silva.de) using blastn (version 2.2.28+; http://blast.ncbi.nlm.nih.gov/Blast.cgi) with standard settings (Camacho et al., 2009). Reads were classified if the value of the function "(% sequence identity + % alignment

coverage)/2" exceeded 93. Classification of each OTU reference read was mapped onto all reads that were assigned to the respective OTU. Those that did not fall within this classification quality were assigned to the group "No Relative." Classifications of the reference sequences were then mapped onto reads within their respective OTUs.

The Illumina dataset was analyzed as a part of the Visualization and Analysis of Microbial Populations Structure (VAMPS) (Huse et al., 2014). An additional assessment of the dataset used the SilvaNGS 1.1 pipeline as described in the previous paragraph (Quast *et al.*, 2013). See Supplemental Discussion for further results.

The maximum likelihood tree in Figure 4 was made by aligning the full-length reference sequences in MEGA and exporting the alignment to PAUP. The sequences were run through a Branch-and-bound search with random seed numbers, and this base topology was used as a constraint on which to add the shorter ANME-1 sequences from the C0014B-2-10 analysis. The three sequences displayed were condensed down from seven C0014B-2-10 sequences, which were nearly identical. They were combined such that ambiguous base pairs were used to adjust the few differences in base pairs.

Analysis of Geochemical Data

Porewater chemistry data (*i.e.* sulfate, methane, alkalinity, and potassium concentrations in Figure A-1) were downloaded from the SIO7 Data Center (http://sio7.jamstec.go.jp) that aims to distribute the science data acquired by International Ocean Discovery Program and Integrated Ocean Drilling Program expeditions of D/V Chikyu. The scientific data recorded are from the J-CORES database. The temperature data points were referenced in the IODP Expedition 331 Proceedings (Takai et al., 2011). The samples analyzed for $\delta^{13}CH_4$ came from sediment plugs stored in gas-tight storage vials (Table A-2). Gas samples were analyzed using a HP 5890 Series

II GC with a flame ionization detector and a custom vacuum inlet system. Daily standard curves are generated using appropriate standards from Scott Specialty Gases. Analytical precision for these samples is better than ± 2%. ~5 nmoles of analyte was injected into a helium carrier stream and purified using a modified PreCon peripheral device before analysis on a Delta V mass spectrometer (Sowers et al. 2005). External precision on this technique is ± 0.3‰ with daily standards providing the means of accurately reporting data directly on the VPDB scale. Modifications to the system correct for the recently discovered Kr interference during $\delta^{13}CH_4$ analyses with an additional chromatography step after the combustion of CH_4 to CO_2 to separate Kr from CO_2 (Schmitt et al. 2013).

Correlation Analysis

The software package IBM® SPSS® Statistics Version 24 was used to run bivariate correlations between dominant microbial taxa and geochemical parameters, using the defaults settings for *Correlation* function. Analyses used relative proportions of taxa available from the 0.30 to 44.85 mbsf. Samples for microbiological and geochemical analyses on IODP Expedition 331 did not come from the same core section. Due to this offset, we used the geochemical information that corresponded closest to the depth of the microbiological sample. The tests produced Pearson Correlation Coefficients and associated significance values (from a 2-tailed test) for each pair of variables tested. The data reported here are only those corresponding to a significance of <0.01.

Results and Discussion

Domains of Life Represented in the Subsurface

From recovered genomic DNA, 28 distinct sediment samples (two samples in duplicate) ranging from the surface to 44.58 mbsf from IODP Expedition 331 Site C0014 were selectively amplified and sequenced for 16S rRNA gene analysis using 454 technology and subsequently classified for taxonomic identification. The amplicon data in Figure 2-1 is presented as a percent relative sequence abundance of classified *Archaea*, *Bacteria*, *Eukaryotes*, and *Not Classified* sequences for each sampled depth horizon, as a proportion from the total sequence yield (see Table A-1 for total sequence yield). The bars in Figure 2-1 generally represent many known subsurface groups, many of which are Archaea, in the upper 16 meters of the sediment column. The dataset also shows isolated peaks in the relative abundance of archaeal sequences to bacterial sequences at depths 10.24, 12.99, and 15.30 mbsf (Figure 2-1). For comparison, IODP Expedition 331 Site C0015, 600 m northwest and upslope of the hydrothermal vent, showed no current hydrothermal activity and was, in this study, shown separately as a control to represent non-hydrothermal conditions within the Iheya North Field. The remainders of the bar graph in Figure 1 represents sequences consistent with those found in the drilling fluid (Yanagawa et al., 2013) and/or extraction blanks (see Appendix A for details) and are excluded from further analyses, as they might not be indigenous.

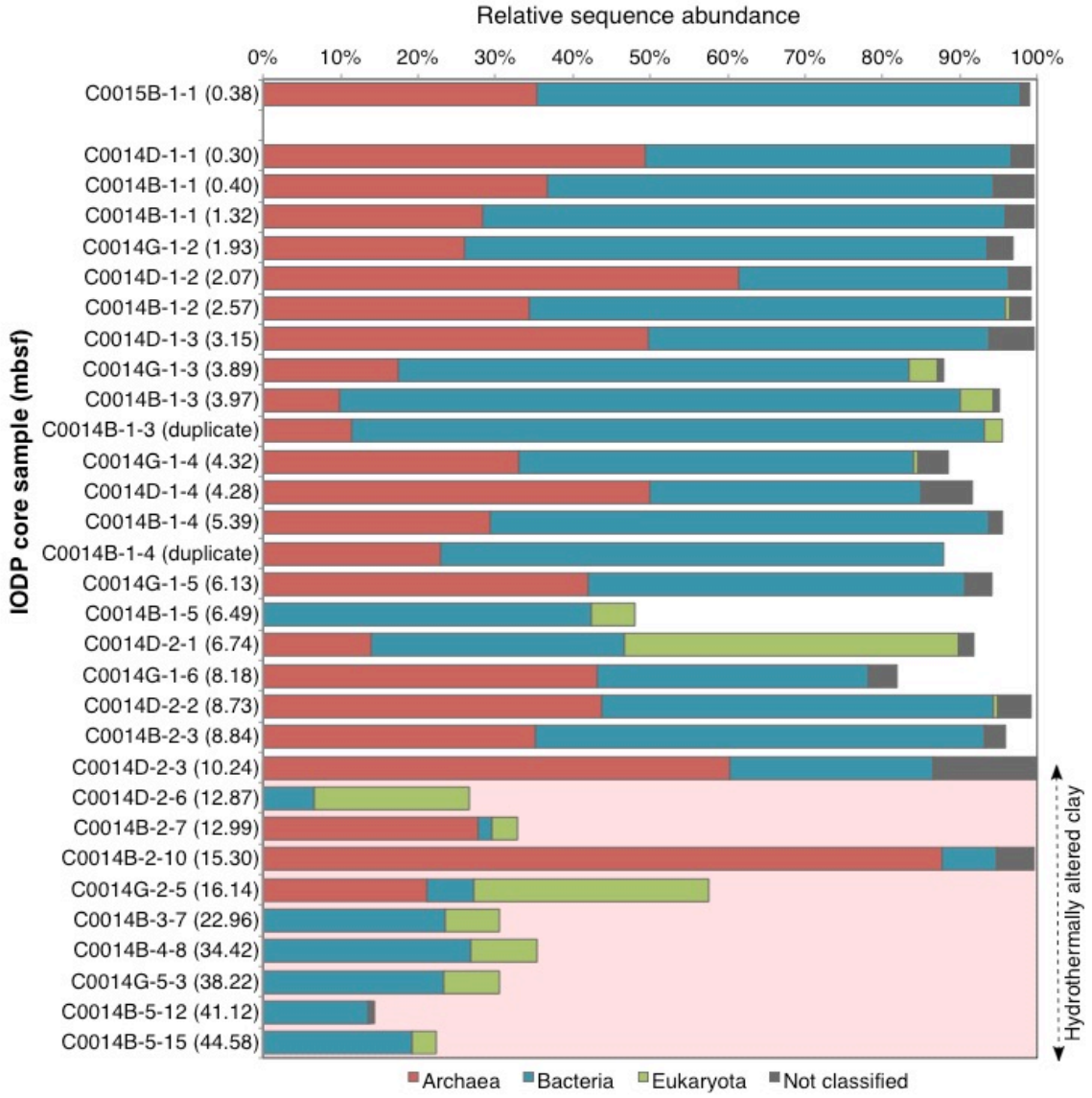

Figure **2-1**: Relative sequence abundance, listed as % across the x-axis, of 16S rRNA gene amplicons from Site C0014 sediment samples classified at the domain level - Archaea, Bacteria, Eukaryota, and Not classified. IODP sample names are listed along the y-axis by increasing depth (meters) below seafloor – depth in parentheses. Site C0015, 600 m northwest and upslope of the hydrothermal vent, (shown separately as the topmost sample) showed no current hydrothermal activity and is being compared to represent non-hydrothermal conditions within the Iheya North Field. The remainders of the horizontal bars represent the relative abundance of sequences in that sample consistent with those found in the drilling fluid (Yanagawa et al., 2013) and/or extraction blanks and are excluded here, as they are less likely to represent indigenous taxa. See Table A-1 for sequence information. The red shaded region corresponds to horizons composed of a hydrothermally altered clay lithology.

In this study, consistent with these challenging samples, there is an increased proportion of sequences that matched sequences from contaminants in drilling fluid and laboratory extraction blanks in the deeper horizons. This trend is observed beginning at the 16.14 mbsf horizon, where the remainders of the bar graph represent a higher percent of recoverable DNA sequences than in shallower horizons (Figure 2-1). Additionally, bivariate correlations between the relative proportion of taxa from each sediment horizon and corresponding environmental variables demonstrate the strongest positive, 0.864 and 0.860, statistically significant ($p<0.01$, 2-tailed test) correlation between these "Non-indigenous sequences" with temperature and depth, respectively (Table 2-1). Eukaryotic sequences also appear in most horizons, despite the archaeal and bacterial specificity of the 16S rRNA primers, with several horizons in significantly high relative abundance represented primarily by Mollusca (6.74 mbsf), Basidiomycota (12.87 mbsf), and Ascomycota (16.14 mbsf) (Figure A-4). Although we did not anticipate that eukaryotic DNA would amplify with our 16S rRNA specific primer set, these sequences most likely represent a combination of amplified indigenous and/or relict environmental DNA. For example, Edgcomb et al., 2011 and Orsi *et al.*, 2013 have demonstrated that the marine subsurface down to at least 35 mbsf in marine sediments is occupied by living eukaryotes, primarily fungi (*i.e. Basidiomycetes*), as well as ancient (2.7 Myr) eukaryotic genomic material (diatoms, Viridiplantae, Alveolata, and Fungi). Also below 16.14 mbsf, there are recovered sequences that likely indicate a signal for terrestrial runoff and environmental DNA. Interestingly, archaeal sequences were not detected in contamination assessments (Yanagawa et al., 2013). Thus, we interpret archaeal DNA to be representative of an indigenous microbial community in this study. We are, therefore, skeptical of the remaining bacterial sequences and extant terrestrial DNA found in the data sets from five deepest horizons, and, at present, we conclude that these horizons do not have substantial microbial. Overall, credible bacterial 16S rRNA gene amplicons, and any archaeal 16S rRNA

gene amplicons, could not be recovered below 16.14 mbsf. Thus, we cannot confidently make detailed conclusions regarding specific taxonomic shifts in the subsurface biosphere below 16.14 mbsf (*in situ* temperature of ca. 55°C) due to the much-reduced DNA yield near this discontinuity.

Subsurface Prokaryotic Diversity

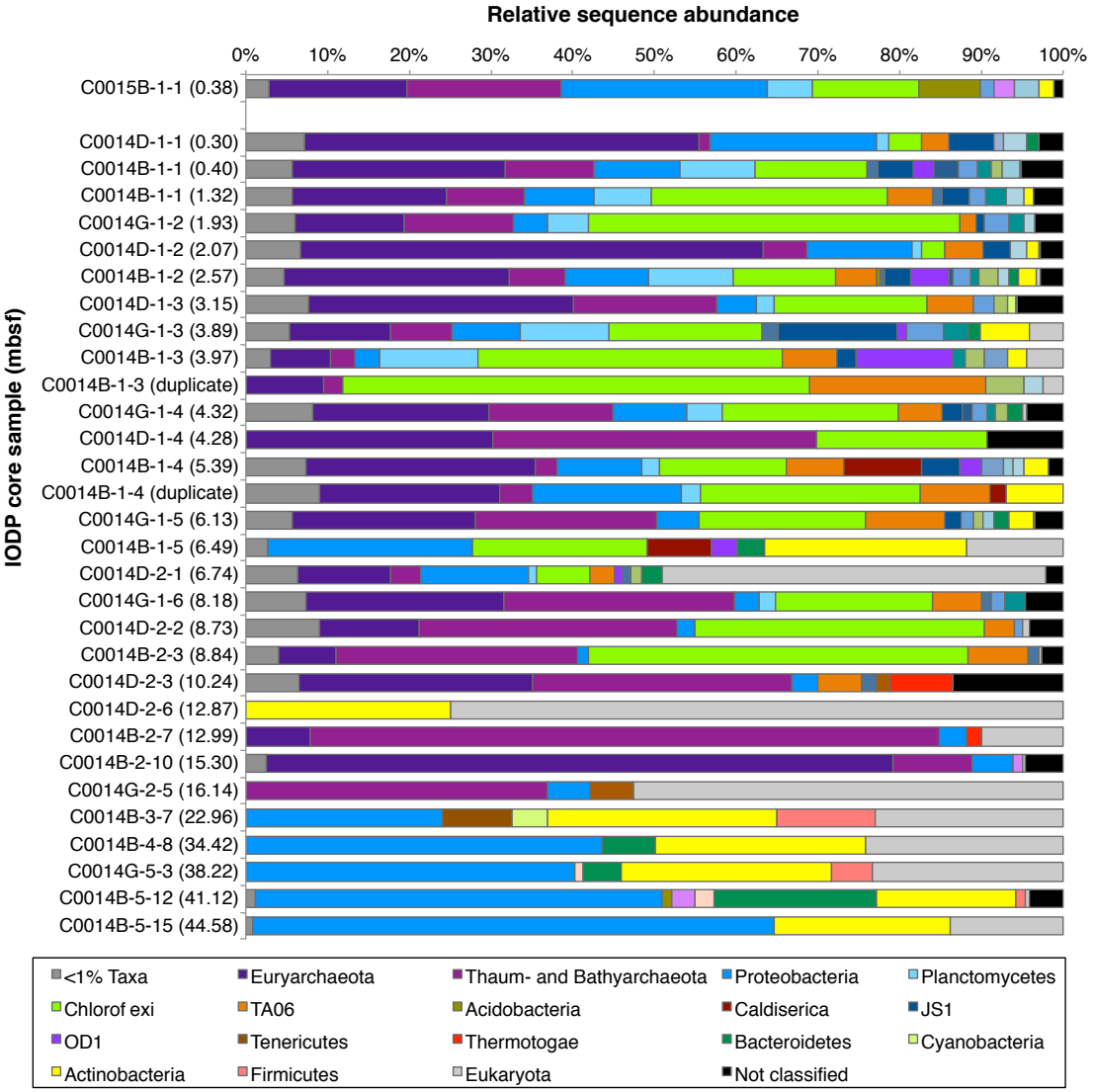

Figure **2-2**: Relative sequence abundance, listed as % across the x-axis, of 16S rRNA gene amplicons from Site C0014 sediment samples classified at the phylum level. Sample horizons are listed by increasing depth below seafloor. IODP sample names, are listed along the y-axis by increasing depth (meters) below seafloor – depth in parentheses. Site C0015, 600 m northwest and upslope of the hydrothermal vent, (shown separately as the topmost sample) showed no current hydrothermal activity and is being compared to represent non-hydrothermal conditions within the Iheya North Field. Sequences included here are those identified as likely to represent indigenous taxa in Figure 1. See Table A-1 for additional sequence information.

The relative abundances of bacterial and archaeal amplicons interpreted at the phylum level (Figure 2-2) reveal a diverse community with distinct community shifts toward higher *in situ* temperatures. For example, the various horizons studied in the top 8.84 mbsf at IODP Expedition Site C0014 show similarities in their most highly represented phyla (*i.e.* Euryarchaeota, Thaum- and Crenarchaeota, Proteobacteria, Planctomycetes, Chloroflexi, and TA06), and they all show a broad similarity to the C0015 control horizon. However, the Chloroflexi sequences, which appear frequently in oceanic subsurface sediments (Rappé and Giovannoni, 2003), is consistently represented as a significant phylotype throughout this interval, but is nearly absent at 10.24 mbsf, and is not present in deeper horizons. The phyla Bacterial Candidatus TA06 and Planctomycetes follow a similar trend as Chloroflexi with less overall abundance, and also abruptly disappears beyond 10.24 mbsf. Bivariate correlations between the relative proportions of taxa from each sediment horizon shown in Table 2-1 demonstrate this significant ($p<0.01$, 2-tailed test) correlation between Chloroflexi and TA06 sequence abundances. Similarly, a high diversity of less abundant bacterial phyla is observed at Site C0014, but only in samples from above 10.24 mbsf. Samples C0014B-1-5 and C0014D-2-6 are unique from their surrounding horizons in that they yielded no archaeal sequences. While the 6.49 and 12.87 mbsf do not show any geochemical or lithological anomalies to infer potentially unique microbial niches, we can only speculate that we have not captured the entire sample diversity in these horizons due to deficient microbial biomass, or that no archaeal community exists at these horizons. The abundance of microbial cells reported from IODP Expedition 331 Site C0014 indicate detectable (on the order of 10^6-10^8 cells/mL sediment) cells down to 2.35 mbsf in Hole B (except C0014B-2-10, approximately 15.3 mbsf, where cells were detected) and 10.17 mbsf in Hole D (Takai et al., 2011). Thus, reduced microbial assemblages at depth as well as method limitations make DNA recovery more tenuous. Deeper in the section, we observe higher relative abundances of DNA amplicons from euryarchaeotic taxa at 12.99 and 15.30 mbsf, and "Cren-

and Thaumarchaeota" at 15.30 mbsf. Additionally, the phylum Thermatogae becomes more abundant at 10.24 and 12.99 mbsf, which may represent a shift towards more optimal conditions for this largely thermophilic phylum. Because the modified primerset appeared to successfully amplify both Euryarchaeota and Cren- and Thaumarchaeota sequences consistently through most horizons above 16.14 mbsf, we do not interpret the signals from 10.24, 12.99, and 15.30 mbsf to be a consequence of primer and/or amplification bias. Furthermore, detectable microbial cells from ~15 mbsf indicates that a substantial microbial assemblage contributes to the DNA recovered.

Table **2-1**: Bivariate correlation values of environmental variables with significant correlation to dominant taxa (second column); taxa that have significant correlations with one another (third column); environmental variables that have significant correlation with depth (bottom row). The chemical species abbreviation or taxon name is listed with the correlation value in parentheses. All correlation values listed here are significant at the 0.01 level (2-tailed).

	Environmental variables	Other phyla
Archaea (domain)	Br(-0.467), Temperature(-0.502), Depth(-0.500)	
Non-indigenous sequences	B(0.755), Ba(0.525), Ca(0.721), K(0.725), Li(0.701), Mg(-0.796), Mn(0.788), NH_4(0.765), Na(-0.586), Rb(0.692), SO_4(-0.533), Si(0.690), Temperature(0.864), Depth(0.860), pH(-0.540)	
Deep Sea Hydrothermal Vent Gpoup 6 (DHVEG-6)	Ca(-0.535), K(-0.504), Li(-0.512), Mg(0.504), NH_4(-0.554), pH(-0.601), Temperature(-0.541), Depth(-0.551)	Planctomycetes(0.551)
ANME-1		
Miscellaneous Crenarchaeotic Group (Bathyarchaeota)	Ca(-0.484), SO_4(0.619)	Proteobacteria(-0.497)
Terrestrial Hot Spring Crenarchaeotic Group (THSCG)		
Bacteroidetes	Ba(0.533), Mn(0.647), Temperature(0.561)	Proteobacteria(0.580)
Chloroflexi	Ca(-0.502), K(-0.562), Li(-0.583), Mg(0.608), Mn(0.528), NH_4(-0.607), SO_4(0.463), Temperature(-0.537), Depth(-0.513)	TA06(0.677)
Proteobacteria	Ca(0.636), Ba(0.615), K(0.593), Li(0.523), Mg(-0.683), Mn(0.735), NH4(0.613), Na(-0.547), SO4(-0.520), Temperature(0.813), Depth(0.825)	
Planctomycetes		
TA06	K(-0.501), Li(-0.512), Mg(0.544), NH4(-0.536), Temperature(-0.469)	
Depth	Ca(0.757), Ba(0.558), K(0.774), Li(0.745), Mg(-0.889), Mn(0.885), NH_4(0.836), Na(-0.692), SO_4(-0.581), pH(-0.500), Temperature(0.997)	

The discontinuation of Chloroflexi and TA06 phyla and general loss of broad microbial diversity below the 10.24 mbsf sample, interestingly, corresponds with a change in clay lithology. Both Chloroflexi and TA06 phyla show significant correlations with depth and temperature, but, more so with K, NH_4, and Mg, which are associated with uptake or exchange of chemical species by changes in clay mineralogy (Figure A-1(F)). The non-hydrothermal hemipelagic ooze shifts to a hydrothermally altered mottled pale gray with alteration products illite and montmorillonite over the course of 9-12 mbsf (Takai et al., 2011). Although the taxonomic richness seems to be affected by the geological and geochemical boundary, archaeal DNA sequences appear to dominate through the transition from temperate to hydrothermal conditions. Overall, our results strongly suggest that this lithologic and temperature transition represents a considerable obstacle for the survival of Chloroflexi and other rare taxa, while certain archaeal taxa are able to persist several meters deeper in hydrothermal clay with an *in situ* temperature of approximately 33°C.

Shifts in Subsurface Archaeal Relative Abundance

Of the total indigenous Prokaryotic sequences within the top 16.14 mbsf at IODP Expedition 331 Site C0014, the recovered archaeal sequences (domain level) increase in relative abundance with depth (Figure 2-3, red diamonds). The bivariate correlation analysis calculated a significant, but weak correlation between the relative proportion of total archaeal sequences recovered and depth and temperature. However, the archaeal relative abundance in the correlation analysis was calculated as a proportion of total sequences recovered. Figure 2-3, on the other hand, is intended to decouple the indigenous prokaryotic sequences from background noise and illustrates a clearer relationship between the proportion of archaeal sequences as a function of depth. Notably, below 10.24 mbsf there is a marked increase in relative abundance in archaeal

sequences, reaching up to 92%, in all but one of the deepest horizons (Figure 2-3). The Site C0015 sample (0.37 mbsf, black diamond) showed an indigenous archaeal fractional abundance of 36% (Figure 2-3), which is similar to the surface sediments of Site C0014. Even though the deepest Site C0014 horizon in which Archaea are found (16.14 mbsf) yields considerably fewer total indigenous sequences, the relative abundance of archaeal sequences is still significantly higher than its surface counterparts. Until recently, Archaea in the marine subsurface were considered to represent an insignificant portion of the active subsurface community (Schippers et al., 2005). However, the data in this study, like more recent findings from Biddle *et al.*, 2006 and Teske and Sørenson, 2008, suggest that the subsurface contains a community with a potentially significant contribution of Archaea. The highest relative abundance of Archaea at 92% occurs at 15.30 mbsf (Figure 2-3), corresponding to an estimated temperature of 55°C (Table A-1). Archaeal sequences are detected at 78% in the subsequent sample at 16.14 mbsf (Figure 2-3), but are not present at all in the sequencing results from deeper horizons. Ten additional efforts to recover and amplify sequences from samples below 16.14 mbsf failed (see Table A-1 for details). These deeper samples between 10.24 and 16.14 mbsf have *in situ* temperatures approaching 57°C (Table A-1) and suggest a shift in community toward archaeal thermophiles.

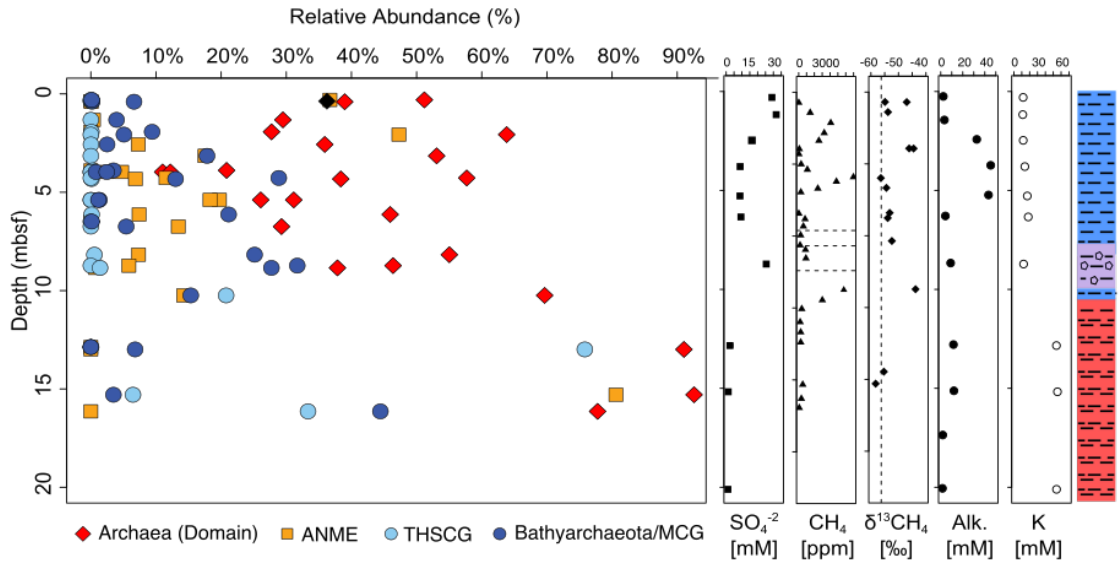

Figure **2-3**: Relative abundance of archaeal (domain) sequences (red diamonds) and three featured subclassifications of archaeal taxa. The percentage values were calculated as a proportion of "indigenous" prokaryotic sequences, which excludes eukaryotic, not classified, and non-indigenous bacterial sequences. Anaerobic methanotrophic archaea (ANME, orange squares) are classified within the Euryarchaeota phylum, while the Bathyarchaeota/MCG (dark blue circles) and Terrestrial Hot Spring Crenarchaeotic Group (THSCG, light blue circles) are within the Thaumarchaeota. Archaeal sequences were not detected below 16.14 mbsf in this study. The panels on the right are porewater geochemistry measurements (sulfate, methane, carbon isotope values of methane, total alkalinity, potassium) and lithologic units. Data extended to further depth in these panels are also displayed in Figure A-1.

Shifts in Subsurface Archaeal Taxa

Figure 2-3 shows the Bathyarchaeota (formerly Miscellaneous Crenarchaeotic Group) (dark blue circles) and Terrestrial Hot Spring Crenarchaeotic Group (THSCG) from the Thaumarchaeota phylum (light blue circles). Both taxonomic groups exhibit an increase in relative abundance within the deeper horizons at IODP Expedition 331 Site C0014. Members of the highly diverse Bathyarchaeota are globally distributed in various marine and continental environments. Recent information on the ecological role of Bathyarchaeota archaea has revealed diverse subgroup of organo-heterotrophic and autotrophic acetogens that are capable of degrading

complex carbohydrate polymers of photosynthetic origin, low-molecular weight carbon substrates, and proteins (Kubo *et al.*, 2013, Lloyd *et al.*, 2013, Meng *et al.*, 2014, Meador *et al.*, 2015, Lazar *et al.*, 2016) and have the cellular machinery for a methane metabolism (Evans *et al.*, 2015). The widespread abundance of Bathyarchaeota throughout Site C0014 (Figures 2-3 and A-2) suggests that they could be less affected by the increasing temperature than other bacterial taxa, or could represent persistent, relict DNA. Similarly, the THSCG become more abundant, particularly below the 8.84 mbsf horizon (Figure 2-3), where Archaea represent the majority of indigenous sequences. In the sample from 12.87 mbsf, THSCG represent ~80% of archaeal sequences. No cultured representatives have been studied from the THSCG, however, the documented samples comprising this clade come from a 1-10 cmbsf, 100°C sediment layer within a middle Okinawa Trough hydrothermal field (Takai and Horikoshi, 1999). Additionally, sequences from THSCG have also been documented in the sediments from Iheya North Hydrothermal field Site C0017, or the presumed site of recharging seawater (Yanagawa *et al.*, 2014). Yanagawa *et al.* showed evidence of THSCG at 141 mbsf, corresponding to 83°C, which is optimal for a hyperthermophilic community. Therefore, it is somewhat surprising that we detected THSCG sequences at a 55°C sediment horizon (Site C0014 – 12.87 mbsf). It is possible that our detection of THSCG could be a microbial relict from a time when this horizon experienced a hotter temperature fluctuation. The appearance of THSCG restricted to these deeper horizons, however, demonstrates the recent establishment of a niche community in response to environmental conditions and supports our first hypothesis in which mesophilic, marine subsurface taxa have been replaced by a high temperature adapted microbial population.

Also shown in Figure 2-3 are the relative abundances observed of DNA amplicons representing the anaerobic methanotrophic archaea (ANME, orange squares). ANME are members of a microbial consortium involved in the anaerobic oxidation of methane (AOM) in anoxic marine sediments (Boetius et al., 2000). Generally, driven by sulfate reduction, the

anaerobic oxidation of methane is a critical control on the flux of methane from marine sediments to the atmosphere. Members of these microbial consortia have not yet been isolated, but archaeal subgroups ANME-1 and -2 have been found to be related to methanogenic Archaea of the Methanosarcinales and Methanomicrobiales, and are often found in association with sulfate-reducing bacterial counterparts (Boetius et al., 2000; Knittel et al., 2005). The anaerobic oxidation of methane is a significant process in coastal marine sediments, and this process is recently getting more attention as part of the trophic ecology of vent ecosystems (*e.g.* Guaymas Basin, Gulf of California, Mexico) (Kulm et al., 1986; Teske et al., 2002; Biddle et al., 2012). The sequences at the 15.30 mbsf horizon represent 81% of the total indigenous sequences. Though the relative abundances of ANME are highly variable throughout the sediment profile at Site C0014 (Figures 2-3 and A-2), the highest relative abundance of ANME in the 15.30 mbsf (c.a. 55°C) suggests a potential methane-oxidizing niche in the thermophilic regime. Further taxonomic evidence of a thermophilic methane-oxidizing taxon is seen in Figure 2-4, where sequences from the C0014B-2-10 horizon were aligned with other published ANME-1 sequences. In order to best resolve the taxonomy of the C0014B-2-10 ANME sequences, full-length 16S rRNA gene alignments of reference sequences were generated as a base topology constraint on which to map the shorter amplicons. The maximum likelihood tree produced in Figure 2-4 reveals that most ANME-1 representatives cluster based on temperature regime. ANME-1 sequences from C0014B-2-10 group most closely with other Iheya Basin clones from high temperature enrichments, and are also part of a larger clade with several clones from Guaymas Basin hydrothermal sediments. Thus, the abundant ANME-1 representatives from 15.30 mbsf at Site C0014 appear more similar to high temperature adapted methanotrophs than those found in cold, methane-seep type environments, and indicate a likely abundant and possibly active thermophilic methane-oxidizing community.

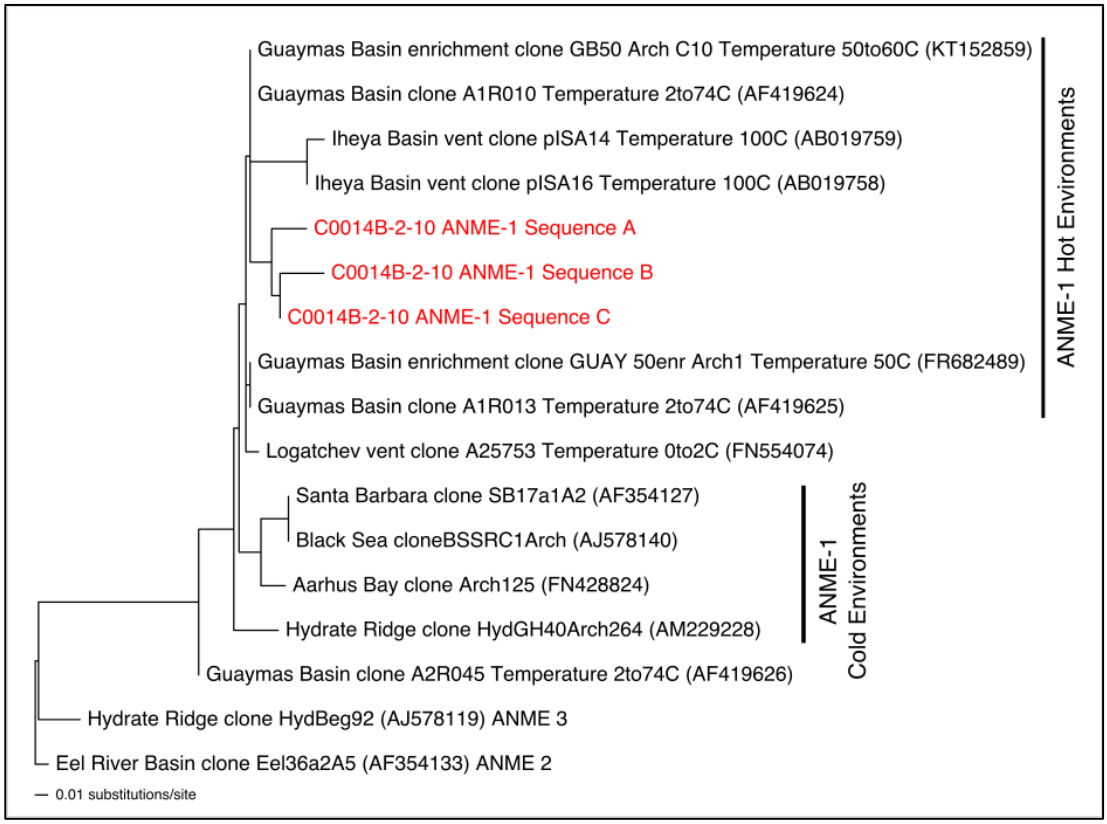

Figure **2-4**: Maximum likelihood of sequence alignment from C0014B-2-10 classified ANME-1 sequences (in red) with other documented ANME-1 sequences. Identifiers in parentheses are the NCBI nucleotide accession numbers. The vertical lines indicate those sequences corresponding to hot or cold temperature regimes.

Generally, geochemical evidence also supports the presence of biological AOM in the top ~15 m of the sediment profile at Site C0014. In the topmost 5 mbsf, which correspond to an estimated temperature range between 5-20°C, there is an inverse relationship between methane and sulfate concentrations that is accompanied by an increase in alkalinity (Figures 2-3 and A-1). Sulfate-methane transition zones provide a niche for microbially driven sulfate-dependent methane oxidation, a process which produces alkalinity (Knittel and Boetius, 2009). ANME sequences are also found in high relative abundances in several of these shallower horizons. Below the peak in methane concentration at 4 mbsf is an overall decrease over the subsequent 4 m, which could be indicative of methane consumption. Additionally, carbon isotopic signatures

observed in methane sampling here are consistent with Rayleigh fractionation during biological anaerobic methane oxidization. One gas headspace measurement at 1.4 mbsf and two at 7.8 mbsf show $\delta^{13}CH_4$ values ranging between -44.5 and -47.0‰ (Table A-2), which are enriched in $\delta^{13}CH_4$ relative to the that of the presumed source gas horizons by 10.1-11.7‰ (Figures 2-3 and A-1(B,C)). These suspected "source" or thermogenic horizons were based on three void gas (extreme degassing of core) measurements at 19.22, 21.36, and 24.81 mbsf and were considered to represent the source gas (an average of -56.27‰) throughout the sediment profile in this study (Figures 2-3 and A-1(C), vertical dashed line). These geochemical observations in the Site C0014 sediment profile are consistent with biological AOM, where biological methane consumption leaves an enrichment of ^{13}C in the remaining methane relative to its source value (Whiticar and Faber, 1986). Elevated concentrations of sulfate also coincide with these two horizons (1.4 and 7.8 mbsf) showing the most enriched $\delta^{13}CH_4$ values (Figures 2-3 and A-1(C)). Pumice lenses have been documented throughout Site C0014 cores and can provide conduits for seawater transport, particularly as significant faulting in this seismically active geologic setting allows for recharging seawater into the system (Takai *et al.*, 2011). Aside from the two independent permeable units at 1.4 and 7.8 mbsf, the characteristic sulfate concentration curve observed here is consistent with other marine sediments in which diffusion is the dominant process controlling sulfate concentrations. Though the sampling resolution for methane isotope measurements is limited, the more depleted $\delta^{13}CH_4$ value at 3 mbsf with a markedly lower sulfate concentration (*i.e.* 16 mM) between these two horizons would suggest that these two seawater intrusions (1.4 and 7.8 mbsf) are localized and do not experience advective communication. However, we interpret the source of entrained seawater at these two horizons as a necessary source of sulfate for biological anaerobic methane oxidation in surrounding horizons of otherwise uniformly, low-conductivity marine clay. Observed ANME sequences throughout this unit, and the particularly high relative

abundances at 2.07 mbsf, 5.39 mbsf, and 6.74 mbsf (Figures 2-3 and A-2), complement the geochemical evidence in support of an active AOM biosphere.

The subsequent 10-15 mbsf horizons, corresponding to a temperature range of 15-55°C, are also accompanied by a decrease in methane concentration with depth, which could be associated with methane consumption. The ANME sequences associated with the 15.30 mbsf horizon are from thermophilic taxa (Figure 2-4) and represent 81% of total sequences (Figure 2-3). Though the taxonomic evidence indicates a thermophilic methane-oxidizing niche, the methane isotope data suggests this community may not be currently wholly active. Beyond 16.14 mbsf, a potential AOM zone at ~27 mbsf is hypothesized from a decrease in methane concentration and an enrichment of $\delta^{13}CH_4$ relative to that of source methane (Figure A-1(B,C)). However, based on the lack of DNA recovery at this depth and an estimated temperature of 96°C, we cannot conclude that this horizon hosts an active biosphere. There are also no reported pumice clasts in this sample horizon to suggest an intrusion of seawater, where $\delta^{13}CH_4$ values of seawater have been reported between -52.0 and -48.2‰ (Sansone *et al.*, 1999). Rather, abiotic AOM could be occurring at or near this depth.

Conclusions

The Okinawa backarc basin is a unique environmental setting to analyze microbial communities through a range of temperatures because of its subsurface hydrothermal network within continental margin-like sediments. This study represents a proxy for the distribution and extent of life in other subsurface environments, where hotter temperatures are more difficult to reach at greater depths. Based on the taxonomic information in this study, the microbial community in the deeper, hotter hydrothermal clay horizons of IODP Expedition Site C0014 is distinct from the shallower, cooler horizons. In this study, we use archaeal sequences as a

confident and conservative estimate of the extent of the biosphere, which extends in this profile down to 16 mbsf. Sequences below, particularly those indicative of relict plant material, suggest that the diminished sequence yield from the indigenous population is dampened by such signals and make detailed conclusions about an indigenous community more tenuous.

The overall heterogeneous community composition in these sediments exhibits similarities to other studied marine sediments, and the results of this study complement the findings from Yanagawa and colleagues (Yanagawa *et al.*, 2016). Many taxa identified here, for example, Chloroflexi and Bathyarchaeota, have ubiquitous distributions in marine subsurface habitats. The widespread distribution of the cosmopolitan Bathyarchaeota persisting into hydrothermal clay likely suggests an ecophysiological flexibility within a wide range of temperature and geological conditions, but could also represent relict DNA. Chloroflexi appear to be restricted to the upper horizons associated with non-hydrothermal marine mud, which reflects a general geochemical, lithological, and/or temperature boundary to much of the microbial diversity. The isolated peaks in abundances of archaeal sequences at depth indicate a recent establishment of potentially better-adapted archaeal community to the prominent hydrothermal conditions over other taxa. The uncultured archaeal taxa identified here (*i.e.* THSCG, thermophilic ANME-1) that are observed here in significant proportion in the hydrothermal clay horizons are also observed in other high temperature sediments (Okinawa Trough and Guaymas Basin) and suggest that high temperature microbial communities may be biogeographically similar among other sedimented hydrothermal vent ecosystems. The appearance of these high temperature taxa support our first hypothesis and demonstrates a transition from mesophilic marine taxa in the top 10 m to the recent establishment of a temperature adapted community restricted to the deeper, hotter horizons. This taxonomic dataset, in combination with geochemical and isotopic data, also suggest that methanotrophy may have once been a significant process occurring in these subsurface sediments, particularly in the thermophilic regime. The

subsurface hydrothermal system in the Iheya North Hydrothermal Field is very dynamic and reflects a diverse subsurface biosphere seemingly adapted to the range of conditions experienced through a hydrothermal gradient, both low and high temperature. Overall, these results support a conceptual model in which a community of cosmopolitan marine subsurface bacteria (*e.g.* Chloroflexi) persists until a lithological/geochemical boundary, but other cosmopolitan archaeal taxa (*i.e.* Bathyarchaeota) persist further into the hydrothermal clay. Additionally, the hydrothermal clay horizons show isolated peaks in abundances of specific high temperature archaeal phylotypes (*i.e.* thermophilic ANME and THSCG) that suggest the recent establishment of high temperature adapted microbial niches maybe have once been supported under different environmental conditions. Lastly, there is also relict DNA in horizons where the conditions suitable for certain communities has ceased (ANME, THSCG, Bathyarchaeota, and plant DNA).

Acknowledgements

Samples for this research were provided by the Integrated Ocean Drilling Program (IODP). We are grateful to the IODP Expedition 331 scientific party and expedition staff onboard the *Chikyu* that assisted with drilling, sampling, measurements, and sample storage during IODP Expedition 331. We thank the National Science Foundation for support for this work through the Ocean Leadership grant BA-40 TO T331B40 and through the Center for Dark Energy Biosphere Investigations (C-DEBI) grant #OCE-0939564. This project was also partially supported by the Penn State Astrobiology Research Center (through the NASA Astrobiology Institute, cooperative agreement #NNA09DA76A). DNA was sequenced in one direction on a quarter plate of a 454 FLX+ Titanium sequencer (454 Life Sciences) at the Pennsylvania State University Center for Genome Analysis, which was partly funded using Tobacco Settlement Funds provided by the Pennsylvania Department of Health. The Roche 454 facility at the Pennsylvania State University

Center for Comparative Genomics and Bioinformatics was funded, in part, by a grant from the Pennsylvania Department of Health using Tobacco Settlement Funds appropriated by the legislature. The Illumina sequencing data were made possible by the Deep Carbon Observatory's Census of Deep Life supported by the Alfred P. Sloan Foundation. Pyrosequencing was performed at the Marine Biological Laboratory (Woods Hole, MA, USA) and we are grateful for the assistance of Mitch Sogin, Susan Huse, Joseph Vineis, Andrew Voorhis, Sharon Grim, and Hilary Morrison at MBL. We thank Dr. Jennifer Biddle (University of Delaware, Lewes) for valuable insights and suggestions during the analyses, and Katsunori Yanagawa (JAMSTEC, Japan) for providing detailed sequencing information on the drilling fluid assessments. We also thank Todd Sowers at the Pennsylvania State University, Department of Geosciences for performing the methane isotope analyses presented in this study.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. (1990) Basic local alignment search tool. *J Molec. Biol.*, **215**, 403-410.

Biddle J.F., Cardman Z., Mendlovitz H., Albert D.B., Lloyd K.G., Boetius A., and Teske A. (2012) Anaerobic oxidation of methane at different temperature regimes in Guaymas Basin hydrothermal sediments. *ISME J.* **6**, 1018-31.

Biddle J.F., Lipp J.S., Lever M.A., Lloyd K.G., Sørensen K.B., Anderson R., Fredricks H.F., Elvert M., Kelly T.J., Schrag D.P., Sogin M.L., Brenchley J.E., Teske A., House C.H. and Hinrichs K.-U. (2006). Heterotrophic archaea dominate sedimentary subsurface ecosystems off Peru. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 3846-3851.

Boetius A., Ravenschlag K., Schubert C.J., Rickert D., Widdel F., Gleseke A., Amann R., Jørgensen B.B., Witte U. and Pfannkuche O. (2000) A marine microbial consortium apparently mediating anaerobic oxidation of methane. *Nature* **407**, 623-626.

Brandt, L.D., Hser Wah Saw, J., Ettema, T., House, C.H. (2015) Marine subsurface microbial communities across a hydrothermal gradient in Okinawa Trough Sediments (B11I-0562). Presented at the American Geophysical Union 2015 Annual Fall Meeting (December 14, 2015, San Francisco, CA).

Camacho C., Coulouris G., Avagyan V., Ma N., Papadopoulos J., Bealer K., and Madden T.L. (2009) BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421-430.

Chang Y.-P., Wang W.-L., Yokoyama Y., Matsuzaki H., Kawahata H., and Chen M.-T. (2008) Millenial-Scale Planktic Foraminifer Faunal Variability in the East China Sea during the Past 40000 Years (IMAGES MD012404 from the Okinawa Trough). *Terr. Atmos. Ocean. Sci.* **19**, 389-401.

Ciobanu, M.-C., Gurgaud, G., Dufresne, A., Breuker, A., Rédou, V., Maamar, S.B., et al. (2014) Microorganisms persist at record depths in the subseafloor of the Canterbury Basin. *ISME J.* **8**: 1370–1380.

DeLong E.F., Wu K.Y., Prezelin B.B., and Jovine R.V.M. (1994) High abundance of Archaea in Antarctic marine picoplankton. *Nature* **370**, 695-697.

Edgcomb V.P., Beaudoin D., Gast R., Biddle J.F., and Teske A. (2011) Marine subsurface eukaryotes: The fungal majority. *Environ. Microbiol.* **13**, 172-183.

Evans, P.N., Parks, D.H., Chadwick, G.L., Robbins, S.J., Orphan, V.J., et al. (2015) Methane metabolism in the archaeal phylum Bathyarchaeota revealed by genome-centric metagenomics. *Science* **350**: 434–438.

Fisher C.R., Takai K., and Le Bris N. (2007) Hydrothermal Vent Ecosystems. *Oceanography* **20**, 14-23.

Herrera, A. and Cockell, C.S. (2007) Exploring microbial diversity in volcanic environments: A review of methods in DNA extraction. *J Microbiol Methods* **70**: 1-12.

Hinrichs K.-U. and Inagaki F. (2012) Downsizing the Deep Biosphere. *Science* **338**, 204-205.

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**, 1-7.

IBM Corp. Released 2016. IBM SPSS Statistics for Mac, Version 24.0. Armonk, NY: IBM Corp.

Inagaki F., Suzuki M., Takai K., Oida H., Sakamoto T., Aoki K., Nealson K.H., and Horikoshi K. (2003) Microbial Communities Associated with Geological Horizons in Coastal Subseafloor Sediments from the Sea of Okhotsk. *Appl. Environ. Microbiol.* **69**, 7224-7235.

Knittel K. and Boetius A. (2009) Anaerobic oxidation of methane: progress with an unknown process. *Annu. Rev. Microbiol.* **63**, 311-334.

Knittel K., Lo T., Boetius A., Kort R., and Amann R. (2005) Diversity and Distribution of Methanotrophic Archaea at Cold Seeps. *Appl. Environ. Microbiol.* **71**, 467-479.

Könneke M., Bernhard A.E., de la Torre J.R., Walker C.B., Waterbury J.B., and Stahl D.A. (2005) Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* **437**, 543-546.

Kubo K., Lloyd K.G., F Biddle J., Amann R., Teske A., and Knittel K. (2012) Archaea of the Miscellaneous Crenarchaeotal Group are abundant, diverse and widespread in marine sediments. *ISME J.* **6**, 1949-1965.

Kulm L.D., Suess E., Moore J.C., Carson B., Lewis B.T., Ritger S.D., Kadko D.C., Thornburg T.M., Embley R.W., Rugh W.D., Massoth G.J., Langseth M.G., Cochrane G.R., and Scamman R.L. (1986) Oregon Subduction Zone: Venting, Fauna, and Carbonates. *Science* **231**, 561–566.

Lazar, C.S., Baker, B.J., Seitz, K., Hyde, A.S., Dick, G.J., Hinrichs, K.-U., and Teske, A.P. (2016) Genomic evidence for distinct carbon substrate preferences and ecological niches of Bathyarchaeota in estuarine sediments. *Environ. Microbiol.*, **18**, 1200-1211.

Lee C.-S., Shor Jr. G.G., Bibee L.D., Lu R.S., and Hilde W.C. (1980) Okinawa Trough: Origin of a Backarc Basin. *Marine Geol.* **35**, 219-241.

Li W. and Godzik A. (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659.

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.

Meador T.B., Bowles M., Lazar C.S., Zhu C., Teske A., and Hinrichs K.-U. (2014) The archaeal lipidome in estuarine sediment dominated by memebers of the Miscellaneous Crenarchaeotal Group. *Environ. Microbiol.* **17**, 2441-2458.

Meyer F., Paarmann D., Souza M.D., Olson R., Glass E.M., Kubal M., Paczian T., Rodriguez A., Stevens R., Wilke A., Wilkening J., and Edwards R.A. (2008) The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* **9**, 386.

Meyer S., Wegener G., Lloyd K.G., Teske A., Boetius A., and Ramette A. (2013) Microbial habitat connectivity across spatial scales and hydrothermal temperature gradients at Guaymas Basin. *Front. Microbiol.* **4**, 207.

Miall A.D. (2000) Interarc and Backarc Basins on Oceanic or Transitional Crust. In, *Principles of Sedimentary Basin Analysis*. Springer, Verlag Berlin Heidelberg, pp. 504–507.

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.

Orsi, W.D., Edgcomb, V.P., Christman, G.D., and Biddle, J.F. (2013) Gene expression in the deep biosphere. *Nature* **499**: 205-208.

Parkes R.J., Cragg B.A., and Wellsbury P. (2000) Recent studies on bactieral populations and processes in subseafloor sediments: A review. *Hydrogeol*. **8**, 11-28.

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

Quast C., Pruesse E., Yilmaz P., Gerken J., Schweer T., Yarza P., Peplies J., and Glöckner F.O. (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590-6.

Rappé M.S. and Giovannoni S.J. (2003) The uncultured microbial majority. *Annu. Rev. Microbiol.* **57**, 369–394.

Rhodes M.E., Oren A., and House C.H. (2012) Dynamics and Persistence of Dead Sea Microbial Populations as Shown by High-Throughput Sequencing of rRNA. *Appl. Environ. Microbiol.* **78**, 2489–2492.

Sanone, F.J., Holmes, M.E., and Popp, B.N. (1999) Methane stable isotope ratios and concentrations as indicators of methane dynamics in estuaries. *Global Biogeochem Cycles* **13**: 463-474.

Schippers A., Neretin L.N., Kallmeyer J., Ferdelman T.G., Cragg B.A., Parkes R.J., and Jørgensen B.B. (2005) Prokaryotic cells of the deep sub-seafloor biosphere identified as living bacteria. *Nature* **433**, 861–864.

Schloss P.D., Westcott S.L., Ryabin T., Hall J.R., Hartmann M., Hollister E.B., Lesniewski R.A., Oakley B.B., Parks, D.H., Robinson C.J., Sahl J.W., Stres B., Thallinger G.G., Van Horn D.J., and Weber C.F. (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75**, 7537-7341.

Schmitt, J., Seth, B., Bock, M., van der Veen, C., Sapart, C.J., Prokopiou, M., Sowers, T., Rockmann, T., and Fischer, H. (2013). "On the interference of 86Kr2+ during carbon isotope analysis of atmospheric methane using continuous flow combustion – isotope ratio mass spectrometry." *Atmos. Meas. Tech. Discuss.* **6**, 1409-1460, 10.5194/amtd-6-1409-2013.

Sowers, T., Bernard, S., Aballain, O., Chappellaz, J., Barnola, J.-M., and Marik, T. (2005) Records of the $\delta^{13}C$ of atmospheric CH4 over the last two centuries as recorded in Antarctic snow and ice. *Global Biogeochem. Cycles*, **19**(GB2002)doi:10.1029/2004GB002408.

Takai K., Mottl M.J., Nielsen S.H., and Expedition 331 Scientists (2011a) Site C0014. *Proc. Integr. Ocean Drill. Progr.*, **331.**

Takai K., Mottl M.J., Nielsen S.H., and The Expedition 331 Scientists (2011b) Expedition 331 summary. *Proc. Integr. Ocean Drill. Progr.*, **331.**

Teske, A. (2013) Marine Deep Sediment Microbial Communities. In, Rosenberg E., DeLong E.F., Lory S., Stackebrandt E., and Thompson F. (eds), *The Prokaryotes – Prokaryotic Communities and Ecophysiology*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 123-138.

Teske A., Hinrichs K.-U., Edgcomb V., de Vera Gomez A., Kysela D., Sylva S.P., Sogin M.L., and Jannasch H.W. (2002) Microbial diversity of hydrothermal sediments in the Guaymas Basin: Evidence for anaerobic methanotrophic communities. *Appl. Environ. Microbiol.* **68**, 1994-2007.

Teske A. and Sørensen K.B. (2008) Uncultured archaea in deep marine subsurface sediments: have we caught them all? *ISME J.* **2**, 3-18.

Teske A.P. (2006) Microbial Communities of Deep Marine Subsurface Sediments: Molecular and Cultivation Surveys. *Geomicrobiol. J.* **23**, 357-368.

Whiticar M.J. and Faber E. (1986) Methane oxidation in sediment and water column environments – Isotope evidence. *Adv. Org. Geochemistry* **10**, 759-768.

Yanagawa, K., Ijiri, A., Breuker, A., Sakai, S., Miyoshi, Y., Kawagucci, S., Noguchi, T., Hirai, M., Schippers, A., Ishibashi, J., Takaki, Y., Sunamura, M., Urabe, T., Nunoura, T., and Takai, K. (2016) Defining boundaries for the distribution of microbial communities beneath the sediment-buried, hydrothermally active seafloor. *ISME J* doi:10.1038/ismej.2016.119

Yanagawa K., Nunoura T., McAllister S.M., Hirai M., Breuker A., Brandt L., House C.H., Moyer C.L., Birrien J.-L., Aoike K., Sunamura M., Urabe T., Mottl M.J., and Takai K. (2013) The first microbiological contamination assessment by deep-sea drilling and coring by the D/V Chikyu at the Iheya North hydrothermal field in the Mid-Okinawa Trough (IODP Expedition 331). *Front. Microbiol.* **4**: 327.

Chapter 3

Temperature-dependent taxonomic and functional changes in microbial communities through a hydrothermal gradient in Okinawa Trough sediments

Abstract

High temperature marine sediments represent an analogue for studying microbial life deep into the subsurface and hold promise into understanding life at the biotic fringe. Marine sediments from Site C0014, IODP Expedition 331 (Iheya North Hydrothermal Field, Okinawa Backarc Basin), provide an ideal profile with which to study potential temperature-dependence and stratification of microbial niches with depth due to its a strong temperature gradient from a local hydrothermal system. The first several meters of sediments resemble those from continental margin sites, but transition to hydrothermally altered clay below 10 meters below sea floor (mbsf). In this study, we show functional and metabolic evidence from six metagenomes (1.32, 3.97, 5.39, 8.84, 12.99, and 15.30 mbsf) that capture a distinct changeover of microbial assemblages between the mesophilic horizons above 10 mbsf and the hottest (55°C) hydrothermal clay horizon. Notably, hyperthermophile-specific reverse gyrase genes are found in only the 15.30 mbsf sample, suggesting an established high temperature adapted community in the deepest horizon of the biosphere. In addition, evidence for both a thermophilic methanotrophic ANME-1 in the same horizon reflects the dynamic nature of this hydrothermal system in which a hyperthermophilic microbial community may have once been better established under different environmental conditions. Taxonomic results from a previous study of these sediments support a model in which ubiquitous marine subsurface taxa persist through the marine mud until punctuated abundances of (hyper)thermophilic archaeal taxa extend into hydrothermally altered clay horizons. Our results presented here provide additional functional information to this model,

demonstrating that the molecular signals represent a responsive microbial community to the increasingly demanding environmental conditions.

Introduction

The discovery of life at hydrothermal vents in the 1970s has since motivated many scientific efforts to explore and understand microbial life and diversity within seafloor environments. The deep subsurface biosphere represents a frontier for the discovery of new microbial life and the extent and versatility under a range of limiting conditions. Due to the challenges and extreme costs of reaching the seafloor, along with low cell concentrations and high proportions of uncultivated microbial lineages, available taxonomic and functional information with respect to marine subsurface biogeography is still very limited (Martino, 2014). Investigations to date suggest that microbes found in the far-reaching seafloor appear to be only distantly related to those identified from surface environments (Inagaki *et al*., 2006, Sorensen *et al*., 2004, Lipp *et al*., 2008). Cultivation studies are accurate approaches to understand an organism's genome and behaviors under controlled conditions; however, the majority of microbes on Earth evade cultivation efforts. Thus, cultivation-independent sequencing of DNA directly from an environmental sample holds great promise and has provided the scientific community with much of its information to date.

Among the methods to gain insight into physiology and genetics of uncultured representatives, metagenomics, the genomic analysis of an assemblage of organisms, has emerged as a powerful tool (Handelsman, 2004). The advent of metagenomics and other "omics" has helped to resolve or circumvent, to an extent, the prerequisite of cultivation to gain full access to the genetic information of individual organisms (Cowan *et al.*, 2015). Studies that focus on specific regions of taxonomic marker genes rely on clone library technology or primer-based

polymerase chain reaction amplifications. It has been shown that many mismatch frequencies in standard archaeal or universal amplification primers bias against many novel phylogenetic lineages of marine subsurface archaea (Teske and Sørensen, 2008). Because of such biases, it has been proposed that a metagenomics analysis produces the most accurate and quantitative perspective of a microbial population (von Mering *et al*., 2007). With the transition of next-generation sequencing technology from Sanger-based sequencing, a number of new sequencing platforms now provide an avenue for millions of DNA strands to be sequenced in parallel, producing a substantially greater throughput with a greatly reduced cost per sequencing yield. Thus, the ability to obtain massive sequencing datasets more efficiently and cheaply has revolutionized not only genomic research, but also the study of unknown biospheres.

One of the systematic approaches in studying microbial communities from seafloor environments has been to comparatively analyze trends in taxonomic or metabolic abundance and diversity with thoroughly measured environmental properties (*e.g.* overlying surface productivity, organic matter deposition, sulfate reduction rates, or porewater geochemistry profiles) and interpolate or make predictions on a more global scale. While many studies have attempted to correlate microbial community composition among sites varying in oxygen content, lithology, and organic matter content (Inagaki *et al.*, 2003; Inagaki *et al.*, 2006; Webster *et al.*, 2007; Biddle *et al.*, 2011), the suggested relationships have not been consistent between studies or sample sites. Even fewer studies have attempted to examine the biogeography of microbial communities at the biotic fringe, or across gradients of that may become increasingly difficult for microbial survival and tenacity. With the initial excitement of life at high temperature hydrothermal vent systems and the potential for chemosynthetic or other extremophilic life, much of the research that emerged from hydrothermal vent systems emanate from hydrothermal vent fluids. Few studies have since focused on hyperthermophilic life in the sedimentary subsurface, and to what extent the biosphere exists under such conditions has yet to be constrained. The Parkes *et al.*, 2000

survey of Juan de Fuca Ridge hydrothermal sediments was one of the first to lay the foundation for exploring hyperthermophilic communities in marine sediments as a proxy for a temperature limit and biotic fringe of subsurface life. While intact cells were found in horizons corresponding to 180°C, Parkes and colleagues relied solely on cell enumerations, and no studies have since followed up with a more extensive, molecular approach. Another set of studies from the hydrothermal sediments of Guaymas Basin, Gulf of California closely examined the methane-oxidizing communities across centimeter-scale temperature ranges (*e.g.* Teske *et al.*, 2002; Holler *et al.*, 2011; Biddle *et al.*, 2012; McKay *et al.*, 2012; Meyer *et al.*, 2013). However, Guaymas Basin represents a unique geologic setting because of its very high organic matter content and accelerated pyrolysis into petroleum-like compounds due to extreme hydrothermal fluid fluxes. Thus, such communities represent localized, metabolically specialized extremophiles that may not be characteristic across other high temperature regimes across the seafloor.

The Iheya North Hydrothermal Field in the Okinawa Backarc Basin represents an ideal environment in which to investigate the biotic temperature fringe of microbial life at depth because of its subsurface hydrothermal activity within its continental margin-type sediment profile. Geographically, the Okinawa Backarc Basin is situated along a continental margin, which is a sediment profile type commonly sampled and studied across the seafloor (*e.g.* Peru Margin, Costa Rica Margin, Cascadia Margin). The hydrothermal network within the subsurface here supplies an additional temperature obstacle to microbial life existing in the sediments. In particular, Site C0014 from IODP Expedition 331 lies approximately 450 m away from the active hydrothermal vent and has an estimated temperature gradient of 3°C/m with a surface temperature resembling that of typical continental margin sediments (~4°C). Thus, the moderate temperature gradient at Site C0014 sediments is gradual enough for the establishment of distinct, temperature-dependent communities ranging from mesophiles to (hyper)thermophiles. In this study, we present taxonomic and functional data from six metagenomes spanning the

hydrothermal gradient (1.32, 3.97, 5.39, 8.84, 12.99, and 15.30 mbsf) that begins to shed light on the relationship between temperature and microbial life deep in the subsurface as temperature becomes increasingly limiting. Previous taxonomic results from these same sediments (Chapter 2) support a general model in which marine subsurface taxa persist into the subsurface at Site C0014 until they are replaced by punctuated abundances of high temperature archaeal taxa in the hydrothermal clay horizons. The metagenomic evidence presented here from several data mining approaches continues to build upon this model in which a temperature-dependent stratification is observed through the hydrothermal gradient.

Experimental Procedures

See Chapter 2 "Experimental Procedures" for Sample Collection and Extraction of DNA. Specific samples (C0014B) for downstream DNA sequencing are presented in Appendix A and are based on visible amplification product on a gel electrophoresis.

Whole Genome Amplification and Sequencing

Because DNA yield from extractions was likely too low for the sequencing chemistry, a whole genome amplification (WGA) was implemented via multiple displacement amplification (MDA; REPLI-g Mini Kit, Qiagen Inc.), according to the manufacturer's instructions. Table B-1 shows the initial sequencing plan, amplification variations, and final sequencing yield. The reaction was carried out using a 5 hour extension time at 30°C for most samples, along with a negative control, to ensure that the amplification reagents did not yield any product. After stopping the reaction as instructed, products were visualized with gel electrophoresis on a 1% agarose gel. Product was visible in most samples except C0014B-1-3, C0014B-1-4, C0015B-1-1.

Duplicate samples C0015B-1-1 and C0014B-2-7 with an 8 hour extension were also submitted. On average, the concentration of double stranded DNA was 0.2-11 ng/µl, as measured by the sequencing facility on a Qubit® fluorometer.

Sequencing was performed at the Penn State Genomics Core Facility – University Park, PA using the Illumina® HiSeq 2500 (NSF-MRI award DBI-1229046 (Axtell *et al.*, 2012)). The sequencing facility prepared DNA libraries using Nextera XT Library Preparation Kits prior to sequencing. Together, the samples were run on one-half of a sequencing plate. This sequencing run was run using rapid-run model that averages up to 300 million single reads, or 600 million paired reads of ~150 nucleotides per rapid-run.

Pipelines for Metagenomic Data Analysis

(1)

For initial assessments, unassembled fastq files were uploaded and paired through the MG-RAST pipeline, using default parameters. The assembled fasta contig files with appended coverage information were also uploaded and put through the MG-RAST pipieline. The analysis tool on this server was then used to gather subsequent taxonomic information. Taxonomic identifications were collected from the annotations of identified protein and rRNA features. Data were compared to the M5NR (protein) or M5RNA (rRNA) databases using the MG-RAST standard analysis parameters using the Best Hit Classification option.

(2)

Thijs Ettema and Jimmy Hser Wah at Uppsala University, Department of Cell and Molecular Biology; Biological Evolution performed metagenomic assembly and gene prediction. Data assembly was performed with Ray Meta de novo assembly program using a k-mer of 65 for all samples. See Table B-1 for assembled contig yield. Gene prediction used Prodigal gene predictor pipeline (Hyatt *et al.*, 2010) with assembled contig data. The amino acid gene-finding output files were used as a database (using the makedb BLAST function) for subsequent protein BLAST (Altshul *et al.*, 1997) "gene fishing" searches with amino acid sequence genes-of-interest from the NCBI database as a query. We then mapped the reads of the BLAST hits back to the contigs to get read coverage, which was calculated as follows: nucleotide contig (fasta) files and their corresponding fastq files were uploaded to Penn State's Hammer network server, which has 288 processor cores, 12 processor cores per server, and 48 GB memory per server; using the Burrows-Wheeler Alignment Tool (bwa) (Li and Durbin, 2009), contigs were indexed and short reads from the fastq were aligned with the contigs; the subsequent paired end file was converted to a sam file; using SAMtools (Li *et al*., 2009), the sam file was converted to a bam file and the bam file was sorted and indexed; the *depth* command generated a position based on the coverage profile, and that output was used to generate a single average coverage value for each contig using the "calc.coverage.in.bam.depth.pl" perl script (provided by Albertsen *et al*., 2013 at http://madsalbertsen.github.io/multi-metagenome/docs/faq.html).

Phylogenetic trees were made by aligning and trimming sequences in MEGA using the ClustalW algorithm default parameters. The alignment exported to PAUP to produce maximum likelihood trees. In most cases, default parameters were used and the branch-and-bound option applied with random start values.

(3)

Only one metagenomic sample was processed for binning. The assembled contigs file was uploaded to Metawatt 1.7 All default parameters were used for processing, including binning with tetranucleotides. The bins were downloaded as nucleotide fasta files and were annotated through the Prokka annotation program (Seeman, 2014). This step was done with the help of Rosa Leon, a postdoc from Dr. Jennifer Biddle's lab, University of Delaware (Lewes) because their computing system had all downloaded software necessary to run Prokka. CheckM (Parks *et al.*, 2015) was also run to assess the quality of the binned metagenome. Marker genes identified at this stage from CheckM were used in a BLASTP search against the non-redundant protein database from NCBI to produce taxonomic hits associated with each one. The BLASTP output files were then uploaded to the MEtaGenomics Analyzer (MEGAN) (Huson *et al.*, 2007) for visual taxonomic parsing. LCA parameters were changed so that the Top Percent was 5% and the LCA percent was 50%. In addition, the Prokka amino acid sequence files were uploaded to the KEGG annotation system (Kanehisa and Goto, 2000; Kanehisa *et al.*, 2016). The output text file was used in the KEGG mapper for metabolic pathway visualization.

Results and Discussion

General Taxonomic Changes through the Hydrothermal Gradient

Using the taxonomic results from Chapter 2 as a model with which to forecast potential differences among metagenomes, we used MG-RAST gene annotations as a first approach to survey the general taxonomy. Both unassembled and assembled reads were processed through the MG-RAST pipeline and compared against one another in order to assess potential biases or differences in metagenome assembly. Generally, the data classified at the domain level show an

overall increase in the relative abundance of represented archaeal sequences, relative to bacterial sequences, with increasing depth (Figure 3-1). There is also agreement in the relative abundances between assembled and unassembled data of each sample – there is a maximum of 20% variability of domain relative abundance classifications between assembled and unassembled datasets, and there is no significant difference in the percent GC content between samples (not shown). While a significant amount of unassembled sequences did not pass the quality control in the MG-RAST pipeline (Appendix B-2), the agreement between datasets does not indicate any conspicuous discrimination or loss of either archaeal or bacterial information at the domain level. Similarly to the taxonomic marker gene assessment from Chapter 2, archaeal sequences are observed in higher relative abundances below 10 mbsf than in shallower horizons. The metagenomic data, however, do not exhibit as drastic a change below 10 mbsf.

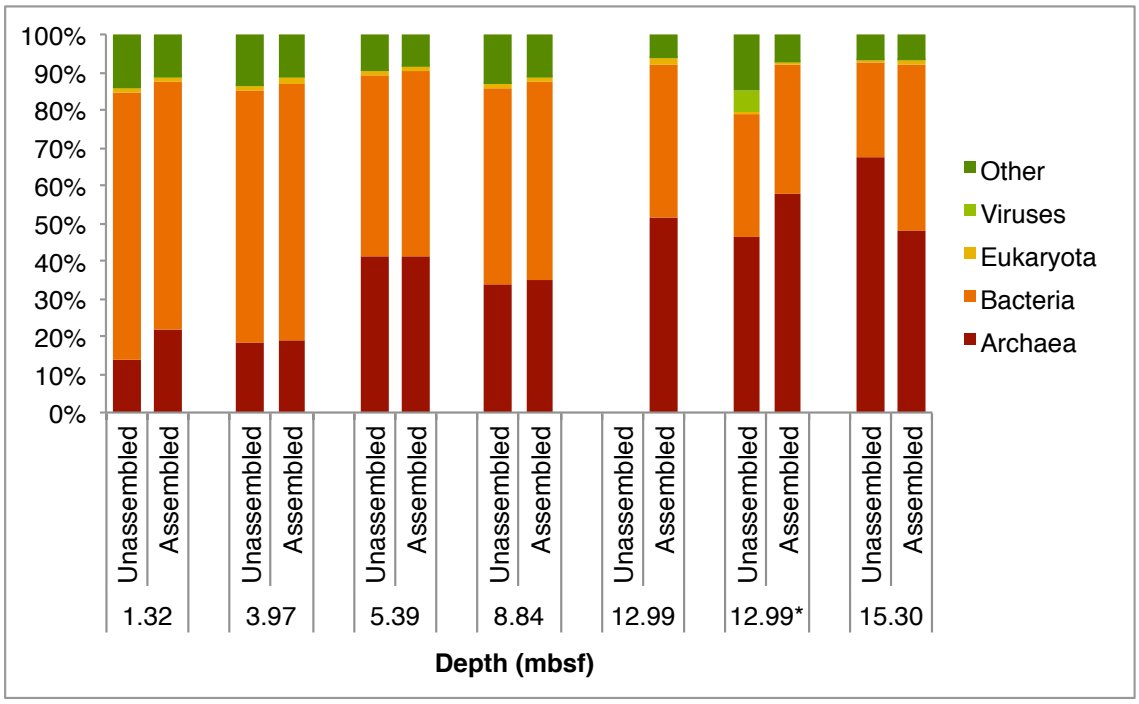

Figure **3-1**: Taxonomic composition classified at the domain based on comparison of metagenomic sequence data to the M5NR database via MG-RAST. Shown are the percentages of total classified reads for each taxonomic grouping per sample analyzed. "Other" refers to sequences that made it through quality control, but were not confidently classified or assigned within one taxa. Each sample depth was analyzed with both unassembled and assembled data

using the same MG-RAST parameters.

A higher taxonomic resolution of the metagenomic data reveals further consistency between the assembled and unassembled sequences (Figure 3-2). Many phyla, such as Proteobacteria, Firmicutes, and Euryarchaeota that are present across all depths correlate well between the assembled and unassembled datasets. Furthermore, both datasets exhibit a shift in taxa with depth, especially in the two horizons below 10 mbsf, where there is a stepwise change in porewater geochemistry and clay composition (Appendix A-1(F)), suggesting a response in the community to this hydrothermally altered environment. For example, Figure 3-2 shows a similar proportion of Chloroflexi in the four shallowest horizons followed by a noticeable decrease in representation in the 12.99 and 15.30 mbsf horizons. The phylum Chloroflexi represents a group of highly abundant and cosmopolitan taxa in deeply buried marine sediments that derives energy through the exploitation of organohalide compounds (Biddle *et al.*, 2012; Seshadri *et al.*, 2005). The abrupt decrease in abundance through the denoted hydrothermal transition indicates a physical and/or chemical boundary that is likely inhibitory on the persistence of a prolific Chloroflexi community. Similarly, other bacterial representatives generally decrease at 12.99 and 15.30 mbsf as archaeal representatives increase. Of the Cren- & Thaumarchaeota, in particular, a large proportion of sequences at 12.99 and 15.30 mbsf are further classified down to the Thermoprotei class level (not shown), whose members are generally known to inhabit environments such as hot springs and hydrothermal vents. The heightened Thermoprotei taxa at depth 12.99 mbsf would suggest a potential thermophilic niche in the deeper, hotter horizons at Site C0014. Interestingly, the greatest relative abundance of Cren- & Thaumarchaeota appears at 12.99 mbsf, rather than 15.30 mbsf. This observation will be discussed later, where an observed spike in presumed thermophilic methanotrophic taxa at 15.30 mbsf has dampened the signal of other taxa. Overall, there is general agreement in taxonomic trends between both the

metagenomic data and the 16S rRNA gene amplicon surveys (Chapter 2), which suggests minimal bias in either approach.

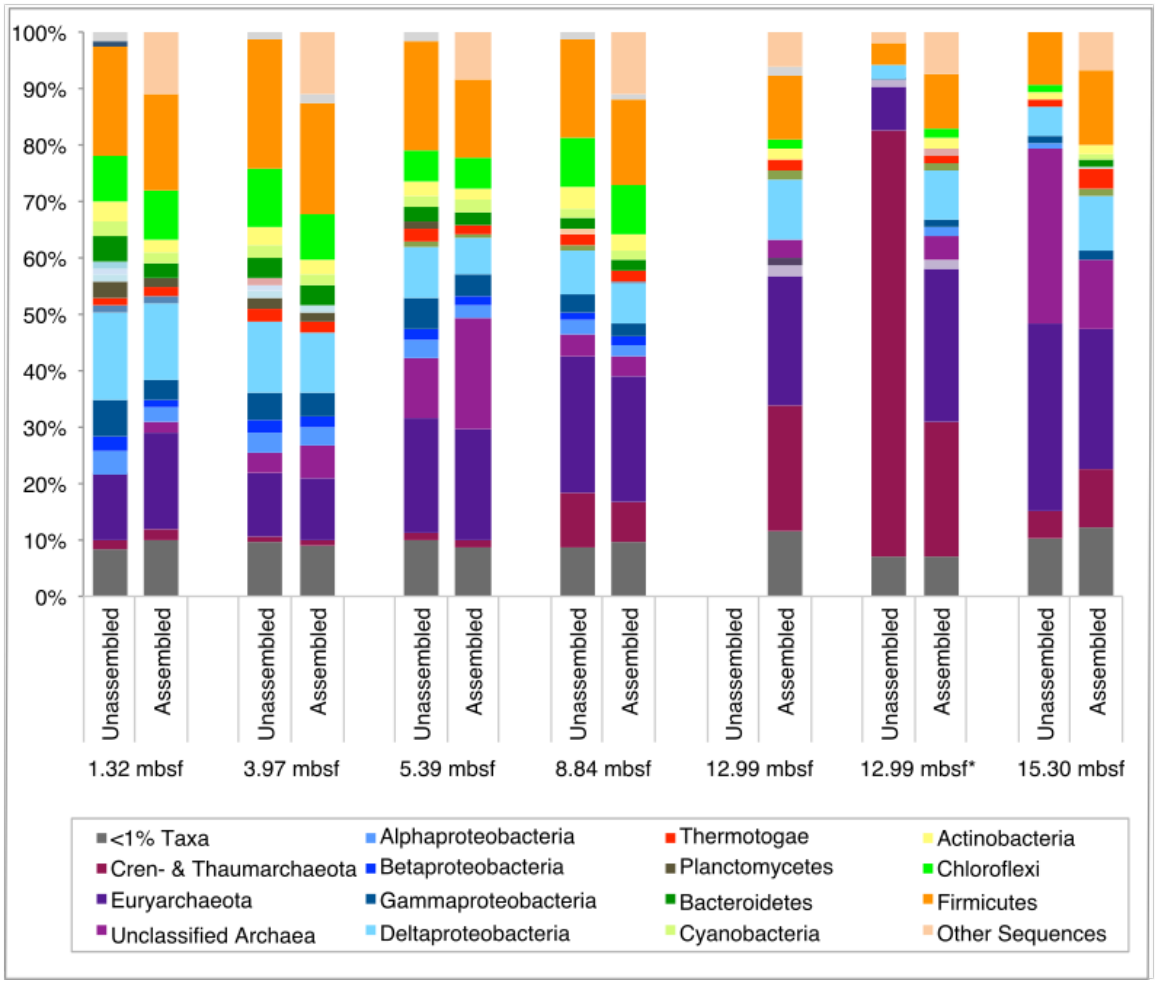

Figure **3-2**: Taxonomic composition classified at the phylum level (or in the case of Proteobacteria, class level) based on comparison of metagenomic sequence data to M5NR database (MG-RAST). Shown are the percent relative abundances of total classified reads for each phylum. Phyla that did not have at least a 1% representation of the total number of sequences in that sample were grouped into "<1% Taxa." Each sample depth was analyzed using both unassembled and assembled data using the same MG-RAST parameters.

High Temperature Adapted Communities Detected through Functional Genes

The taxonomic classifications from Figure 3-2 suggest a general shift in the microbial community composition from shallow, cooler sediments at Site C0014 through the hydrothermal gradient between ~10-15 mbsf. However, it is unclear from Figure 3-2 whether these microbial communities are functionally different or simply a relict of burial from the sediment surface. To explain the taxonomic diversity along this gradient, we have employed BLAST searches to identify certain genes with inherent function among specific taxonomic populations. The following sections will discuss the results from Table 3-1, which shows the results of BLAST searches of specific genes among each metagenome.

Table **3-1**: Table listing each metagenomic sample by depth and the number of contig hits of each gene from a BLASTP search. The numbers in parentheses represent the contig coverage, or range of contig coverage for each BLASTP search. The sample marked (*) refers to the replicate metagenome sample with an extended whole genome amplification time..

	Protochlorophyllide	*RdhA*	*Dihydrouridine Synthase*	*Reverse Gyrase*	*mcrA*	*Nitrogenase*
C0014B-1-1 (1.32 mbsf)	0	8 (8-52)	5 (4-50)	0	1 (82)	1 (64)
C0014B-1-3 (3.97 mbsf)	0	5 (12-23)	1 (32)	0	7 (10-31	3 (7-77)
C0014B-1-4 (5.39 mbsf)	0	4 (23-125)	1 (54)	0	6 (19-69)	7 (10-43)
C0014B-2-3 (8.84 mbsf)	0	4 (13-69)	0	0	0	2 (13-48)
C0014B-2-7 (12.99 mbsf)	0	0	0	0	0	0
C0014B-2-7* (12.99 mbsf)	0	0	0	0	0	0
C0014B-2-10 (15.20 mbsf)	0	0	1 (25)	2 (57-118)	1 (1898)	2 (41-2220)

Protochlorophyllide Reductase and Reductive Dehalogenase

To determine the extent to which relict DNA contributes to molecular signals within the environment, we compared two genes that help to decouple the DNA signals from allochthonous and indigenous marine subsurface bacteria. Protochlorophyllide reductase is an enzyme involved

in the biosynthesis of chlorophyll precursors. Organisms that require these large biomolecules use sunlight and reside in photic habitats on land or surface waters. Therefore, a presence of this gene in the deep, marine subsurface would suggest the deposition, burial, and preservation of such surface life DNA, as it is improbable that active photosynthetic life persists in the deep subsurface biosphere. Even though many marine subsurface studies have found genes relating to Cyanobacteria taxa (*e.g.* Mills *et al.*, 2012), the significance of these molecular signals is presently unknown. Based on unsuccessful culturing attempts of phototrophs from such sediments (Martino, 2013), it is most likely that these molecular structures have merely evaded considerable degradation in sediments. BLAST searches for protochlorophyllide, as well as other photosynthesis genes identified in Partensky *et al.*, 1999, did not return any hits to photosynthetic genes. Specifically, returned BLAST results yielded hits to nitrogenase genes, all of which corresponded to non-photosynthetic organisms. These results are somewhat surprising, as previous 16S rDNA results showed taxonomic evidence indicative of relict plan material recovered from meters into the subsurface. The lack of recoverable data in this study indicative of photosynthetic life suggests that any extant plant or surface DNA deposited to the seafloor may be too degraded to produce robust assembled contigs. With respect to molecular signals contributing to the Site C0014 sediment samples, those from relict plant and surface life appear to be negligible in relation to those from marine subsurface prokaryotic life and help to describe the high-energy nature of this environment regarding long-term biomolecule preservation.

In contrast to allochthonous molecular signals, reductive dehalogenase (*Rdh*) is a key functional enzyme in dehalorespiration, a process that employs halogenated organic compounds as terminal electron acceptors, and is commonly found in marine sediments. Halogenated compounds are generally recalcitrant and effectively buried in marine sediments, but are used by dehalorespiring bacteria, such as those from genus *Dehalococcoides*. In many 16S rRNA gene surveys from marine sediments, Chloroflexi are one of the most frequently detected phyla, and

many sequences within these Chloroflexi are closely related to *Dehalococcoides*. Furthermore, reductive dehalogenase homologous genes related to those sequences from *Dehalococcoides* have been documented in marine margin and open ocean sediments and suggest that dehalorespiration represents an ecologically important biogeochemical process in the marine subsurface (Futagami *et al.*, 2009). As a signal for an indigenous microbial population in Site C0014 marine sediments, we searched for *Rdh* as a proxy to determine the extent of an indigenous biosphere and the extent of dehalorespiration in the hydrothermal subsurface. Table 3-2 shows that *Rdh* was found in the metagenomes from 1.32, 3.97, 5.39, and 8.84 mbsf, but was not detected in the two deeper horizons associated with hydrothermal clay. The restriction of *Rdh* to the upper, marine clay horizons suggests that the lithologic/geochemical change below this depth (T = 15-22°C) represents a physical boundary in which the conditions have become unfavorable for the survival of a dehalorespiring community. Prior to a previous high temperature event that produced this hydrothermal clay section *in situ*, a prolific subsurface Chloroflexi community may have once extended deeper. However, a pulse of high temperature fluid through this sediment section likely extinguished many organisms not capable of enduring such conditions and any molecular evidence that they once existed. Even though the current measured temperatures in this core section (10-15 mbsf) are within the mesophilic and thermophilic regimes, it is possible that organohalide molecules have become chemically altered and no longer bioavailable here from such an event, which has prevented a dehalorespiring community from recolonizing. Together, the data from both protochlorophyllide reductase and reductive dehalogenase genes have contributed to an initial perspective of the extent of the microbial ecosystem downcore at Site C0014 and the potential contribution of allochthonous molecular signals.

tRNA-Dihydrouridine Synthase and Reverse Gyrase

To test functionally whether or not hyperthermophilic taxa are present at depth, we searched within all metagenomes for tRNA-dihydrouridine synthase and reverse gyrase genes, whose absence and presence, respectively, have direct implications for finding high temperature adapted life. In a wide range of life, the posttranscriptional incorporation of 5,6-Dihydrouridine into tRNA by a tRNA-dihydrouridine synthase increases its conformational flexibility at low temperatures (Dalluge *et al.*, 1997; Noon *et al.*, 2003; Saunders *et al.*, 2003; Cavicchioli, 2006). It has been observed that dihydrouridine is generally absent from hyperthermophilic prokaryotes and has also been found at higher levels in some psychrotolerant Archaea relative to other Archaea (Noon *et al.*, 2003). Thus, we anticipate a marked absence of tRNA-dihydrouridine synthase in the metagenomes whose sequences come from a predominantly (hyper)thermophilic microbial community. In contrast, reverse gyrase is the only protein found exclusively in hyperthermophilic organisms to date (Forterre, 2002). Functionally, reverse gyrase is responsible for 1) positively supercoiling DNA to prevent excess local unwinding of the double helix at high temperatures (Déclais *et al.*, 2001, Forterre, 2002), and 2) recognizing nicked DNA and recruiting a protein coat to the site of damage through cooperative binding at high temperatures (Kampmann and Stock, 2004). Thus, the presence of reverse gyrase genes is an indication of which horizons host hyperthermophiles.

The results for the tRNA-dihydrouridine synthase gene BLASTP search show multiple hits in the surface site C0014B-1-1, followed by one hit in each of the two subsequent depths, and one hit in the deepest sample (Table 3-1). Within C0014B-1-1, one hit had a 97% identical sequence to an uncultured archaeon from a study investigating the anaerobic oxidation of methane in Hydrate Ridge cold seep sediments (Meyerdirks *et al.*, 2005). The other hits within C0014B-1-1 were to a sulfate-reducing bacterium within the Desulfobacterales order, an

anaerobic ammonia-oxidizing (anammox) bacterium within the Planctomycetes phylum, a representative within the Alphaproteobacteria class, and a representative within the Clostridia class, all with 41-55% sequence identity. The hits returned from the deeper horizons produced low sequence identity hits to a sulfate-reducing bacterium, a representative from the Caldiserica phylum, and a representative from the Clostridia class (3.97, 5.39, and 15.30 mbsf horizons, respectively). At first glance, the tRNA-dihydrouridine synthase sequence hits to most horizons throughout the sediment profile suggest that this gene may not represent a sensitive proxy for detecting changes in temperature-dependent adaptations. However, previous studies have found 16S rRNA gene sequences classified within Clostridia from drilling mud samples and extraction blanks (Yanagawa *et al.*, 2013 and Appendix A). Thus, amplicon or metagenomic sequences classified alike (*i.e.* the tRNA-dihydrouridine synthase hit from 15.30 mbsf) should not be confidently considered part of the indigenous population. Excluding the Clostridia representative from 15.30 mbsf, tRNA-dihydrouridine synthase appears to be restricted to the upper, unaltered sediment horizons (c.a. 8-20°C) and absent in the hotter, hydrothermal clay horizons. These results support a trend in which functional requirements for tRNA flexibility are more imperative in psychro- and mesophilic conditions than in hotter horizons.

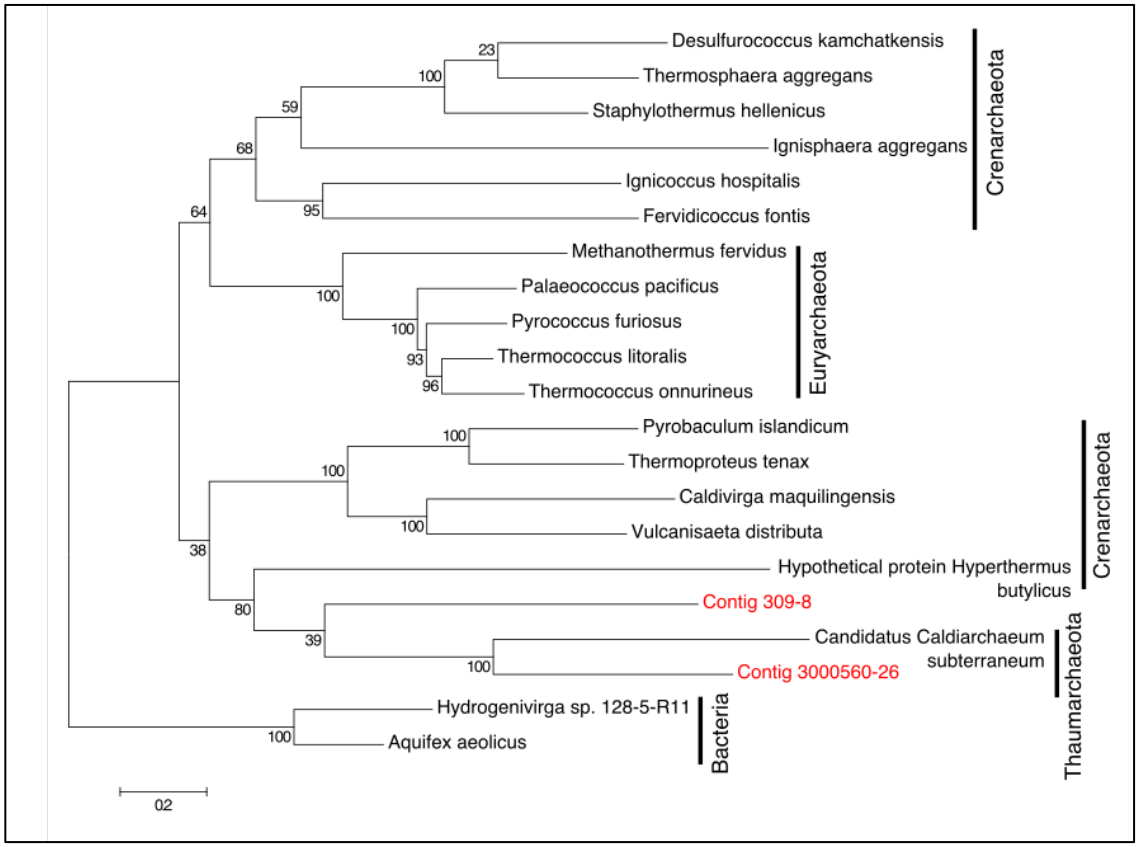

Figure **3-3**: Maximum likelihood tree of reverse gyrase protein sequences produced with MEGA default parameters, referenced at amino acid positions 9-1,049 of Candidatus Caldiarchaeum subterraneum. The two contigs found in the C0014B-2-10 metagenome are indicated in red. Bootstrap values are displayed at the branch bifurcations. The archaeal class is denoted by the vertical bars on the right side.

The BLASTP search of reverse gyrase genes produced two contigs hits from only the 15.30 mbsf horizon (Table 3-1). The observed contigs found exclusively in the deepest sample suggest that a hyperthermophilic microbial community exists in this 55°C temperature horizon, and is not likely a relict of endured burial. Both contigs were aligned with other published reverse gyrase genes to determine the taxonomic identity of these organism(s). In Figure 3-3, full-length reverse gyrase gene alignments of published sequences were generated as a base topology constraint on which to map the two contig sequences from the C0014B-2-10 metagenome. Among both bacterial and archaeal reverse gyrase sequences represented in the tree, Contigs 309-

8 and 3000560-26 appear to have an archaeal origin, clustering closest to Candidatus Caldiarchaeum subterraneum of the Thaumarchaeota class. To further resolve the identity of these contigs, the alignment of these sequences were visually compared to a sequence representative from each of the Euryarchaeota, Thaumarchaeota, and Crenarchaeota classes (Figure B-2). Highlighted regions along the alignment in Figure B-2 represent insertions or deletions (indels) of each contig and are described by the reference sequence(s) with a corresponding indel. Within Contig 3000560-26, most of its indels corresponded to those of Candidatus Caldiarchaeum subterraneum than any other reference sequence (Table 3-2). The agreement between these two sequences supports the placement of Contig 3000560-26 in Figure 3-3 and suggests that Contig 3000560-26 likely represents a Candidatus Caldiarchaeum subterraneum population. Contig 309-8, however, was comprised of many indels that did not correspond to any of the other sequences. For example, six of its 30 indels corresponded to Thermococcus gemmatolerans, while 18 were unique to itself and did not correspond to any reference sequence. To test whether ambiguities in the alignment are a function of variability in either the helicase or topoisomerase domain of the gene, we trimmed and realigned the sequences from Figure 3-3 according to the amino acid positions of the aforementioned domains (Figures B-3 and B-4). In Figures B-3 and B-4, the placement of Contig 300056-26 with Candidatus Caldiarchaeum subterraneum remains unchanged, which agrees with the previous comparisons of Contig 3000560-26 to Candidatus Caldiarchaeum subterraneum. In the maximum likelihood tree of the helicase domain in Figure B-4, however, Contig 309-8 does not cluster with Candidatus Caldiarchaeum subterraneum and indicates variability in this region of the sequence alignment. Furthermore, Contig 309-8 appears unique in that it does not cluster with any reference sequence, suggesting that Contig 309-8 may altogether represent a reverse gyrase sequence not currently documented in protein databases.

Table **3-2**: The number of insertion-deletions that Contigs 300056-26 and 309-8 have in common with four reference sequences when the six full-length sequences are aligned together. See Figure B-2 for visual full-length alignment. The Unique column refers to indels that did not correspond to any reference sequence. The Total Indels are lower than the sum of the indels listed because three positions on Contig 300056-26 corresponded to both Candidatus Caldiarchaeum subterraneum and Methanothermus fervidus; three positions on Contig 309-8 corresponded to both Hyperthermus butylicus and Thermococcus gemmatolerans.

Contig	Thaumarchaeota	Crenarchaeota	Euryarchaeota		Unique	Total Indels
	Candidatus Caldiarchaeum subterraneum	Hyperthermus butylicus	Methanothermus fervidus	Thermococcus gemmatolerans		
3000560-26	18	0	3	0	1	19
309-8	2	4	3	6	18	30

In addition to analyzing the sequence alignment of Contig 309-8, we also examined BLAST hits of the other genes on the same assembled contig to better resolve the taxonomic identity of this hyperthermophilic population (Figure 3-4). Figure 3-4 shows a compilation of top BLAST hit classifications from all predicted genes on Contigs 309 and 3000560. Nearly 40% of genes on Contig 3000560 produced top BLAST hits to Candidatus Caldiarchaum subterraneum, which is continued support of Contig 3000560 representing a hyperthermophilic Thaumarchaeota population in the C0014B-2-10 horizon. Contig 309 is comprised of fewer gene sequences, where Contig 309-8 represents only one of 11 recognized genes on assembled Contig 309 (Table B-2). Unlike Contig 3000560, none of the BLAST hits from Contig 309 predicted genes corresponded to Candidatus Caldiarchaeum subterraneum, though nearly half of the top hits were classified within the Thaumarchaeota. Among the predicted genes on Contig 309, Contig-11 represents a reverse gyrase gene that was not detected from the previous gene search (Table B-2). The top BLAST hits of Contig-11 and Contig 309-8 correspond to taxa within different archaeal classes. However, when the two contigs are aligned with other reference reverse gyrase gene sequences, Contig 309-11, though shorter, is nearly identical to Contig 309-8. Thus, it is likely that the sequences described by the nearly-identical Contigs 309-8 and 309-11 represent a novel reverse

gyrase gene not documented in protein databases. Ultimately, the two distinct reverse gyrase sequences presented and discussed here indicate two potentially different hyperthermophilic populations constrained within the 15.30 mbsf horizon.

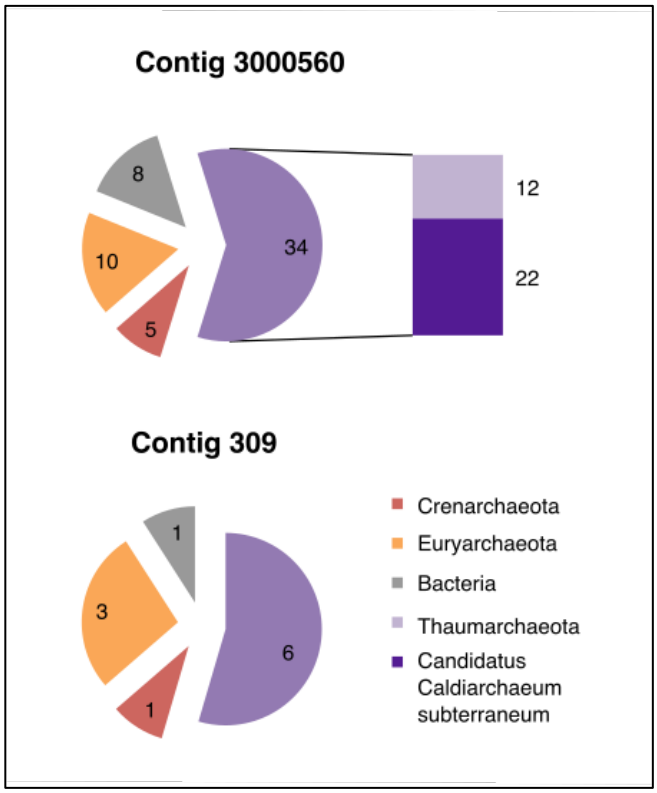

Figure **3-4**: Taxonomic classification of genes on Contig 3000560 and 309. The numbers on each pie slice represent the number of genes within the taxonomic group. In both cases, the purple pie section represents Thaumarchaeota, which was further classified down to Candidatus Caldiarchaeum subterraneum in Contig 3000560.

The 16S rDNA taxonomic analysis discussed in the previous chapter captured a high relative abundance of Terrestrial Hot Spring Crenarchaeotic Group (THSCG, of the Thaumarchaeota) sequences in the C0014B-2-7 sediment horizon, with a lower relative abundance in horizon C0014B-2-10. Though members of this taxonomic group have not been studied in pure culture, the documented 16S rRNA gene sequences comprising this clade come from a 1-10 cmbsf, 100°C sediment layer within a middle Okinawa Trough hydrothermal field

(Takai and Horikoshi, 1999). Organisms that grow optimally in these upper temperature ranges are operationally defined as hyperthermophiles. Thus, we had anticipated the detection of reverse gyrase genes in the metagenomes from both C0014B-2-7 and C0014B-2-10, potentially associated with THSCG. The surprising absence of reverse gyrase in C0014B-2-7 may be explained by the overall poor sequence coverage (Table B-1). Table B-1 shows that both C0014B-2-7 samples produced a total sequence count within a similar range of all other samples. However, the assembly process produced a very low yield of contigs of greater than 100 nucleotides in length – approximately 4x (C0014B-2-7) and 8x (C0014B-2-7*) less contigs than C0014B-2-10. Ultimately, this low assembly yield greatly reduces the functional information that can be obtained from these horizons. In the C0014B-2-10 horizon, however, we presented evidence of reverse gyrase genes confidently associated with Candidatus Caldiarchaeum subterraneum. Though Contig 3000560 and THSCG, and likely Contig 309, are classified within Thaumarchaeota, the lack of functional and genomic information of THSCG limits our capability to correlate it with the two hyperthermophilic populations found in this study.

Methyl Coenzyme M Reductase (MCR)

A high abundance of thermophilic anaerobic methanotrophic archaeal 16S rRNA gene sequences from sample C0014B-2-10 (Chapter 2) was the motivation for a subsequent investigation of a potential thermophilic methane-oxidizing community within the sediment profile at C0014. A BLASTP search for methyl coenzyme M reductase (MCR), an enzyme expressed by known methanogens and methanotrophs, yielded hits in horizons C0014B-1-1, C0014B-1-3, C0014B-1-4, and C0014B-2-10 (Table 3-1). Among the three sites in the top 5.39 mbsf, contig coverage ranged between 10 and 82, whereas, the coverage calculated from the deepest 15.30 mbsf sample was 1,898 – two orders of magnitude greater than the shallower

samples. This is also tracked by a similar observed trend in nitrogenase gene coverage (Table 3-1). Nitrogenase genes, which enable fixation of atmospheric nitrogen, have been studied in an archaeal and sulfate-reducing bacteria methane-oxidizing consortium (Pernthaler *et al*., 2008, Dekas *et al*., 2009). One study even detected the incorporation of ^{15}N from labeled N_2 in archaeal substituents, showing that certain methanotrophic archaea are also diazotrophs (Dekas *et al*., 2009). The nitrogenase sequence recovered from C0014B-2-10 with the highest contig coverage best identified with an anaerobic methanotroph (ANME) top BLASTP hit, which is consistent with the prolific ANME-1 taxonomic classification from the 16S amplicon dataset. In general, the sizeable spike in MCR and nitrogenase gene abundances in C0014B-2-10, relative to the shallower horizons, is continued functional support for an abundant and predominant niche of methanotrophic archaea inhabiting the deepest horizon of the predicted biosphere.

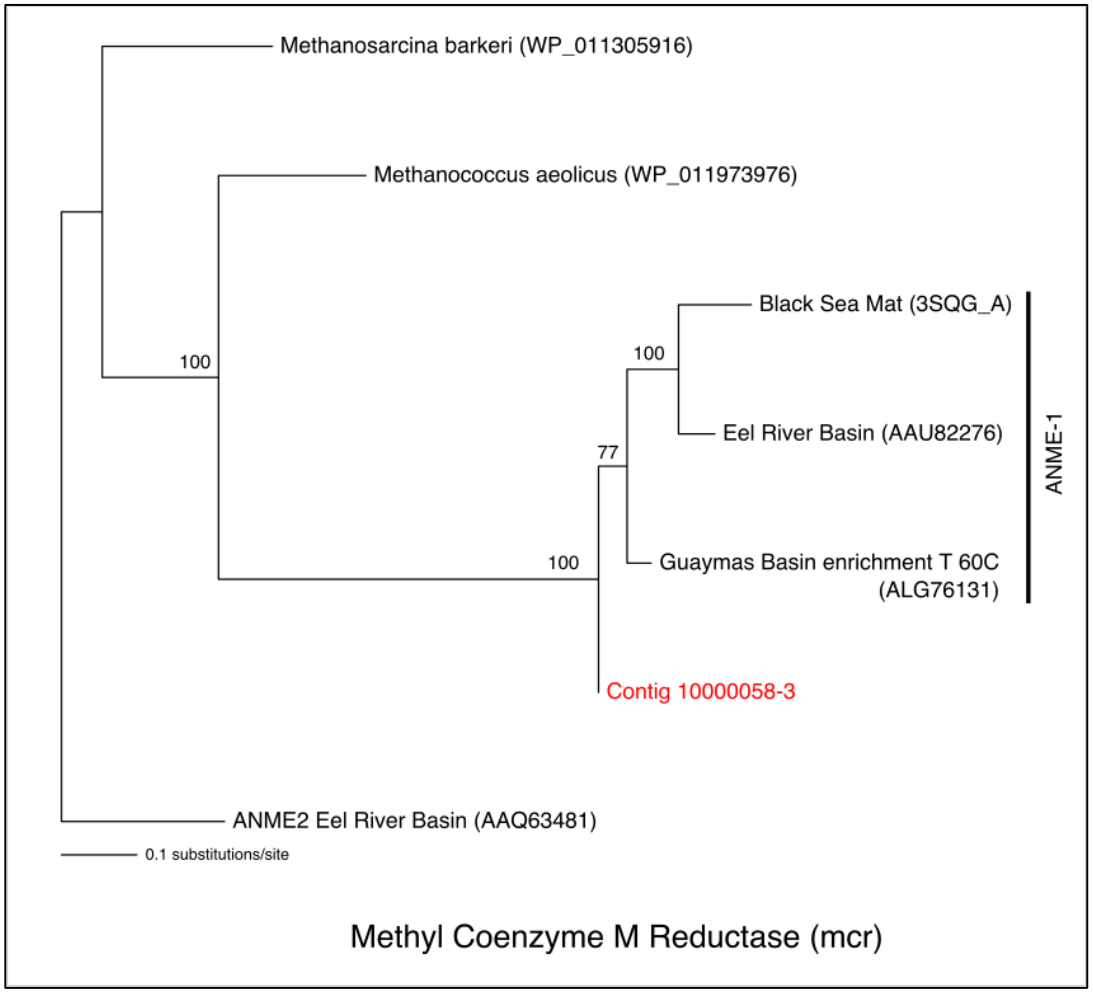

Figure **3-5**: Maximum likelihood tree of *mcrA* amino acid sequences calculated using PAUP branch-and-bound parameters. Bootstrap values are indicated on the branches. The accession numbers for the NCBI protein database are indicated within the parentheses. ANME-1 from the Black Sea and Eel River Basin are cold-seep environments, whereas the Guaymas Basin sample is from hydrothermal sediments. ANME-2 represents the outgroup, while the other species are cultured methanogens.

Additional evidence that the methane-oxidizing community in the C0014B-2-10 horizon is better adapted for higher temperatures comes from an extensive classification of the MCR sequences throughout the sediment profile. Both the sequence alignment and top BLASTP hit (94% identity) for the MCR subunit A (*mcrA*) gene C0014B-2-10 Contig 10000058-3 exhibit a close identity to a thermophilic ANME-1 enrichment (grown at 60°C) (Figure 3-5). The other two ANME-1 *mcrA* sequences from the Black Sea and Eel River Basin in Figure 3-5 come from cold

environments that cluster together. The *mcrA* gene from C0014B-2-10 is also distinct from other depth horizons, indicating that it is not an amplified relict of buried/preserved methanotrophs through the sediment profile. For example, all MCR sequences found from 1.32, 3.97, and 5.39 mbsf corresponded to Eel River Basin (cold seep sediments) methanotrophic archaea or other methanogen species. Therefore, C0014B-2-10 is clearly taxonomically distinct from shallower horizons and indicates that a methane-oxidizing community has recently established in this localized horizon. Furthermore, the contig associated with the C0014B-2-10 *mcrA* gene is the 27th most abundant contig in the entire metagenome. This distinction and intensification of *mcrA* at 15.30 mbsf relative to shallower horizons emphasizes the inherent metabolic capabilities of high-temperature methane-oxidization here as a localized horizon of thermophilic methane-oxidizing community. The absence of similar high temperature ANME-1 representatives in the cooler and shallower horizons at Site C0014 is evidence that the thermophilic, methanotrophic niche at 15.30 mbsf is not a function of long-term endured burial, but rather the establishment of an active and adapted community within its hydrothermal surroundings.

Metagenomic Binning: Metabolic Insight into Individual Microbial Populations

Up to this point, the metagenomic data has provided a general perspective into the sampled community as a whole. Function-identity relationships can be further resolved by binning the data to parse contigs into partial genomes of very closely related populations (Strous *et al.*, 2012). The following sections discuss the binning analysis intended to capture a more complete view of the microbial populations within with the C0014B-2-10 horizon, and particularly those populations associated with the previously identified reverse gyrase and MCR genes. The contigs on which these genes are found coincide with the two most complete bins (Table B-3). A bin identified as a bacterial population with relatively high completeness was also

analyzed, as it may provide another aspect of the microbial community not previously considered. The fourth bin was analyzed as a final attempt to resolve the identity of the organism associated with reverse gyrase gene (Contig 309). Prior to analyzing the functional capabilities of each bin, single-copy marker genes detected within the four selected bins were classified (Figure 3-6) to provide further taxonomic resolution of the selected microbial populations.

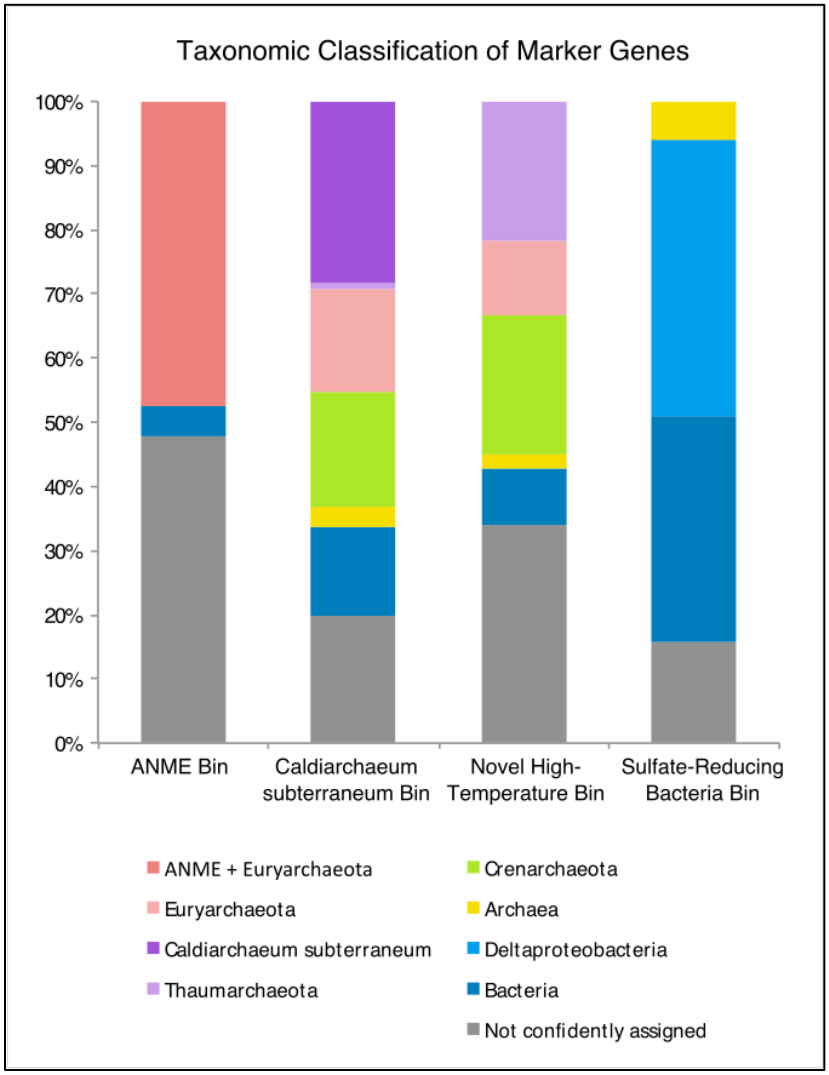

Figure **3-6**: Relative abundance of taxonomic parsing of BLASTP output of identified CheckM marker genes from each of the four metagenomic bins. The taxa are color coded, where Bacteria and phyla within the Bacteria are represented by blue, Euryarchaeota and its subphyla in pink, Thaumarchaeota and its subphyla in purple, Crenarchaeota and its subphyla in green, and unconfidently assigned in gray.

ANME Bin

Based on the taxonomic classification from marker genes in Figure 3-6, almost 30% of ANME Bin is comprised of ANME classified genes, while another 20% is comprised of other archaeal taxa. It is clear based on most annotatable marker genes that ANME represents an archaeal, and most likely an anaerobic methane-oxidizing, population. Among all contigs assigned to the ANME Bin are many families of genes involved in methane metabolism and nitrogen fixation, including those MCR and nitrogenase genes with very high contig coverage identified in the previous section. Other annotated genes involved in methane metabolism found in ANME Bin include formate dehydrogenase, tetrahydromethanopterin, formyltransferase, coenzyme F_{420}, CO dehydrogenase/acetyl-CoA synthetase. The ANME Bin contigs were also compared against annotated contigs of an ANME-1 (accession number FP565147) metagenome via a BLASTN search. Out of 196 contigs in ANME Bin, 98 unique contigs produced hits to 249 annotated genes from the reference ANME-1 metagenome, where many short sequence segments along a contig corresponded to multiple annotated genes. Thus, half of ANME Bin was identifiable with the reference ANME-1 metagenome. Many of the contigs hits (>70% sequence identity) were identified with genes in the ANME-1 reference metagenome involved in methane metabolism (*e.g.* CoB-CoM heterodisulfide reductase subunits, coenzyme F_{420} reducing hydrogenase, methyl-coenzyme M reductase). The reference ANME-1 metagenome contains 3,537 total annotated genes, which may not seem convincing of a comprehensive ANME-1 representation. However, no ANME-1 representatives have been isolated in pure culture and very few full ANME-1 metagenomes are available. The ANME-1 reference metagenome used in this study comes from a low-temperature Black Sea methanotrophic mat, and it is possible that the uncultured thermophilic archaeon found here is distinct enough to limit confident identity matches. . against a reference Methanosarcina genome only yielded one contig hit.

Within the annotated genes in the ANME Bin are several contigs corresponding to CRISPR associated genes. The clustered regularly interspaced short palindromic repeats (CRISPRs) found within host genomes have been identified as an adaptive microbial immune system in which an organism incorporates sequences derived from foreign viruses into a small-RNA-based repertoire interspaced among the repeating sequences (Karginov and Hannon, 2010). The CRISPR enzymatic complex recognizes those foreign sequences and signals mechanisms to destroy the invader by cutting the target DNA (Karginov and Hannon, 2010). The CRISPR system was investigated in hydrothermal vent samples, which showed a high number of CRISPR sequences in thermophile genomes than in mesophiles, suggesting that the marine vent virome has a more significant role in high temperature communities (Anderson *et al.*, 2011; Cowan *et al.*, 2015). Anderson *et al.* hypothesized that because there was not a higher diversity among viruses infecting thermophilic hosts and there was not a higher rate of infection in hotter environments, it is possible that CRISPRs are the predominant immunity system in thermophiles, whereas mesophiles favor other types of immunity mechanisms. Alternatively, higher rates of horizontal gene transfer at higher temperatures could explain the higher abundance of CRISPR in thermophiles (Anderson *et al.*, 2011). Although CRISPR has been found in 40% of bacterial and most archaeal sequenced genomes (Karginov and Hannon, 2010), the presence of CRISPR in the predominant archaeal population in the most extreme sample of the C0014B biosphere leads to considering the processes in the subsurface involving phage as well as transfer of genetic material. In the BLAST search of the ANME Bin against the annotated reference ANME-1 metagenome, the CRISPR associated genes were not identified to those of ANME-1, which indicates a localized and unique adaptation of the specific population of Bin 1 to its particular environment and viral interactions.

Though anaerobic methanotrophs have evaded cultivation efforts, there are many parallels in terms of function and taxonomic comparison of the Site C0014B-2-10 (15.30 mbsf)

ANME Bin contig sequences to one of few ANME-1 metagenomes. The presence of genes involved in methane metabolism and nitrogen fixation, both with extremely high calculated contig coverage, suggests that the organism/population inferred from ANME Bin is likely the most abundant in the 15.30 mbsf, 55°C horizon and is most similar to an ANME-1 than other known similar methanogen species (*i.e.* Methanosarcina acetivorans). As we have begun to see from the discussed taxonomic and functional analyses from Site C0014B, the gradient-dominated nature of Site C0014B sets up a series of microenvironments, providing niches for diverse communities. In such a dynamic environment, it is evident that the organisms representative of ANME Bin have adapted an anti-viral mechanism for their survival under already strenuous conditions.

Candidatus Caldiarchaeum subterraneum Bin

Several lines of evidence from the annotated genes in the second most complete bin point to a Candidatus Caldiarchaeum subterraneum population. Firstly, nearly half of the marker genes associated with this bin are confidently assigned to Candidatus Caldiarchaeum subterraneum, which is the largest relative proportion of classified genes among the four bins with such resolved taxonomic specificity. Secondly, genes specific to Candidatus Caldiarchaeum subterraneum are also found in this bin. The genome of a cultivated Candidatus Caldiarchaeum subterraneum has been previously annotated and points to the existence of a new phylum of Archaea (Nunoura *et al.*, 2011; Koonin and Yutin, 2014). Distinct from other members within the Thaumarchaeota, Candidatus Caldiarchaeum subterraneum harbors an ubiquitin-like protein modifier system consisting of Ub, E1 and E2 and small Zn RING finger family proteins with structural motifs specific to those from the eukaryotic system proteins (Nunoura *et al.*, 2011). Similarly, the bin discussed here contains a gene subunit each from the Ubiquitin-activation enzyme E1 and

Ubiquitin-conjugating enzyme E2 (Figure B-5). Other genes from the Ubiquitin Mediated Proteolysis pathway were absent from this bin, which may not immediately make for a compelling argument for the presence of a Candidatus Caldiarchaeum subterraneum. However, no other genes from the Ubiquitin Mediated Proteolysis were identified among the other three bins. Furthermore, the genes identified in this particular pathway had top BLAST hits to those from Candidatus Caldiarchaeum subterraneum. Thirdly, the Candidatus Caldiarchaeum subterraneum Bin also contains the reverse gyrase gene congruous with that of a Candidatus Caldiarchaeum subterraneum (Contig 3000560).

The absence of many genes in other pathways, such as energy or carbon metabolism, compounds the continued difficulty in constraining the function-identity relationship of the population inferred from this bin. Though the Probable Candidatus Caldiarchaeum subterraneum Bin is seemingly incomplete, we have encountered evidence for a partial representation of a Candidatus Caldiarchaeum subterraneum from the presence of a reverse gyrase gene, ubiquitin-like protein modifier system genes, and classified single-copy marker genes. These data have revealed additional information about the microbial community present the 15.30 mbsf horizon at C0014B independent from previous 16S rRNA gene amplicon data. Candidatus Caldiarchaeum subterraenum from this bin has emerged as a taxon 1) not previously considered in high temperature marine sediments, as it was originally collected from a high temperature subsurface mine, and 2) not predicted from the 16S amplicon (Chapter 2) or MG-RAST (*Taxonomic Changes Inferred through Metagenomics*) classifications. Additionally, the lack of Thaumarchaeota – Candidatus Caldiarchaeum subterraneum – from the MG-RAST pipeline (Figure 3-2) and amplicon data suggests that either the representatives from this bin do not actually comprise a significant proportion of the biosphere despite their functional and taxonomic support to with high temperatures, or that the complexity of the newly classified Thaumarchaeota phylum has not been wholly reflected in general classification pipelines. Under the *in situ*

thermophilic conditions, and possibly at one point hyperthermophilic conditions, Candidatus Caldiarchaeum subterraneum, or other presumed hyperthermophile, is an apt constituent of the microbial community, inferred through both taxonomy and function, to persist in the hottest horizon of the subsurface biosphere. This study was intended to capture a snapshot in time of the environment under the measured conditions, and we speculate that this horizon (c.a. 55°C) previously experienced a more optimal temperature to sustain a population of Candidatus Caldiarchaeum subterraneum and/or like hyperthermophiles. Likewise, the current temperature and conditions have likely prompted the establishment of a new, thriving and abundant ANME-1 niche.

Novel High Temperature Bin

The Novel High Temperature Bin was found to include Contig 309, which incorporates two reverse gyrase genes, but also demonstrates taxonomic ambiguity (see *Reverse Gyrase* discussion). Predicted marker genes from this however exhibit a similar trend, where gene classifications are classified amongst three archaeal phyla, almost equally (Figure 3-6). Because the Novel High Temperature Bin is small relative to the other three examined bins, no full metabolic pathways were found (Figure B-6). Half of the genes required for Coenzyme F_{420} biosynthesis were found (Figure B-7), which suggests the potential for other methane metabolism. However, a BLAST of the sequences associated with the F_{420} biosynthesis pathway did not yield consistent methanotrophic archaeal hits. The objective of examining this bin was to supplement the analysis of Contig 309 from the previous section and further resolve the taxonomic and metabolic potential of the hyperthermophilic organism associated with the reverse gyrase genes. However, the lack of metagenomic information from the binning process as well as

taxonomic heterogeneity within Contig 309 is insufficient evidence to understand the context of this organism.

Sulfate-Reducing Bacterium Bin

The Probable Sulfate-Reducing Bacteria Bin represents the third most complete bin, and is almost entirely comprised of bacterial genes, with approximately 50% of its marker genes from Deltaproteobacteria. Generally, the Deltaproteobacteria encompass a diversity of sulfate-reducers in marine sediments. Two functionally conserved and phylogenetically informative key genes, dissimilatory sulfite reductase (*dsrA* and *dsrB*) and adenosine-5'-phosphosulfate reductase (*aprA*), have been widely used in gene surveys to selectively identify sulfate-reducing bacteria against considerably more abundant background microbial populations (Teske, 2013). We used BLAST searches and the KEGG mapping annotation pipeline to search for these two functional genes; however, only *aprA* was detected among the Sulfate-Reducing Bacteria Bin genes (Figure 3-7). Figure 3-8 shows the relationship between the *aprA* gene sequence from this bin to other *aprA* genes published in protein databases. The *aprA* gene sequence from this bin clusters with a Syntrophobacter (Deltaproteobacterium) taxon from methane-rich estuary sediments as well as a Deltaproteobacterium from a 60°C Hot Seep (Guaymas Basin) enrichment. While Syntrophobacter are most common in cold, shallow marine sediments, the Deltaproteobacterium from Hot Seep-1 has been found to form ANME-1-associated clusters in thermophilic AOM Guaymas Basin sediments (Holler *et al.*, 2011). Holler *et al.*, 2011 examined hydrothermal Guaymas Basin sediments in a 50-60°C enrichment and found that the most abundant 16S rDNA bacterial sequences were from the "HotSeep-1" Deltaproteobacteria clade, which were distantly related to *Desulferlla* spp. Additionally, fluorescence *in situ* hybridization showed ANME-1 cell aggregates associated with HotSeep-1 Deltaproteobacteria. 16S rRNA gene sequences were also

previously obtained from surficial Guaymas hydrothermal sediments rich in ANME-1, and without detectable ANME-2 (Teske *et al.*, 2002). Based on the presence of a distinct thermophilic ANME-1 observed in this same C0014B-2-10 sediment horizon and the findings from both the Holler *et al.* 2011 and Teske *et al.*, 2002 studies, it is possible that the Sulfate-Reducing Bacteria Bin may represent a syntrophic thermophilic HotSeep-1 Deltaproteobacterium. Continued interest and research efforts in biological anaerobic methane oxidation in marine sediments have demonstrated many phylogenetically distinct microbial members involved in AOM. In hydrothermal vent environments, specifically, a distinct thermophilic ANME-1 and sulfate-reducing bacterium have evolved to withstand the extreme conditions. Here, a likely high temperature adapted sulfate-reducing bacterium in the same horizon as a thermophilic ANME-1 is not only further evidence for a thermophilic AOM niche in the 15.30 mbsf horizon, but demonstrates also that this thermophilic AOM community may be biogeographically similar among other sedimented hydrothermal vent ecosystems.

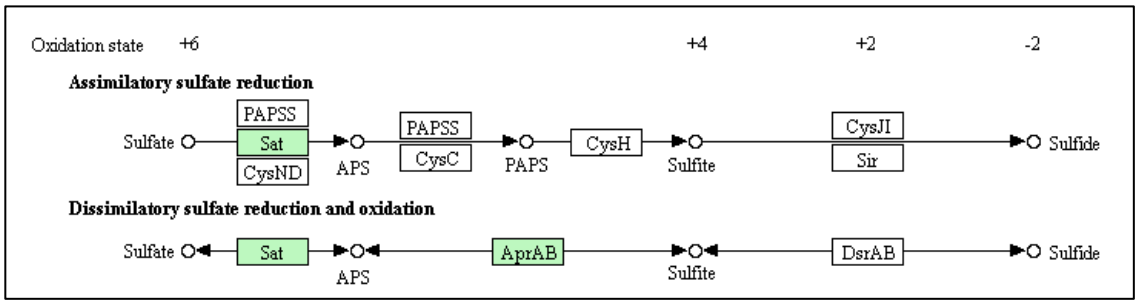

Figure **3-7**: KEGG pathway map showing the dissimilatory sulfate reduction and oxidation pathway and genes necessary for functional operation. The genes highlighted in green are present in the Sulfate-Reducing Bacterium Bin.

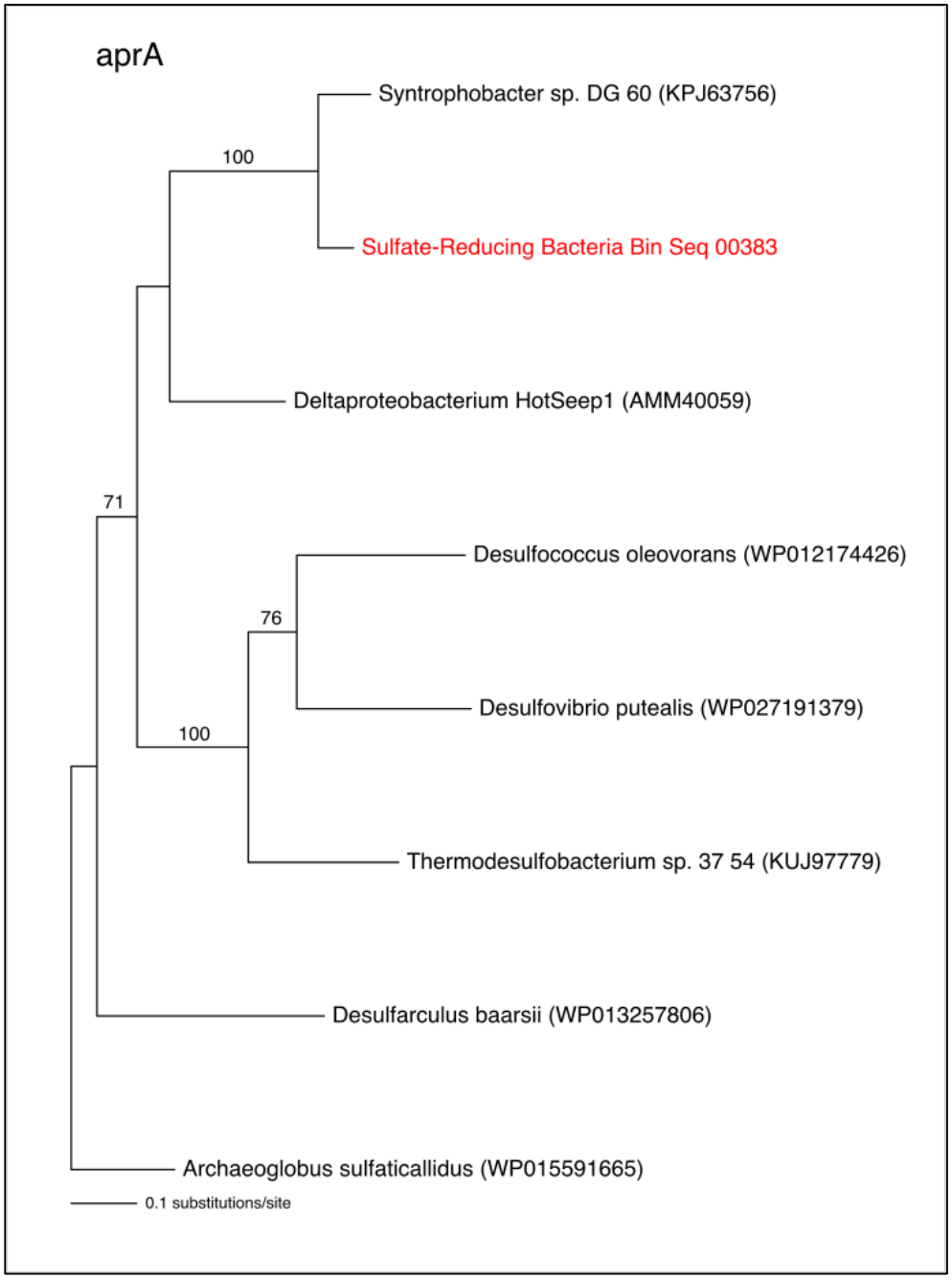

Figure **3-8**: Maximum likelihood tree (MEGA) from the alignment of *apr* genes with that from the Sulfate-Reducing Bacterium Bin. Identifiers in the parentheses represent NCBI protein accession numbers. Bootstrap values are presented on the branches.

Conclusions

Sediments within hydrothermal vent systems represent an ideal proxy for investigating the temperature-dependent biotic fringe of microbial life within the marine subsurface. Site C0014 from the IODP Expedition 331: Iheya North hydrothermal field, Okinawa backarc basin exhibits a temperature and geochemical gradient within the sediment profile that transitions from psychro- and mesophilic, continental margin-type sediments to hydrothermally altered clay. The hydrothermal gradient in Site C0014 sediments is gradual enough for stratigraphic establishments of microbial niches, but significant enough to reach the hyperthermophilic window that would otherwise be difficult for the sample collection of sediments in a geothermal gradient. In this study, we analyzed several metagenomes through this hydrothermal gradient, and found evidence for functionally and taxonomically distinct microbial communities between the cold, shallow and hot, deep temperature regimes. Our results indicate that these molecular signals are from a largely indigenous microbial community. For example, no evidence for protochlorophyllide reductase genes, which would imply a significant contribution of sequences from photic zone or terrestrial life, was found in any of the metagenomes. Furthermore, other genes, such as reductive dehalogenase and reverse gyrase, appeared to be restricted to either marine mud or hydrothermal clay horizons, respectively, which reflects a response in the functional capabilities of the microbial community to the altered conditions.

In a broad taxonomic overview of the metagenomic data, the relative abundance of archaeal sequences increases with depth, suggesting a more abundant and better-adapted archaeal community with the increasing hydrothermal conditions. A more in-depth gene survey demonstrated temperature-dependent functional adaptations between the shallowest and deepest sample horizons. The tRNA-dihydrouridine synthase gene, associated with conformational flexibility of tRNA at colder temperatures, was most abundant in the shallowest 1.32 mbsf

metagenome sample (C0014B-1-1). With a measured temperature of 8°C, the 1.32 mbsf sample is considered to be within the psychrophilic temperature regime and has the most potential for a cold-adapted community, which appears to dwindle as temperature increases. In contrast, two hyperthermophilic-specific reverse gyrase genes were detected in only the deepest metagenomic sample (55°C). The 1) alignment of both sequences to other reverse gyrase genes, 2) taxonomic classification of genes along each of the contigs, and 3) the analysis of associated bins reveal that one population is likely representative of a Candidatus Caldiarchaeum subterraneum, while heterogeneities and poor bin quality within the other were insufficient to resolve its identity between a hyperthermophilic Thaumarchaeota or Euryarchaeota. These two temperature-dependent genes, however, confirm a stratigraphic change in taxonomy and function in the microbial community composition between the two end-unit temperature regimes.

Metabolically, the detection of extremely high contig coverage of MCR in the deepest sample is evidence of an abundant methane-utilizing archaeal community. While the MCR genes in all other metagenome samples corresponded to cold-seep methanotrophic archaea or other methanogen species, the *mcrA* gene recovered from C0014B-2-10 is most similar to a thermophilic ANME-1 studied from Guaymas Basin hydrothermal sediments. Furthermore, a potentially thermophilic Deltaproteobacterium identified from a metagenomic bin, which has previously been observed in a syntrophic partnership with thermophilic ANME-1 (Holler *et al.*, 2011), is further support of an active and abundant methane-oxidizing community at 15.30 mbsf. It is clear that C0014B-2-10 hosts both ANME-1 (thermophile) and Candidatus Caldiarchaeum subterraneum (hyperthermophile) communities, thus, an overlap in optimal temperature requirements. Though this study represents a snapshot in time in this particularly dynamic environment, it is possible that the taxonomic and functional observations, here, reflect a recent change in ambient temperature at 15.30 mbsf, where a Candidatus Caldiarchaeum subterraneum

population may have once been the dominant community in a much hotter temperature regime, while ANME-1 has since found its niche among a cooler, thermophilic-type temperature.

The results of this study are important for understanding the temperature limits of microbial life in the deep marine subsurface, where sampling is extremely challenging. Despite the challenges in DNA recovery, we have interpreted the results from the metagenomic datasets to reflect an indigenous community. Though biomass concentrations are predictably much lower deeper within marine sediments, this study demonstrates that, barring other potential physical obstacles, microbial communities have the potential to persist up to (hyper)thermophilic conditions within the deep subsurface.

Acknowledgements

Samples for this research were provided by the Integrated Ocean Drilling Program (IODP). We are grateful to the IODP Expedition 331 scientific party and expedition staff onboard the *Chikyu* that assisted with drilling, sampling, measurements, and sample storage during IODP Expedition 331. We thank the National Science Foundation for support for this work through the Ocean Leadership grant BA-40 TO T331B40 and through the Center for Dark Energy Biosphere Investigations (C-DEBI) grant #OCE-0939564. We also thank the staff at the Penn State Genomics Core Facility – University Park, PA (NSF-MRI award DBI-1229046 (Axtell *et al.*, 2012)). Many of these analyses could not have been done without the help of Thijs Ettema and Jimmy Hser Wah (Uppsala University, Department of Cell and Molecular Biology; Biological Evolution) and Rosa Leon (University of Delaware, Lewes) from Jennifer Biddle's laboratory.

References

Altschul S.F., Madden T.L., Schäffer A.A., Zhang J., Zhang Z., Miller W., Lipman D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl. Ac. Res.* **25**, 3389-3402.

Anderson R.E. Brazelton W.J., and Baross J.A. (2011) Using CRISPRs as a metagenomic tool to identify microbial hosts of a diffuse flow hydrothermal vent viral assemblage. *FEMS Microbiol. Ecol.* **77**, 120-133.

Baker B.J., Lazar C.S., Teske A. and Dick G.J. (2015) Genomic resolution of linkages in carbon, nitrogen, and sulfur cycling among widespread estuary sediment bacteria. *Microbiome* **3**, 14.

Biddle J.F., White J.R., Teske A.P. and House C.H. (2011) Metegenomics of the subsurface Brazos-Trinity Basin (IODP Site 1320): comparison with other sediment and pyrosequenced metagenomes. *ISME J.* **5**, 1038-1047.

Biddle J.F., Sylvan J.B., Brazelton W.J., Tully B.J., Edwards K.J., Moyer C.L., Heidelberg J.F. and Nelson W.C. (2012) Prospects for the study of evolution in the deep biosphere. *Front. Microbiol.* **2**, 285-292.

Boetius A. Ravenschlag K., Schubert C.J., Rickert D., Widdel F., Gieseke A., Amann R., Jørgensen B.B., Witte U., and Pfannkuche O. (2000) A marine microbial consortium apparently mediating anaerobic oxidation of methane. *Nature* **407**, 623-626.

Cavicchioli R. (2006) Cold-adapted archaea. *Nature Rev. Microbiol.* **4**, 331-343.

Cowan D.A., Ramond J.-B., Makhalanyane T.P., and De Maayer P. (2015) Metagenomics of extreme environments. *Curr. Op. Microbiol*. **25**, 97-102.

Dalluge J.J., Hamamoto T., Horikoshi K., Morita R.Y., Stetter K.O., McCloskey J.A. (1997) Posttranscriptional modification of transfer RNA in psychrophilic bacteria. *J. Bacteriol*. **179**, 1918-1923.

Déclais A.-C., Bouthier da la Tour C., and Duguet M. (2001) Reverse Gyrases from Bacteria and Archaea. *Methods Enz.* **334**, 146-162.

Dekas A.E., Poretsky R.S., and Orphan V.J. (2009) Deep-Sea Archaea Fix and Share Nitrogen in Methane-Consuming Microbial Consortia. *Science* **326**, 422-426.

Forterre P. (2002) A hot story from comparative genomics: reverse gyrase is the only hyperthermophilic-specific protein. *Trends Gen.* **18**, 236-238.

Futagami T., Morono Y., Terada T., Kaksonen A.H., and Inagaki F. (2009) Dehalogenation Activities and Distribution of Reductive Dehalogenase Homologous Genes in Marine Subsurface Sediments. *Appl. Environ. Microbiol.* **75**, 6905-6909.

Handelsman J. (2004) Metagenomics: Application of Genomics to Uncultured Microorganisms. *Microbiol. Mol. Biol. Rev.* **68**, 669-685.

Hyatt, D., Chen, G.L., Locascio, P.F., Land, M.L., Larimer, F.W., and Hauser, L.J. (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119.

Holler T. Widdel F. Knittel K., Amann R., Kellermann M.Y., Hinrichs K.-U., Teske A., Boetius A., and Wegener G. (2011) Thermophilic anaerobic oxidation of methane by marine microbial consortia. *ISME J.* **5**, 1946-1956.

Huson D.H., Auch A.F., Qi J., and Schuster S.C. (2007) MEGAN analysis of metagenomic data. *Genome Res.* **17**, 377-386.

Inagaki F., Suzuki M., Takai K., Oida H., Sakamoto T., Aoki K., Nealson K.H., and Horikoshi K. (2003) Microbial Communities Associated with Geological Horizons in Coastal Subseafloor Sediments from the Sea of Okhotsk. *Appl. Env. Microbiol.* **69**, 7224-7235.

Inagaki, F., Nunoura, T., Nakagawa, S., Teske, A., Lever M., Lauer, A., Suzuki, M., Takai, K., Delwiche, M., Colwell, F.S., Nealson, K.H., Horikoshi, K., D'Hondt, S. and Jørgensen, B.B. (2006) Biogeographical distribution and diversity of microbes in methane hydrate-bearing deep marine sediments on the Pacific Ocean Margin. *Proc. Natl. Acad. Sci. USA* **103**, 2815-2820.

Kampmann M. and Stock D. (2004) Reverse gyrase has heat-protective DNA chaperone activity independent of supercoiling. *Nucl. Ac. Res.* **32**, 3537-3545.

Kanehisa M. and Goto S. (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucl. Ac. Res.* **28**, 27-30.

Kanehisa M., Sato Y., Kawashima M., Furumichi M., and Tanabe M. (2016) KEGG as a reference resource for gene and protein annotation. *Nuc. Ac. Res.* **44**, D457-D462.

Karginov F.V. and Hannon G.J. (2010) The CRISPR System: Small RNA-Guided Defense in Bacteria and Archaea. *Molec. Cell* **37**, 7-19.

Koonin E.V. and Yutin N. (2014) The Dispersed Archaeal Eukaryome and the Complex Archaeal Ancestor of Eukaryotes. *Cold Spring Harbor Persp. Biol.* **8**, 5.

Li., H. and Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**,1754-60.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009) The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics* **25**, 2078-9.

Lipp J. S., Morono Y., Inagaki F. and Hinrichs K.-U. (2008) Significant contribution of Archaea to extant biomass in marine subsurface sediments. *Nature* **454**, 991–994.

Martino A.J. (2014) Assessment and advancement of approaches for the study of subseafloor microbial communities. PhD Thesis, The Pennsylvania State University, University Park, PA.

McKay L.J., MacGregor B.J., Biddle J.F., Albert D.B., Mendlovitz H.P., Hoer D.R., Lipp J.S., Lloyd K.G., and Teske A.P. (2012) Spatial heterogeneity and underlying geochemistry of

phylogenetically diverse orange and white *Beggiatoa* mats in Guaymas Basin hydrothermal sediments. *Deep-Sea Res.* **67**, 21-31.

Meyer S., Wegener G., Lloyd K.G., Teske A., Boetius A., and Ramette A. (2013) Microbial habitat connectivity across spatial scales and hydrothermal temperature gradients at Guaymas Basin. *Front. Microbiol.* **4**, 207.

Meyerdierks A., Kube M., Lombardot T., Knittel K., Bauer M., Glöckner F.O. Reinhardt R., Amann R. (2005) Insights into the genomes of archaea mediating the anaerobic oxidation of methane. *Environ. Microbiol.* **7**, 1937-1951.

Mills H.J., Reese K.R., Shepard A.K. Riedinger N., Dowd S.E., Morono Y., and Inagaki F. (2012) Characterization of metabolically active bacterial populations in subseafloor Nankai Trough sediments above, within, and below the sulfate-methane transition zone. *Front. Microbiol.* **3**, 113.

Noon K.R., Guymon R., Crain P.F., McCloskey J.A., Thomm M., Lim J., and Cavicchioli R. (2003) Influence of Temperature on tRNA Modification in Archaea: *Methanococcoides burtonii* (Optimum Growh Temperature [T_{opt}], 23°C) and *Stetteria hydrogenophilia* (T_{opt}, 95°C). *J. Bacteriol.* **185**, 5483-5490.

Nunoura T., Takaki Y., Kakuta J., Nishi S., Sugahara J., Kazama H., Chee G.-J., Hattori M., Kanai A., Atomi H., Takai K., and Takami H., (2011) Insights into the evolution of *Archaea* and eukaryotic protein modifier systems revealed by the genome of a novel archaeal group. *Nucl. Ac. Res.* **39**, 3204-3223.

Orphan V.J., House C.H., Hinrichs K.-U., McKeegan K.D., and DeLong E.F. (2002) Multiple archaeal groups mediate methane oxidation in anoxic cold seep sediments. *Proc. Natl. Ac. Sci. USA* **99**, 7663-7668.

Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., and Tyson, G.W. (2015) CheckM: assessing the quality of microbial genomces recovered from isolates, single cells, and metagenomes. *Genome Research* **25**, 1043-55.

Parkes J.R., Cragg B.A. and Wellsbury P. (2000) Recent studies on bacterial populations and processes in subseafoor sediments: A review. *Hydrogeol. J.* **8**, 11-28.

Partensky F., Hess W.R., Vaulot D. (1999) *Prochlorococcus*: A Marine Photosynthetic Prokaryote of Global Significance. *Microbiol. and Molec. Biol. Rev.* **63**, 106-127.

Pernthaler A., Dekas A.E., Brown T., Goffredi S.K., Embaye T., Orphan V.J. (2008) Diverse synthrophic parternships from deep-sea methane vents revealed by direct cell capture and metagenomics. *Proc. Natl. Acad. Sci. USA* **105**, 7052-7057.

Saunders N.F.W., Thomas T., Curmi P.M.G., Mattick J.S., Kuczek E., Slade R., Davis J., Franzmann P.D., Boone D., Rusterholtz K., Feldman R., Gates C., Bench S., Sowers K., Kadner K., Aerts A., Dehal P., Detter C., Glavina T., Lucas S., Richardson P., Larimer F., Hauser L., Land M., and Cavicchioli R. (2003) Mechanisms of Thermal Adaptaion Revealed from the Genomes of Antarctic *Archaea Methanogenium frigidum* and *Methanococcoides burtonii*. *Genome Res.* **7**, 1580-1588.

Seeman T. (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068-2089.

Seshadri R., Adrian L., Fouts D.E., Eisen J.A., Phillippy A.M., Methe B.A., Ward N.L., Nelson W.C., Deboy R.T., Khouri H.M., Kolonay J.F., Dodson R.J., Daugherty S.C., Brinkac L.M., Sullivan S.A., Madupu R., Nelson K.E., Kang K.H., Impraim M., Tran K., Robinson J.M., Forberger H.A., Fraser C.M. Zinder S.H., and Heidelberg J.F. (2005) Genome sequence of the PCE-dechlorinating bacterium *Dehalococcoides ethenogenes*. *Science USA* **307**, 105-108.

Sørensen K.B., Lauer A. and Teske A. (2004) Archaeal phylotypes in a metal-rich and low-activity deep subsurface sediment of the Peru Basin, ODP Leg 201, Site 1231. *Geobiol.* **2**, 151-161.

Strous, M., Kraft, B., Bisdorf, R., Tegetmeyer, H.E. (2012) The binning of metagenomic contigs for microbial physiology of mixed cultures. *Frontiers in Microbiology* **3**, 410.

Swofford D.L. (2002) PAUP*. Phylogenetic Analysis Using Parsimony ("and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Teske A., Hinrichs K.-U., Edgcomb V., de Vera Gomez A., Kysela D., Sylva S.P., Sogin M.L., and Jannasch H.W. (2002) Microbial Diversity of Hydrothermal Sediments in Guaymas Basin: Evidence for Anaerobic Methanotrophic Communities. *Appl. Environ. Microbiol.* **68**, 1994-2007.

Teske A. and Sørensen K. (2008) Uncultured archaea in deep marine subsurface sediments: have we caught them all? *ISME J. **2**, 3-18.*

Teske, A. (2013) Marine Deep Sediment Microbial Communities. In E. Rosenberg *et al.* (eds), *The Prokaryotes – Prokaryotic Communities and Ecophysiology* (pp. 123-138) Springer-Verlag, Berlin, Heidelberg.

Von Mering C., Hugenholtz P., Raes J., Tringe G., Doerks T., Jensen J., Ward N., and Bork P. (2007) Quantitative Phylogenetic Assessment of Microbial Communities in Diverse Environments. *Science USA* **315**, 1126-1130.

Webster G., Yarram L., Freese E., Köster J., Sass H., Parkes R.J., and Weightman A.J. (2007) Distribution of candidate division JS1 and other Bacteria in tidal sediments of the German Wadden Sea using targeted 16S rRNA gene PCR-DGGE. *FEMS Microbiol. Ecol.* **62**, 78-89.

Yanagawa K., Nunoura T., McAllister S.M., Hirai M., Breuker A., Brandt L., House C.H., Moyer C.L., Birrien J.-L., Aoike K., Sunamura M., Urabe T., Mottl M.J., and Takai K. (2013) The first microbiological contamination assessment by deep-sea drilling and coring by the D/V *Chikyu* at the Iheya North hydrothermal field in the Mid-Okinawa Trough (IODP Expedition 331). *Front. Microbiol.* **4**, 327.

Chapter 4

Investigating the Active Microbial Populations in Near Hydrothermal Vent Sediments

Abstract

The Okinawa Backarc Basin sediments encompass spatial diversity of physical and geochemical inputs from the subsurface hydrothermal network of the Iheya North Hydrothermal Field. Previous microbial community gene-based assessments reflect stratified and diverse communities of Bacteria and Archaea between cored IODP Expedition 331 sites. Furthermore, temperature constrains in sediments affected by hydrothermal fluid flow (Site C0014) restrict the biosphere to the upper meters of sediments. In this study, we attempt to supplement the gene-based and metagenomic studies from Sites C0014 (heated sediments) and C0017 (recharge zone sediments) with an RNA-based approach to selectively investigate the active microbial populations. The results from sequenced reverse transcribed 16S rRNA demonstrate the sensitivity of extraction and amplification to external nucleic acids and attest to the challenges in working with low biomass environmental samples, but may also reveal certain bacterial taxa (*e.g.* Bacteroidetes and Alphaproteobacteria) recovered from select sediment horizons. However, inconsistencies between primersets and replicates make conclusions about the results of this RNA study extremely tenuous.

Introduction

The exploration of the deep marine biosphere has relied heavily on molecular biological approaches to understand the uncultivated, phylogenetically and functionally diverse microbial biosphere. Over the past several decades, a plethora of studies have attempted to constrain

taxonomic and functional composition of microbial life with energy, nutrient, lithological, and geochemical parameters. As the interest for deep-sea drilling for microbiological objectives has piqued, so have the number of samples and sequenced datasets that contribute to our overall knowledge of life in the deep ocean. Not only have such studies communicated the significance of the microbial biosphere to global biogeochemical cycle, they have also prompted attention to examine life at the biotic fringe or the extent under extreme conditions to which life can exist.

IODP Expedition 331 to the Okinawa Backarc Basin provided the first opportunity to drill into an active hydrothermal system and associated deposits within a sediment-filled backarc basin in a continental margin setting (Takai *et al*, 2011). Sediments in hydrothermal vent systems represent a proxy in which to study the temperature limits, thus extreme depths, of life in the sedimentary subsurface. Sediment cores recovered from IODP 331 included Site C0014, a sediment profile 450 m away from the hydrothermal vent exhibiting characteristics of hydrothermal alteration within the top 10 meters below seafloor (mbsf), and Site C0017, a site of subsurface recharging flow of oxygenated, low-temperature seawater into the permeable zone 1.5 km away from the hydrothermal vent (Takai *et al.*, 2011). Prior microbial community surveys have applied gene-based DNA (Site C0017, Yanagawa *et al.*, 2014; Site C0014, Chapter 2) and metagenomic (Site C0014, Chapter 3) tools to investigate the stratigraphic taxonomic and functional characteristics through the sediment profiles. While all studies have found distinct, diverse and stratified community structure rich in bacterial and archaeal taxa between the two sites, Site C0014 studies revealed that the biosphere is restricted to much shallower horizons, as the intensified temperature gradient is presumed to represent a significant physical barrier to survival (Chapter 2). The DNA-based analyses have provided valuable information regarding the microbial community structure in two physically and chemically distinct environments around a hydrothermal vent. However, DNA-based gene surveys do not distinguish between living populations and dead cells or preserved DNA (Sørenson and Teske, 2006). RNA assays, on the

other hand, can be used to characterize active members of a community because RNA is more reactive and readily degraded by enzymatic attack than DNA once a cell has been compromised.

Detection of RNAs, which have short residence times once a cell has been compromised, provides a strong indication of specific gene expression at the time of sampling (Nogales *et al*., 2002). By analyzing ribosomal RNA (rRNA) rather than genomic DNA in sediments, it is possible to target the active portion of the microbial community (Sørenson and Teske, 2006). 16S rRNA, like the 16S rRNA gene from DNA, has both conserved and variable regions that permit the discrimination of taxa at multiple taxonomic levels (Hill *et al*. 2000). Because ribosomes are the sites of protein synthesis within a cell, the cellular ribosome content is directly correlated with metabolic activity and growth rate (Wagner, 1994; Sørenson and Teske, 2006). Additionally, rRNA is present in higher numbers than rDNA because cells can contain many 100ibosomes (Moran *et al*., 1993). In this study, we attempt to not only distinguish only the active proportion of the microbial biosphere at Sites C0014 and C0017 using extractable RNA and subsequent 16S rRNA targeting, but spatially compare the community structure between the two geochemically and geographically distinct sites. Here, we demonstrate 1) the challenges in working with environmental RNA, 2) the necessity for quality controls to interpret the phylotypes representing the indigenous population, and 3) the potential positive taxonomic identities across Sites C0014 and C0017 sediments.

Experimental Procedures

RNA Extraction

The following protocol was modified from the RNA extraction protocol used in the Teske and MacGregor labs, UNC. An extraction blank was enforced for each extraction.

Day 1

Between 8-15 g sediment was mixed with lysis buffer (33.5 mL 4.5 M Sodium TCA, 2.5 mL 1M Tris-Cl pH 7, 1.5 mL 0.5 M Na-EDTA pH 7.5-8, 5 mL 10% N-laurylsarcosine, fill to 50 mL with DEPC or molecular grade water). Sterile 2 mL bead-beating quality tubes were filled with 0.02 g PVPP, 0.5 g 0.1 mm glass beads, 0.1 g 0.5 mm glass beads, and 1-1.5 mL of the sediment and lysis buffer slurry. Half of the sample tubes were homogenized for 40 s, while the other half 2x40 s and kept over ice. All slurry was transferred to larger centrifuge tubes and centrifuged for 25 min at 2500 rpm at 4°C. The supernatant was transferred and one volume of isopropanol was added. After vortexing, the sample was stored over night at -20°C.

Day 2

The sample was centrifuged for 30 min at 2500 rpm at 4°C. The supernatant was discarded and the pellet was washed in 10 mL col 70% ethanol. The sample was centrifuged for 10 min at 2500 rpm at 4°C and the supernatant removed. The pellet was air dried and resuspended in 10 mL DEPC or molecular grade water. One volume of phenol (pH 4.5-5.5) was added and vortexed. After 10 min of centrifugation at 2500 rpm at 4°C, the aqueous phase was transferred to a new tube. This addition of phenol to the transferred aqueous phase was repeated about 6 times until the film between phases was diminished. An equal volume of 1:1 phenol:chloroform was added, the sample was vortexed and centrifuged, and the aqueous phase transferred. An equal volume of 24:1 chloroform: isoamyl alcohol was added, the sample was vortexed and centrifuged, and the aqueous phase transferred. To the sample, 0.5 volumes 7.5M ammonium acetate, 0.7 volumes isopropanol were added. $MgCl_2$ was added to a final concentration of 2mM. The sample was vortexed and incubated overnight at -20°C.

Day 3

The sample was centrifuged for 30 min at 2500 rpm at 4°C. The supernatant was discarded and the pellet was washed in 10 mL cold 70% ethanol. The sample was vortexed and centrifuged for 10 min at 2500 rpm at 4°C. The supernatant was discarded and the pellet was air-dried. The pellet was resuspended in 100 µL water. A 50 µL aliquot of the sample was stored at -80°C and the other 50 µL were Dnased, purified, and reverse transcribed immediately.

Post-extraction Treatments of RNA

Dnase Treatment and Purification

The 50 µL aliquot of RNA extract was treated using the Invitrogen™ Ambion™ Turbo DNA-free Kit by the manufacturer's instructions. The other 50 µL sample was frozen at -80 °C. The Dnased sample was purified using the Qiagen® Rneasy Mini Kit according the manufacturer's instructions except for the last step. 35 µL water was added to the spin column for a final 1 minute spin at >8000g. The eluate was pipetted back onto the spin filter to sit for 1-5 minutes. A final centrifuge for 5 minutes yielded the final purified and Dnased product.

Preparation of cDNA Library and Amplification of cDNA

A cDNA library was made using the Qiagen® QuantiTect Reverse Transcription Kit, following the manufacturer's instructions for a 2 µL sample. A blank sample was run along side all samples and extraction blanks. cDNA library samples were kept at 8°C once completed. A polymerase chain reaction was executed on 2 µL cDNA using the KAPABiosystems® KAPA2G Robust HotStart ReadyMix PCR Kit according to the manufacturer's instructions and

thermocycler conditions for 32 cycles. Primerset pairs used in amplification encompassed the V6-V9 hypervariable regions of the 16S rRNA gene – 906F (5'- AAACTYAAAKGAATTGRCGG-3') and a modified version of 1392R (5'-ACGGGCGGTGTGTRC-3') (Rhodes et al., 2012). Additionally, another pair used encompassed the V4 region of the 16S rRNA gene – 515F (5'-GTGCCAGCMGCCGCGGTAA-3') and 806R (5'-GGACTACHVGGGTWTCTAAT-3'). A PCR control (only reagents) was run along side all samples. The products were visualized with gel electrophoresis on a 1% agarose gel. Samples were chosen for sequencing based on bands in the samples with significantly fainter bands in the controls. In all cases the extraction blank and cDNA control had amplified product, but were fainter than the sample product.

DNA Sequencing

Table 4-1 shows the samples sequenced. RNA refers to extract that was Dnased, purified, and reverse transcribed before PCR. DNA refers to extract that was only amplified by PCR. A total of 11 samples had been taken through the extraction and subsequent processes. The samples in Table 4-1 are those that produced product of what was considered more concentrated than the extraction blank and cDNA control. In many cases, if no amplification worked at 32 PCR cycles, 35 cycles was attempted. Sequencing was performed at the Penn State Genomics Core Facility – University Park, PA using the Illumina® MiSeq 2500. The sequencing facility prepared DNA libraries using Nextera XT Library Preparation Kits prior to sequencing. Together, the samples (Table 4-1) were run on one sequencing plate with 84 samples total. This sequencing runs of the MiSeq averages up to 300 million single reads, or 30 million paired reads of ~250 nucleotides per run.

Analysis of 16S rRNA (Gene) Amplicons

Sample demultiplexing was performed in Mothur (v.1.30.1), as well as some preliminary quality controls eliminating sequences shorter than 100 basepairs, with more than one mismatch in the barcode sequence, with more than two mismatches in the primer sequence, or with more than eight homopolymers. In addition, sequences were screened by quality score using the "qwindowaverage" function, set at a quality of 35 (Schloss et al., 2009).

The resulting individual fasta files were then processed as a single job with the NGS analysis pipline of the SILVA rRNA gene database project (SILVAngs 1.3) according to the default parameters (Quast et al., 2013). This pipeline included alignment with the SILVA Incremental Aligner (SINA v1.2.10 for ARB SVN (revision 21008)) (Pruesse et al., 2012) against the SILVA SSU rRNA SEED and quality controlled (Quast et al., 2013). Specifically, reads that had fewer than 50 aligned nucleotides and/or more than 2% ambiguities or homopolymers were excluded from further processing, as were putative contaminations, artifacts, and reads with a low alignment quality. The remaining sequences were dereplicated and clustered into operational taxonomic units (OTUs) on a per sample basis, using cd-hit-est (version 3.1.2; http://www.bioinformatics.org/cd-hit) (Li and Godzik, 2006) running in *accurate mode*, ignoring overhangs, and applying identity criteria of 1.00 and 0.98, respectively. Classification was performed by a local nucleotide BLAST search against the non-redundant version of the SILVA SSU Ref dataset (release 123; http://www.arb-silva.de) using blastn (version 2.2.30+; http://blast.ncbi.nlm.nih.gov/Blast.cgi) with standard settings (Camacho et al., 2009). Reads were classified if the value of the function "(% sequence identity + % alignment coverage)/2" exceeded 93. Classification of each I reference read was mapped onto all reads that were assigned to the respective I. Those that did not fall within this classification quality were assigned to the group

"No Relative." Classifications of the reference sequences were then mapped onto reads within their respective OTUs.

Results and Discussion

The objective of this study was to profile 16S rRNA downcore at Sites C0014 and C0017 to understand the spatial and stratigraphic distribution of microbial community composition between two geochemically and geographically distinct sediment profiles in the Iheya North Hydrothermal Field. After extraction attempts on 11 different sample horizons and numerous post-extraction manipulations, we were only confident in the integrity of 4 sample horizons as representing extractable RNA in greater concentration than quality controls (Table 4-1). The following sections are discussions of the quality of sequencing attempts, vetting of taxonomic phylotypes, and overall significance of the presented data in the context of the Iheya North Hydrothermal Field.

Table **4-1**: Sequenced IODP sample horizons and corresponding depth in meters below seafloor. Each sample was amplified with two different primersets (See Experimental Procedures). The Raw Sequence Yield shows the number of sequences after quality control that were classifiable. Each extraction had a corresponding extraction blank. Samples C0017D-1-3 and C0014D1-1 were extraction in parallel. DNA refers to extract that was only PCR-amplified.

Sample	Primer Set	Raw Sample Yield		Control Yield
C0017A-1-1 (0.735 mbsf)	V4	RNA	50747	77282
		DNA	794	
	V6-V9	RNA	36759	37744
		DNA	166	
C0017D-1-3 (63.77 mbsf)	V4	RNA	16009 16333	285
		DNA	3940	
	V6-V9	RNA	98 7089	16608
		DNA	139	
C0014D-1-1 (3.075 mbsf)	V4	RNA	1145 16433	see C0017D-1-3
		DNA	444	
	V6-V9	RNA	11353 19361	see C0017D-1-3
		DNA	7	
C0014B-1-1 (1.22 mbsf)	V4	RNA	108067	16892
		DNA	133	
	V6-V9	RNA	20816	10527
		DNA	92	
cDNA control	V4			11560
	V6-V9			10686

General Sequencing Yield

Prior to analyzing the sequence classifications from the IODP Sites C0014 and C0017 samples, an initial evaluation of the sequenced extraction blanks and reverse transcription controls, relative to the samples, is important in assessing the quality of the dataset as a whole. While part of the library preparation process at the sequencing facility attempts to use relatively equal DNA concentrations of each sample, we still expected that extraction blanks and controls would be significantly lower than the samples. Ideally, the control samples should represent the minor component of traceable sequences from the ambient environment or reagents that we must account for when parsing taxa from the indigenous communities. Table 4-1 shows each sample's

sequencing yield as represented by extractable and reverse transcribed RNA as well as DNA, which was PCR amplified only extract. Additionally in Table 4-1 are the extraction blank sequencing yields corresponding to each sample. Table 4-1 shows that the reverse transcription control (cDNA control) produced significant sequence product, suggesting that ambient RNA has been integrated into the control reagents to become subsequently reverse transcribed and amplified by PCR primers. Alternatively, DNA present would also produce amplification in the subsequent PCR step. Moreover, in the cases of C0017A-1-1, C0017D-1-3 (V6-V9), and C0014D-1-1 (V6-V9), the sequencing yield of the controls exceeds those of their corresponding samples, indicating severe exposure to ambient RNA in the extraction process.

According to Table 4-1, all samples representing extractable DNA produced much lower sequence yields, relative to those from reverse transcribed RNA. Because the extraction protocol is selective for RNA in certain low pH steps, it is not surprising that DNA product would be diminished. The results here are reassuring in knowing that this extraction protocol produces a significantly higher yield of RNA, which should, ultimately, dampen any signal from background DNA. Nevertheless, when working with extremely difficult clays and low biomass sample, such as those from the Okinawa Backarc Basin, having both datasets is advantageous to parse taxonomic representatives found in DNA from those in RNA, representing the active community.

Contaminant Vetting: Taxonomic Classification of Sequencing Results

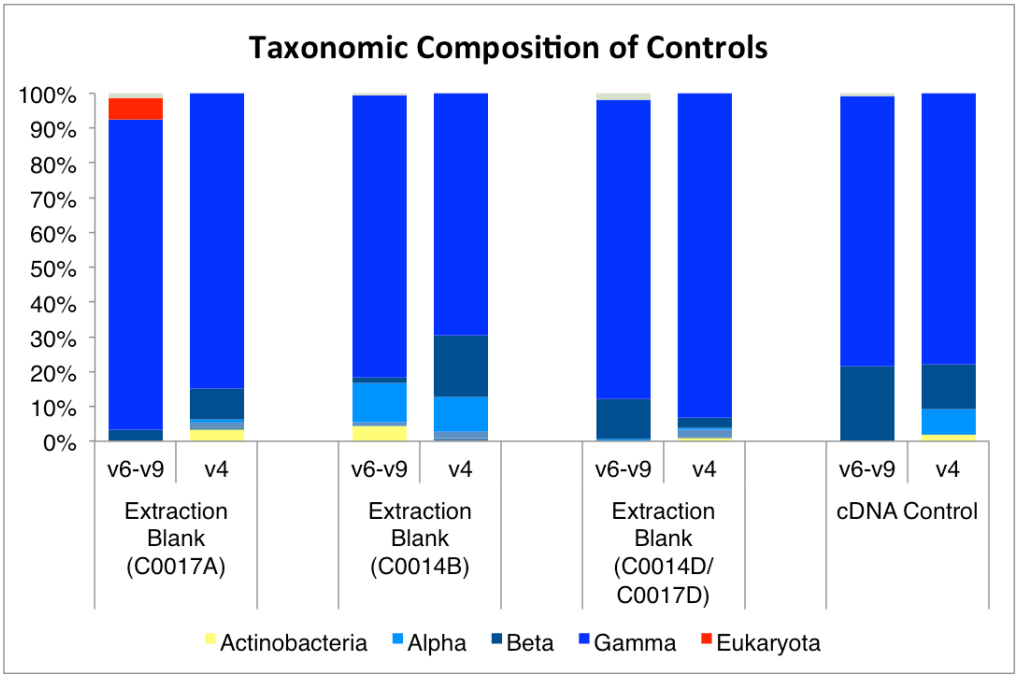

Figure **4-1**: Relative abundance chart showing the taxonomic classifications of extraction blanks and reverse transcription (cDNA) control. Each was amplified with two primersets.

Extraction blank and cDNA controls were examined in order to determine the extent of contamination and also to parse the sequence data from the environmental samples. Figure 4-1 shows the percent relative abundance of phyla (and Eukaryota) that comprise each control sample. Interestingly, Gammaproteobacteria – Escherichia Shigella, specifically – represent the majority of sequences found in both extraction blanks and cDNA controls. The consistency among all samples between Gamma- and Betaproteobacteria is an indication that the input of RNA (or DNA) is not situational, but rather consistent throughout the laboratory procedures. Escherchia *shigella* is a subclassification within the Gammaproteobacteria that is known as a human pathogen (Hale, 1991). Generally, the Gammaproteobacteria phylum encompasses a wide range of microorganisms and are often encountered readily in 16S rRNA gene libraries from deep subsurface sediments (Fry *et al.*, 2008). For example, Gammaproteobacteria belonging to Halomonas and Marinobacter have even been isolated from Peru Margin ODP sites 1228 and

1229 (Biddle *et al.*, 2005). Specific to the IODP Expedition 331 cruise, however, seawater samples were sequenced to account for intrusion of microbial input from coring and revealed sequences belonging to orders Alteromonadales, Oceanospirillales, Pseudomonadales, Xanthomonadales, and mostly Vibrionales (Yanagawa *et al.*, 2013). Table 4-2(A) shows several BLASTN top hits to several sequences classified as Escherchia *shigella* from one extraction blank sample. The results from Table 4-2 and the overwhelming abundance of Escherichia *shigella* over any other Gammaproteobacteria imply a non-seawater input, but rather a likely external contribution from reagent contamination.

Table **4-2**: Table listing the top BLAST hits of several randomly selected sequences from (A) C0017A-1-1 classified Escherichia Shigella and (B) C0014B-1-1 classified Soil Crenarchaeotic Group. Both sequence queries were from V4 primer amplified and sequenced datasets. The accession numbers are from the NCBI nucleotide database.

Top BLASTN Hits to Select Samples

A

C0017A-1-1 (V4): Escherichia Shigella	
Accession Number	Sample Type
JF244572	Intestinal microbiota
KX034560	Textile industry effluent
AY268309	Shower curtain biofilm microbes
KU365417	Multidrug and heavy metal resistant isolates from Ganges water
KU821025	Industrial effluent
EU771609	Elephant feces
KM499057	Bat digestive tract
KM126460	Argentinian Lake Naheul Huapi pumice
KM882689	Sugarcane bagasse metagenome

B

C0014B-1-1 (V4): Soil Crenarchaeotic Group	
Accession Number	Sample Type
KX077532	Freshwater estuarine wetland
EU365286	Vegetable soil
LN864946	Karst system
FJ784251	Drained alkaline, saline Lake Texcoco, Mexico
KP328069	Rice soil tidal marsh, East China
KM273624	Rice roots
JQ978473	Shule River soil
JN205398	Iron-ore mine

After taxonomic vetting of the C0014 and C0017 environmental samples from taxa identified from controls (see methods), it is clear that the extraction technique was extremely susceptible to non-indigenous taxa. Figure 4-2 shows the percent relative abundance of sequences remaining after those identified from extraction and amplification controls were removed. In all but one sample, less than 40% of sequences were recovered as potential for identifying taxa from the indigenous sample. Samples marked with stars represent extracted, DNased, and purified RNA that was subsequently reverse transcribed and amplified. In most cases, the RNA yield is greater that sequence yield from extractable DNA, which was previously discussed as an affirmative outcome of the RNA selectivity of the extraction protocol. Surprisingly, almost all RNA and DNA sequences are identified as bacterial taxa (Figure 4-2). In general, archaeal taxa are prevalent from marker gene and metagenome datasets from marine subsurface sediments, and even in RNA studies (*e.g.* Biddle *et al.*, 2006). According to the 16S amplicon and metagenomic studies discussed in the previously chapters, recovered archaeal sequences constituted a significant proportion of the total presumed indigenous community. Though the explanation for such a loss in information is not conclusive, it is possible that 1) Archaea in these horizons may no longer be an active part of the community; 2) the extraction protocol may not have been optimized for the complex sediment matrix and clay composition that the Iheya North Hydrothermal Field experiences; or 3) due to the lack of murein in archaeal cell membranes (Kandler, 1994; Kandler and Konig, 1998), archaeal cells may be more resistant to the chemical lysing mechanisms endured in this particular extraction protocol. The former is less likely a function of whether an archaeal community population suddenly became extinct, but rather a consequence of freezer malfunctions thawing the samples to room temperature, which may have eradicated the entire active microbial community.

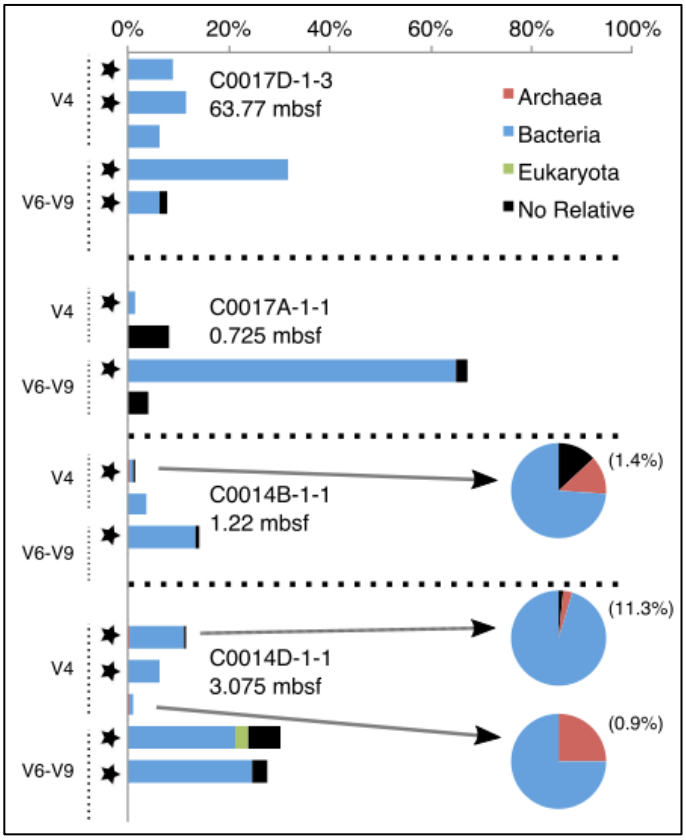

Figure **4-2**: Relative abundance of domain-classified taxa of each IODP 331 sample after contaminant taxa were removed (*i.e.* the remainder of the bar). Each sample was amplified using two different primer sets (V4 and V6-V9). The stars indicate reverse transcribed, sequenced RNA samples. Bars with no stars indicate sequenced DNA, or PCR amplified only extract. Several samples from C0014 contained archaeal sequences – the entire sample is displayed in the pie chart to visualize the relative abundance of archaeal sequences. The number in parentheses represents the percentage of sequences remaining after contaminant vetting.

Among the remaining sequences represented in Figure 4-2, a deeper classification was executed in order to resolve any trends between samples and/or sites (Figure 4-3). Within sample C0017D-1-3, the Betaproteobacteria order represents a large percentage of the RNA sequences. Those from the V4 primers were classified down to the Neisseriales order, while the RNA sample from the V6-V9 primers was classified down to the Methylophilales order. Neisseriales is often associated with the pathogens transmitting meningitis and gonorrhea, however, many of the sequences found in this study were associated with the Kong *et al*., 2012 study investigating the

microbiome shifts in skin conditions. It is, therefore, unlikely that the extractable RNA amplified from the V4 primers is from an indigenous marine environment. The RNA sample amplified by the V6-V9 primers, however, shows BLASTN hits to Betaproteobacteria samples from unpublished, environmental studies (*e.g.* NCBI accession numbers KR339083: Antarctic snow microbial community, and KM221356: geothermal spring). The duplicate RNA sample produced most sequences classified as Deltaproteobacteria of the Oligoflexales order, which has been studied in wetland, and hypersaline sediments (Kim *et al*., 2012; Bouali *et al*., 2014). While it is possible that the V6-V9 dataset may represent an indigenous population at C0017D-1-3, the inconsistency between duplicate samples suggests that the sample size is overly small to account for such diversity.

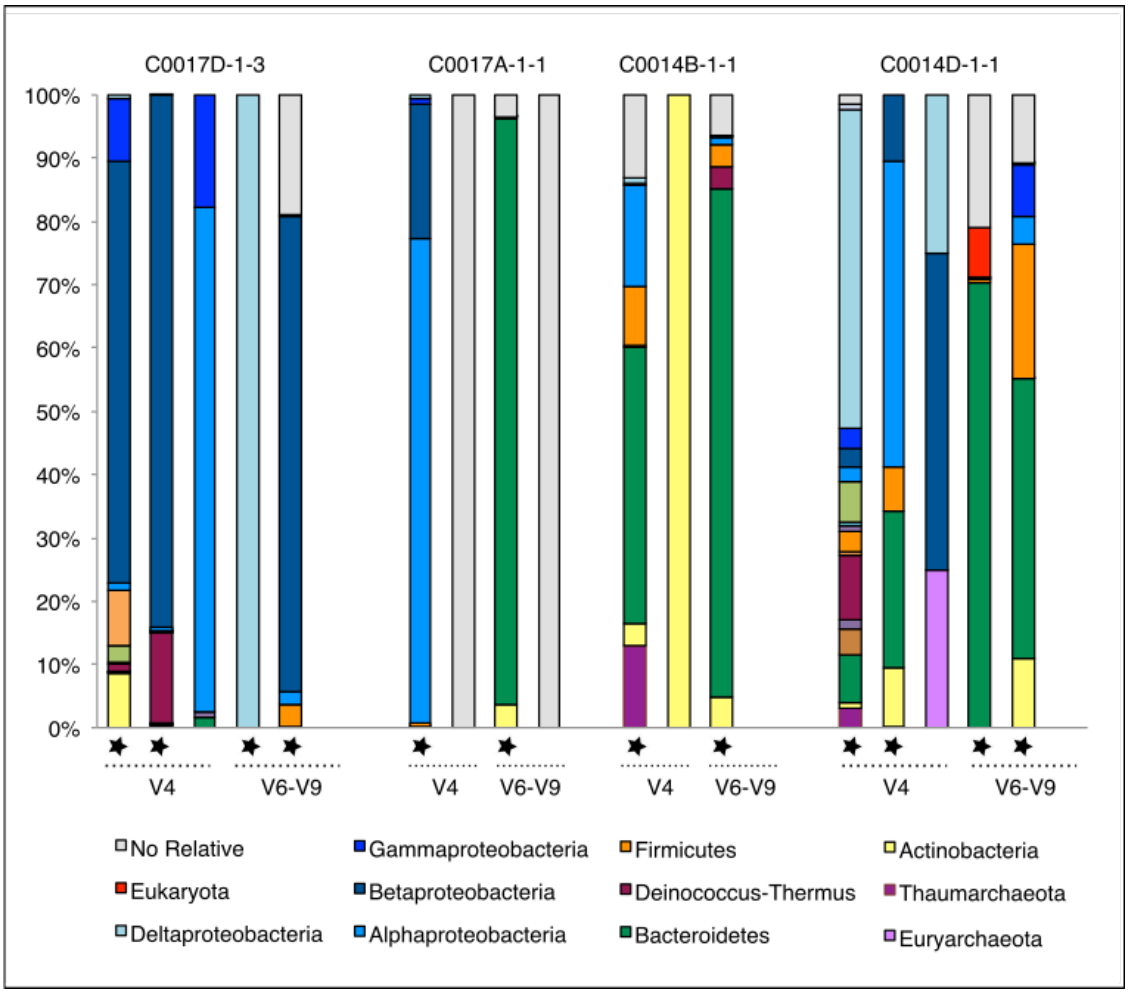

Figure **4-3**: Relative abundances of classified phyla among sequenced samples. Each sample was amplified using two different primer sets (V4 and V6-V9). The stars indicate reverse transcribed, sequenced RNA samples. Bars with no stars indicate sequenced DNA, or PCR amplified only extract. In some DNA samples, no sequences remained after vetting.

Another seemingly abundant taxon among the V4 primer amplified samples is from the Alphaproteobacteria order (Figure 4-3) – Rhodospirillales. Rhodospirillales is found across a range of deep, marine environments, ranging from sediments, basalt, hydrothermal deposits, and deep water (Orcutt *et al*., 2011) and could potentially represent part of the indigenous microbial community between Sites C0014 and C0017. Interestingly, Rhodospirillales is found in RNA samples from the surface sites, C0014B-1-1, C0014D-1-1, and C0017A-1-1, but found in the

DNA in the deeper subsurface Site C0017D-1-3 (63.77 mbsf). A previous study from Site C0017 found Alphaproteobacteria in classified 16S rRNA gene fragments, but only in the 26-30 mbsf depth horizon (Yanagawa *et al.*, 2014). While the DNA classification of Alphaproteobacteria does not agree exactly with the Yanagawa *et al.*, 2014 study, the presence of this phylotype in the RNA from surface sites implies a possible advantage of an active Rhodospirillales community to the upper sediments in the Iheya North hydrothermal field.

The V6-V9 primerset shows an apparent bias in the amplification of Bacteroidetes in the RNA samples from the surface sites of C0014 and C0017 (Figure 4-3). Species from the phylum Bacteroidetes exist in many environments from surficial sediments, basalts, whale carcasses, and ultra-mafic rock-hosted sulfide, but are rarely found in deep sediments and active hydrothermal sulfide chimneys (Orcutt *et al*., 2011). The significant Bacteroidetes surface abundances presented here support this observation and should be further considered as possible indigenous inhabitants. Almost all of the recovered Bacteroidetes sequences here are further classified as Flavobacteria, which are an uncultivated but a metabolically versatile genera, including halophilic, aerobic chemoheterotrophs, and many pathogens. A BLASTN search of several sequences classified as Flavobacteria produce a range of results that span the aforementioned heterogeneity of metabolic potential and environments, which imply a diverse Bacteroidetes community in these samples. However, among the C0014D-1-1 RNA V6-V9 replicates, one sample was classified as Flavobacteria, while the other as the Sphingobacteria class. The Sphingobacteriales order includes bacterial species found in oxic, anoxic, sulfidic, and acidic environments, and is recognized as fish pathogens (Orcutt *et al.*, 2011). Sequences associated with the Flavobacteria and Sphingobacteria sequences of the Bacteroidetes in Sites C0014 and C0017 surface sediments are unsurprising, considering their abundances in other deep, marine settings. However, as the initial classification of V6-V9 sequences at the phylum level demonstrates a degree of spatial homogeneity across sites in terms of Bacteroidetes, replicates of

the same RNA sample do not exhibit reproducibility at a deeper taxonomic level. These results are perplexing and indicate an unreliable and insufficient sampling of the population.

Although the lack of archaeal sequences is unexpected from previous amplicon and metagenomic analyses, one V4 amplified RNA sample from C0014B-1-1 (Figures 4-2 and 4-3) produced an archaeal signal with 193 sequence reads classified as Thaumarchaeota. Two other samples in C0014D-1-1 show an archaeal sequence contribution to the relative abundance of remaining possible indigenous sequences (Figure 4-3), however, the quantity of sequences recovered after contaminant vetting was extremely low (1 and 4 archaeal sequences per sample). Almost all Thaumarchaeota sequences within the C0014B-1-1 are classified further as Soil Crenarchaeotic Group. Table 4-2(B) shows the top BLASTN hits of these sequences and reveals taxa from a number of studies from non-marine sampling locations. Not only are these sequences associated with environments independent of the deep marine realm, they are also not present in any quantity in previous amplicon study (Chapter 2). If terrestrial input were significant in this basin environment, it would be reflected in the taxonomic fingerprint of DNA from the surface samples. For example, the detection of eukaryotic and fungal taxa, despite the prokaryotic 16S specific primerset, demonstrated that inputs from the overlying water column were deposited in the Site C0014 sediments and buried. In Chapter 2, we used archaeal sequences as an indication of in indigenous population because no archaeal sequences were detected in drilling fluid (Yanagawa *et al.*, 2013) or extraction blanks (Appendix A). The archaeal taxa were also comparable to ubiquitous marine taxa. But, the same assumptions cannot be made in this study – the data presented here, and as the Soil Crenarchaeotic Group name and sampling locations of studies suggest, do not appear to be of marine origin or of a community deposited from nearby terrestrial surroundings.

RNA dataset comparison to published DNA amplicons

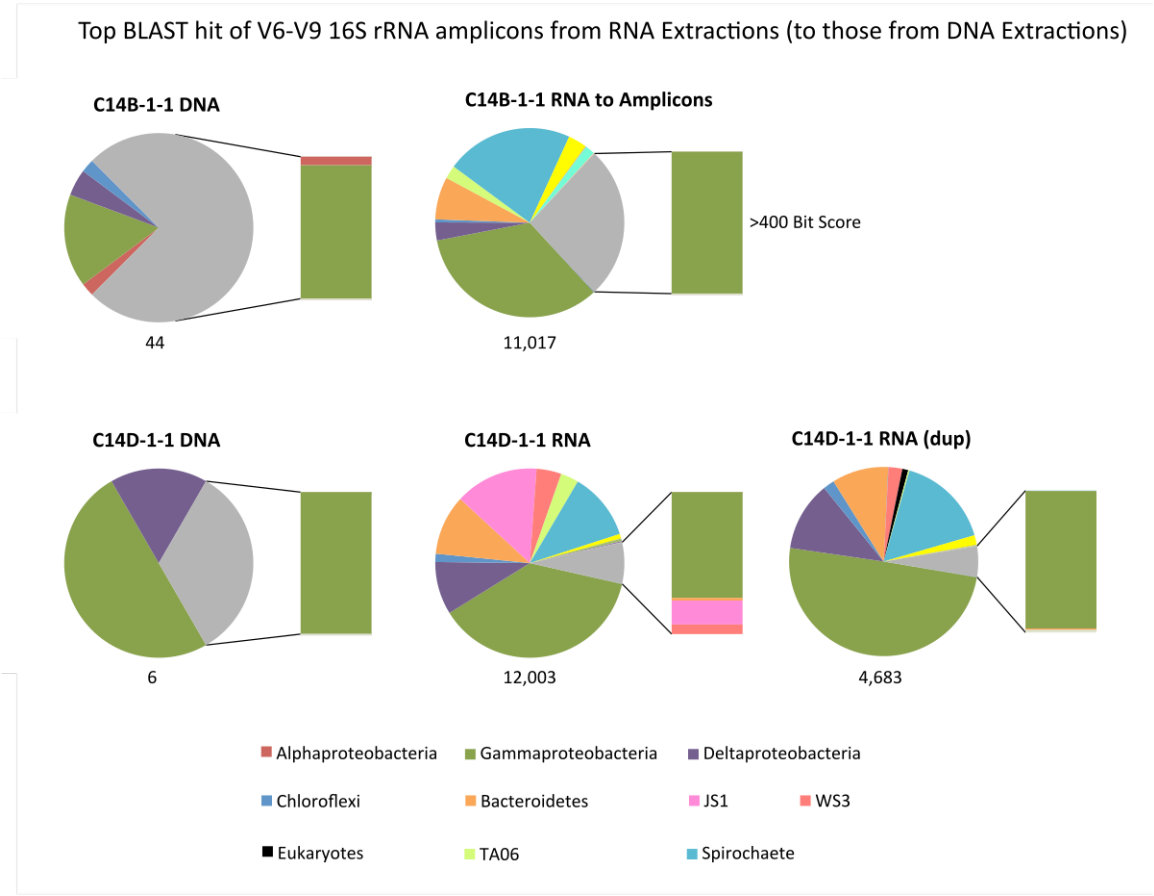

Figure **4-4**: Top BLAST hits of V6-V9 RNA and DNA samples against the corresponding Brandt and House, 2016 DNA amplicon samples The gray area represents hits with bit scores above 400 – the expanded section of this area is depicted as the vertical bar to the right side of the pie. The numbers listed below the pie charts are the number of sequences.

With most sequence processing pipelines, such as SILVAngs, sequences are compared against databases of published sequence datasets and cultured organisms. In this study, we were interested in the taxonomic makeup of sequences in how they compare to not just global databases, but to previously published DNA sequence results from the same samples (*i.e.* Brandt and House, 2016). We used BLAST to compare the V6-V9 C0014 samples in this dataset to

libraries of their corresponding Brandt and House, 2016 DNA datatsets. The top BLAST hits of each sample are presented in Figure 4-4 and show that the majority of top hits correspond to Gammaproteobacteria in most samples. As discussed with Figure 4-1, Gammaproteobacteria represent the predominant taxa found in extraction blanks and controls and should be interpreted in samples with caution. Of the BLAST hits with >400 bit score, one C0014D-1-1 RNA sample showed hits to JS1, which is a bacterial phyla observed in marine sediments. Though JS1 hits represent a small proportion of the total sequences, we were interested in exploring these sequences further as potential representatives from an active microbial community in this sample. We took the sequences corresponding to JS1 through a BLAST search against the NCBI database to determine whether they could be confidently identified as such. However, the results displayed in Table 4-3 indicate that those sequences are not likely JS1 representatives. These results were also repeatable with other taxa from Figure 4-4 (WS3, for example, shown in Table 4-3). Thus, the taxonomic classifications of sequences from the RNA dataset do not appear to support taxa found in marine sediments.

Table **4-3**: The following table shows the C0014D-1-1 RNA sequences whose hits were to two potential marine taxa when BLAST against Brandt and House, 2016 (Column 1); how those hits were initially classified according to SILVAngs (Column 2); the top hit classification when BLAST against the NCBI nucleotide database (Column 3); the number of sequences making up the hits in Column 1 (Column 4).

BLAST hit to DNA amplicons	RNA sequence classification	BLAST hit to NCBI	Number sequences
JS1	Firmicutes	Firmicutes	82
WS3	Actinobacteria	Actinobacteria	19
	Firmicutes	Firmicutes	13

Conclusions

Marine subsurface sediments host a widely diverse microbial assemblage, which has been readily detected and characterized by molecular methods and DNA sequencing. The IODP Expedition 331 to the Okinawa Backarc Basin collected sediment cores from two different environments surrounding a hydrothermal vent – a gently warmed sediment profile affected by subsurface hydrothermal fluid migration and a much cooler sediment profile representing a zone of recharged seawater into the subsurface. Both sampling sites have been previously studied in terms of microbial community structure through sequenced and taxonomically classified DNA marker genes. The study presented here was intended to provide an additional perspective of the microbial community composition by studying reverse transcribed DNA from extractable environmental RNA, which represents a more accurate view of intact, thus active, microbial life.

The classified sequencing results from both RNA and DNA are variable and, often, inconsistent. When working with novel environments and, especially, low biomass samples, it is crucial to account for background DNA (in the case of specific RNA interests) and external inputs (*i.e.* reagent or background contamination), which is why extraction blanks and amplification controls were also sequenced. Though it was valuable to obtain sequences from such sources, the sequencing yield of the controls in most cases exceeded that of the environmental samples, indicating significant background noise from extraction contaminants. Background DNA appeared to be minimal, which implies that the RNA extraction process did, indeed, select for predominantly RNA.

Among the entire dataset, the majority of sequences were associated with taxa from the controls. The remaining sequences did not exhibit taxa that were predicted from previous assessments (*i.e.* Chloroflexi, JS1, Planctomycetes, and Archaea). Even when the sequences were compared to published datasets from the same samples (Brandt and House, 2016) and appeared to

have hits to marine taxa, the sequences were not confirmed as such. A deeper classification of the abundant phyla revealed that certain taxa, especially those associated with pathogens or skin microbiome, were likely not indigenous to the environmental sediments. Other taxa within the Alphaproteobacteria and Bacteroidetes showed consistency to other marine or aquatic studies and could be representative of the active population. From a spatial standpoint, some overlap of taxa, such as Bacteroidetes and Alphaproteobacteria, occurred among the surficial sediments across both Sites C0014 and C0017. The deeper sample at Site C0017 exhibited predominantly Beta- and Deltaproteobacterial taxa, some of which provoked skepticism in validity of origin. Overall, replicate samples were not reproducible, as were the results between two different primersets. It is probable that a significant loss of diversity occurred when the freezer malfunction compromised the preservation of samples, which is supported by high detection of contaminant in classified taxa, overall low (RT)PCR amplification yield among samples, low diversity, and lack of predicted marine phylotypes present. Nonetheless, this study demonstrates the challenges and required meticulousness associated with environmental RNA work and the need for quality controls when working in such challenging environments.

Acknowledgements

Samples for this research were provided by the Integrated Ocean Drilling Program (IODP). We are grateful to the IODP Expedition 331 scientific party and expedition staff onboard the *Chikyu* that assisted with drilling, sampling, measurements, and sample storage during IODP Expedition 331. We thank the National Science Foundation for support for this work through the Ocean Leadership grant BA-40 TO T331B40 and through the Center for Dark Energy Biosphere Investigations (C-DEBI) grant #OCE-0939564. Sequencing was performed at the Penn State Genomics Core Facility – University Park, PA (NSF-MRI award DBI-1229046 (Axtell *et al.*,

2012)). Lastly, we thank Jennifer Biddle and Zena Cardman for their technical support and advice in RNA extraction protocol troubleshooting.

References

Biddle J.F., House C.H., and Brenchley J.E. (2005) Microbial stratification in deeply buried marine sediment reflects changes in sulfate/methane profiles. *Geobiol.* **3**, 287-295.

Biddle J.F., Lipp J.S., Lever M.A., Lloyd K.G., Sørensen K.B., Anderson R., Fredricks H.F., Elvert M., Kelly T.J., Schrag D.P., Sogin M.L., Brenchley J.E., Teske A., House C.H., and Hinrichs K.-U. (2006) Heterotrophic Archaea dominate sedimentary subsurface ecosystems off Peru. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 3846–3851.

Bouali M., Zrafi I., Backhrouf A., Chaussonnerie S., and Sghir A. (2014) Bacterial structure and spatiotemporal distribution in a horizontal subsurface flow constructed wetland. *Appl. Microbiol. Biotechnol.* **98**, 3191-3203.

Camacho C., Coulouris G., Avagyan V., Ma N., Papadopoulos J., Bealer K., and Madden T.L. (2009) BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421.

Fry J.C., Parkes J.P., Cragg B.A., Weightman A.J., and Webster G. (2008) Prokaryotic biodiversity and activity in the deep subseafloor biosphere. *FEMS Microbiol. Ecol.* **66**, 181-196.

Hale T.L. (1991) Genetic Basis of Virulence in *Shigella* Species. *Molec. Rev.* **55**, 206-224.

Kandler O**.** (1994) Cell wall biochemistry in Archaea and its physical implications. *J. Biol. Phys.* **20,**165-169.

Hill G.T., Mitkowski N.A., Aldrich-Wolfe L., Emele L.R., Jurkonie D.D., Ficke A., Maldonado-Ramirez S., Lynch S.T. and Nelson E.B. (2000) Methods for assessing the composition and diversity of soil microbial communities. *Appl. Soil Ecol.* **15**, 25-36.

Kandler O. and Konig H. (1998) Cell walls polymers in Archaea (Archaeabacteria). *Cell Mol. Life Sci*. **54,** 305-308.

Kim J.S., Makama M., Petito J., Park N.H., Cohan F.M., and Dungan R.S. (2012) Diversity of Bacteria and Archaea in hypersaline sediment from Death Valley National Park, California. *MicrobiologyOpen* **1**, 135-148.

Kong H.H., Oh J., Deming C., Conian S., Grice E.A., Beatson M.A., Nomicos E., Polley E.C., Komarow H.D., NISC Comparative Sequence Program, Murray P.R., Turner M.L., and Segre J.A. (2012) Temporal shifts in the skin microbiome associated with disease flares and treatment in children with atopic dermatitis. *Genome Res.* **22**, 850-859.

Li W. and Godzik A. (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658-1659.

Moran M.A., Torsvik V.L., Torsvik T. and Hodson R.E. (1993) Direct extraction and purification of rRNA for ecological studies. *Appl. Environ. Microbiol.* **59**, 915-918.

Nogales B., Timmis K.N., Nedwell D.B. and Osborn A.M. (2002) Detection and diversity of expressed denitrification genes in estuarine sediments after reverse transcription-PCR amplification from mRNA. *Appl. Environ. Microbiol.* **68**, 5017-5025.

Wagner, R (1994) The regulation of ribosomal RNA synthesis and bacterial cell growth. *Arch. Microbiol.* **161**, 100-109.

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.

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

Quast C., Pruesse E., Yilmaz, P., Gerken J., Schweer T., Yarza P., Peplies J., and Glöckner F.O. (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590-D596.

Rhodes M.E., Oren A., and House C.H. (2012) Dynamics and Persistence of Dead Sea Microbial Populations as Shown by High-Throughput Sequencing of rRNA. *Appl. Environ. Microbiol.* **78**, 2489-2492.

Schloss P.D., Westcott S.L., Ryabin T., Hall J.R., Hartmann M., Hollister E.B., Lesniewsi R.A., Oakley B.B., Parks D.H., Robinson C.J., Sahl J.W., Stres B., Thallinger G.G., Van Horn D.J., and Weber C.F. (2009) Introducing mothur: open-source, platform-independent, community-

supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75**, 7537-7541.

Sørenson K.B., and Teske A. (2006) Stratified Communities of Active Archaea in Deep Marine Subsurface Sediments. *Appl. Environ. Microbiol.* **72**, 4596-4603.

Takai K., Mottl M.J., Nielsen S.H., and the Expedition 331 Scientists (2011) Expedition 331 summary. *Proc. IODP* **331**.

Wagner R. (1994) The regulation of ribosomal RNA synthesis and bacterial cell growth. *Arch. Microbiol.* **161**, 100-109.

Yanagawa K., Nunoura T., McAllister S.M., Hirai M., Breuker A., Brandt L., House C.H., Moyer C.L., Birrien J.-L., Aoike K., Sunamura M., Urabe T., Mottl M.J., Takai K. (2013) The first microbiological contamination assessment by deep-sea drilling and coring by the D/V Chikyu at the Iheya North hydrothermal field in the Mid-Okinawa Trough (IODP Expedition 331). *Front. Microbiol.* **4**, 327.

Yanagawa K., Breuker A., Schippers A., Nishizawa M., Ijiri A., Hirai M., Takaki Y., Sunamura M., Urabe T., Nunoura T., and Takai K. (2014) Microbial Community Stratification Controlled by the Subseafloor Fluid Flow and Geothermal Gradient at the Iheya North Hydrothermal Field in the Mid-Okinawa Trough (Integrated Ocean Drilling Program Expedition 331). *Appl. Environ. Microbiol.* **80**, 6126-6135.

Chapter 5

Conclusions

The work presented here will contribute greatly to the ongoing attempts to understand the microbial biosphere deep within seafloor sediments in terms of a temperature-dependent and biogeochemical context. Since the beginning of marine subsurface research, which originally stemmed from the discovery of life at deep-sea hydrothermal vents, a central question pursued internationally has been the vertical extent of Earth's habitable zone and whether temperature represents the most significant boundary of biosphere reaches. While studies from high temperature hydrothermal vent fluids have attempted to resolve upper temperature limits of microbial life, few studies have examined such hyperthermophilic life deep beneath marine sediments. This set of studies investigated the potential for taxonomic and functional changes across a temperature gradient with depth induced from subsurface hydrothermal fluid flow with the intent to distinguish between established horizons of meso-, thermo-, and hyperthermophilic zones. Taxonomic marker gene and metagenomic sequencing datasets both showed a significant contribution of Archaea in the sediment profile, particularly in higher temperature horizons. Both datasets and geochemical data exhibit evidence for an active methane-oxidizing community, and a distinct and localized thermophilic methane-oxidizing community in the thermophilic zone. Metagenomic data also revealed evidence for a specific hyperthermophilic community that may have once been more prevalent with potentially hotter temperatures. Overall, the presented research has implications for the tenacity of life under in the marine subsurface, the extent of the biosphere in terms of a geothermal gradient, and the significance of methane cycling by deep subsurface microbial communities.

This work also speaks to the challenges and complexities faced when working with sediments where biomass is generally low and organisms generally only distantly relate to

documented surface life. The importance of quality controls and phylotype vetting was not only evidenced by the DNA datasets, but especially in the attempts of extracting RNA from environmental samples. Though the outcome of the RNA study was unanticipated, these studies have provided a wealth of new 16S rRNA gene sequences and metagenomic datasets, and potentially new taxa, that can continue to be used for further molecular biological studies of marine research.

Appendix A

Supplemental Materials for Chapter 2

Table **A-1**: Sample information: IODP identifier (Site-core-section), corresponding depth, sequencing yields, and sedimentary and temperature information. Samples with less than 16 sequences (the extraction blank yield) were not used in further analyses. Samples with no amplification were not sequenced.

IODP Sample Number	Top Depth (mbsf)	Bottom Depth (mbsf)	No amplification from PCR	454 Sequencing Yield (V6-V9 universal primers)	Illumina Sequencing Yield: Bacterial; Archaeal (V6 bacterial and archaeal primers)	Sediment Type	Estimated Temp. (°C)	Measured Temp. (°C)	MG-RAST ID
C0015B-1H-1, 30.0-45.0 cm	0.3	0.45		4850		Pumiceous Gravel			4633466
C0014D-1H-1, 23.0-38.0 cm	0.23	0.38		3349	646766; 554173	Clay	5.26		4633456
C0014G-1H-1, 28.0-40.0[#]	0.28	0.4				Clay	5.42		
C0014B-1H-1, 35.0-45.0 cm	0.35	0.45		5437	592568; 757769	Silty Clay	5.66		4633437
C0014B-1H-1, 122.0-142.0 cm	1.22	1.42		1848		Clay	8.53		4633452
C0014G-1H-2, 45.0-57.0 cm	1.87	1.99		6287		Silty Clay	10.67		4633471
C0014D-1H-2, 57.0-77.0 cm	1.97	2.17		1739		Sandy Silt	11.00		4633464
C0014B-1H-2, 110.0-120.0 cm	2.52	2.62		4637	289603; 341982	Clay	12.82		4633438
C0014D-1H-3, 27.0-42.0 cm	3.075	3.225		1024	351767; 633536	Pumiceous Gravel - Matrix	14.65		4633442
C0014G-1H-3, 100.0-112.0 cm	3.83	3.95		514		Sandy Clay	17.14		4633446
C0014B-1H-3 (454 duplicate),	3.86	4.08		141/44	292952; 95121	Clay	17.24		4633439.3 ;
C0014G-1H-4, 0.0-20.0 cm	4.225	4.425		3360		Pumiceous Gravel - Matrix Supported	18.44		4633447
C0014D-1H-4, 0.0-10.0 cm	4.225	4.325		60	179358; 596865	Pumiceous Grit	18.44	21	4633469
C0014B-1H-4 (454 duplicate),	5.33	5.45		242/292	341807; 35808	Clay	22.09		4633440.3 ;
C0014G-1H-5, 42.0-57.0 cm	6.06	6.21		2765		Hydrothermal Clay - Horizon with hydrothermal origin	24.50		4633472

C0014B-1H-5, 75.0-95.0 cm*	6.39	6.59		1901		Silty Clay	25.59	22	4633459
C0014D-2H-1, 17.0-32.0 cm	6.67	6.82		1389	241181; 42238	Pumiceous Gravel - Clast	26.51		4633470
C0014G-1H-6, 105.0-120.0 cm	8.1	8.25		5614		Pumiceous Gravel - Clast Supported	31.23		4633448
C0014D-2H-2, 75.0-95.0 cm	8.63	8.83		837	176555; 297037	Pumiceous Gravel - Matrix	32.98		4633467
C0014B-2H-3, 20.0-35.0 cm	8.77	8.92		2045	270289; 203916	Pumiceous Grit - Matrix	33.44		4633460
C0014D-2H-3, 88.0-103.0 cm	10.17	10.32		1809	237182; 370577	Gradation from Clayey Hydrothermal Sand with Mineralized	38.06	15	4633468
C0014D-2H-4, 66.0-81.0 cm	11.38	11.53		13		Hydrothermal Clay - Horizon with hydrothermal origin	42.05		4633443
C0014D-2H-6, 19.0-34.0 cm	12.795	12.945		70		Hydrothermal Clay - Horizon with hydrothermal origin	46.72		4633444
C0014B-2H-7, 50.0-70.0 cm*	12.89	13.09		2403		Hydrothermal Clay - Horizon with hydrothermal origin	47.04		4633461
C0014B-2H-10, 30.0-44.0 cm	15.225	15.365		670	168476; 364439	Hydrothermal Clay - Horizon with hydrothermal origin	54.74		4633462
C0014G-2H-5, 127.0-142.0 cm	16.065	16.215		34		Hydrothermal Clay - Horizon with hydrothermal origin	57.51	55	4633449
C0014B-3H-2, 62.0-77.0 cm	17.46	17.61		0		Hydrothermal Clay - Horizon with hydrothermal origin	62.12		
C0014E-1H-4, 70.0-90.0 cm	19.67	19.87		8		Hydrothermal Clay - Horizon with hydrothermal origin	69.41		4633445
C0014B-3H-5, 0.0-20.0 cm	19.855	20.055		12		Hydrothermal Clay - Horizon with hydrothermal origin	70.02		4633441
C0014G-3H-2, 65.0-81.0 cm	19.99	20.15	x			Hydrothermal Clay - Horizon with hydrothermal origin	70.47		
C0014B-3H-7, 99.0-119.0 cm*	22.865	23.065		2194		Hydrothermal Clay - Horizon with hydrothermal origin	79.95		4633463

C0014B-3H-9, 47.5-67.5 cm	24.76	24.96	x		Hydrothermal Clay - Horizon with hydrothermal origin	86.21		
C0014B-4H-3, 17.0-37.0 cm	27.64	27.84	x		Clayey hydrothermal sand; Poorly sorted clay and sand comprising hydrothermally altered and mineralized material	95.71		
C0014E-2H-6, 85.0-100.0 cm	30.805	30.955	x			106.16		
C0014G-4H-5, 60.0-75.0 cm	31.01	31.16	x		Hydrothermal Clay - Horizon with hydrothermal origin	106.83		
C0014B-4H-6, 93.0-103.0 cm	31.47	31.57	x		Hydrothermal Clay - Horizon with hydrothermal origin	108.35		
C0014E-2H-7, 65.0-80.0 cm	32.02	32.17	x			110.17		
C0014E-2H-8, 55.0-70.0 cm	33.33	33.48	x			114.49		
C0014B-4H-8, 100.0-115.0 cm*	34.34	34.49		2445	Hydrothermal Clay - Horizon with hydrothermal origin	117.82		4633453
C0014G-5H-3, 11.0-26.0 cm	38.145	38.295		2606	Hydrothermal Clay - Horizon with hydrothermal origin	130.38		4633465
C0014B-5H-12, 49.0-59.0 cm	41.075	41.175		1003	Hydrothermal Clay - Horizon with hydrothermal origin	140.05		4633454
C0014B-5H-15, 65.0-80.0 cm*	44.51	44.66		1585	Hydrothermal Gravel - Matrix Supported	151.38	150	4633455
C0014G-21H-3, 0.0-15.0 cm	99.105	99.255	x		Hydrothermal Gravel - Matrix Supported	331.55		
C0014G-24T-2, 39.0-54.0 cm	110.09	110.24	x		Hydrothermal Gravel - Matrix Supported	367.80		
Extraction Blank - 28 PCR cycles				16				4633450
Extraction Blank - 35 PCR cycles				405				4633451
* notes 34 PCR cycles								

Supplemental Discussion of Methane Data

Figure A-1 shows several shipboard porewater chemistry measurements. The data shown in Figure A-1(C) are the carbon isotopic measurements of methane samples collected from cores for land-based analyses. The three open diamonds represent safety gas samples, or samples in which methane and sulfide were noticeably degassing on the core cutting deck, implying extremely high concentrations of methane. Thus, it should be noted that the methane concentration measurements in Figure A-1(B) at those depths are likely not accurate, as it was necessary for the core to sit on deck to degass. These depths are indicated in Figure A-1(B) by horizontal dashed lines. The average of these three void gas measurements, also considered to represent the source gas in this study, is represented by the vertical dashed line in Figure A-1(C).

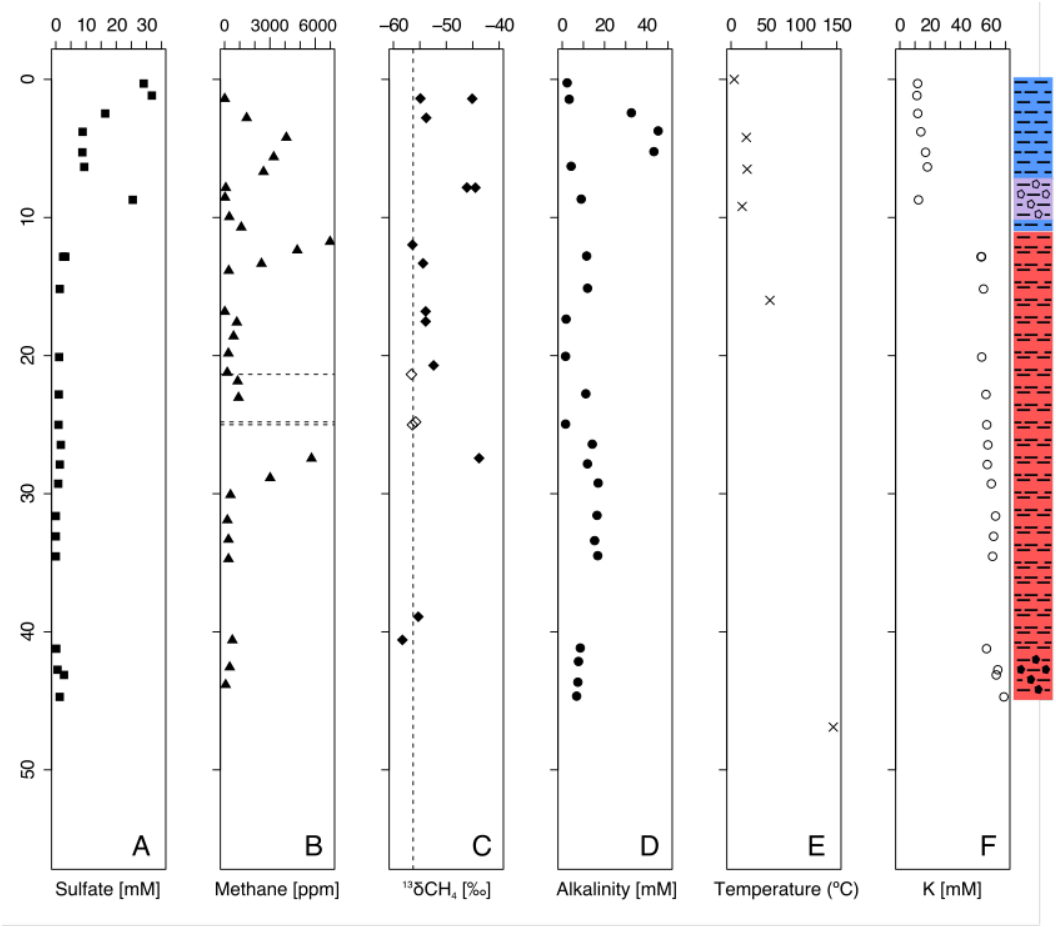

Figure **A-1**: Geochemical profiles with depth of IODP Site C0014 core. (A) Sulfate concentrations from Site C0014B reported in mM. (B) Methane concentrations from headspace gas samples of Site C0014B reported in ppm. The dashed lines represent the depths of collected safety gas samples (noticeable degassing on core cutting deck). (C) $\delta^{13}CH_4$ measurements from Site C0014 (Holes B and D) samples reported in ‰ VPDB. The open diamonds represent the values of the safety-gas samples. The dashed vertical line is the average of the safety gas values. (D) Total alkalinity reported in mM. (E) Temperature measurements in °C. (F) Potassium concentrations reported in mM. Abrupt change in K reflects the change in clay lithologies with depth. The lithologic representation is a modification from Takai *et al.* 2011. The first blue unit represents dark grayish brown silty clay. The purple unit represents pumiceous gravel/grit with dark grayish brown clay matrix. The first red unit represents a pale gray, heavily undurated hydrothermally altered clay. The deepest red unit represents a pale gray, heavily undurated hydrothermally altered clay with indurated mud clasts present.

Table **A-2**: List of all IODP Expedition 331 samples plotted in Figure A-1(C) and their corresponding depth and $\delta^{13}CH_4$ measurements. The depths are the averages of the Top and Bottom Depths.

IODP sample name	Depth (mbsf)	$\delta^{13}CH_4$ (‰ VPDB)
C0014D-1H-1 WR, 136.0-140.0 cm	1.38	-54.9
C0014B-1H-1 WR, 138.0-142.0 cm	1.40	-47.0
C0014D-1H-2 WR, 136.5-140.5 cm	2.79	-53.8
C0014B-2H-1 WR, 133.0-137.0 cm	7.85	-46.1
C0014D-2H-1 WR, 134.0-138.0 cm	7.86	-44.5
C0014D-2H-4 WR, 123.5-127.5 cm	12.00	-56.4
C0014B-2H-7 WR, 93.0-97.0 cm	13.34	-54.4
C0014B-3H-1 WR, 80.0-84.0 cm	16.82	-53.2
C0014E-1H-2 WR, 136.5-140.5 cm	17.54	-53.9
C0014G-3H-2 WR, 137.0-141.0 cm	20.73	-52.4
C0014B-4H-2 WR, 136.5-140.5 cm	27.45	-43.8
C0014G-5H-4 WR, 31.5-35.5 cm	38.93	-55.3
C0014B-5H-12 WR, 0.0-4.0 cm	40.60	-58.3
C0014G-16T-2 WR, 0.0-4.0 cm	76.70	-56.7
C0014E-1H-4 WR, 25.0 cm*	19.22	-56.4
C0014E-1H-5 WR, 98.0 cm*	21.36	-56.6
C0014G-3H-7 WR, 34.0 cm*	24.81	-55.8
* Presumed source gas horizons		

Identification of External or Background DNA

Due to the low concentrations of DNA of most sediment samples, a negative control carried through the extraction process was sequenced to account for any background DNA from the extraction kits. To account for any signal from the extraction kit in all samples, classification of reads was examined at the “fully expanded” taxonomic depth from the SILVA pipeline output,

and all lineages present at the "order" level in the extraction blank in any amount were flagged and in all samples. Similarly, to account for external contamination from drilling processes, taxonomic "orders" identified from the seawater gel 16S rRNA clone analysis in Yanagawa *et al.*, 2013 were flagged if they 1) represented 5% or more of clones from their contamination analysis (includes data from holes B, E, and G), or 2) appeared in more than one hole from IODP Expedition 331 Site C0014.

Taxonomic Classification Discrepancies

The recent emergence of Thaumarchaeota, the deeply branching phylum within the Achaea, has spawned some archaeal classification disparities within 16S rRNA databases, namely, the SILVA SSU and the RefSSU databases. While most non-Euryarchaeota sequences were classified under the Thaumarchaeota phylum in the 454 dataset, the Illumina dataset referenced them as Crenarchaeota. To circumvent the ambiguities that arise, we have referred to this phylum as a combination of the two ("Thaum- and Crenarchaeota").

Supplemental Discussion of Amplicon Data

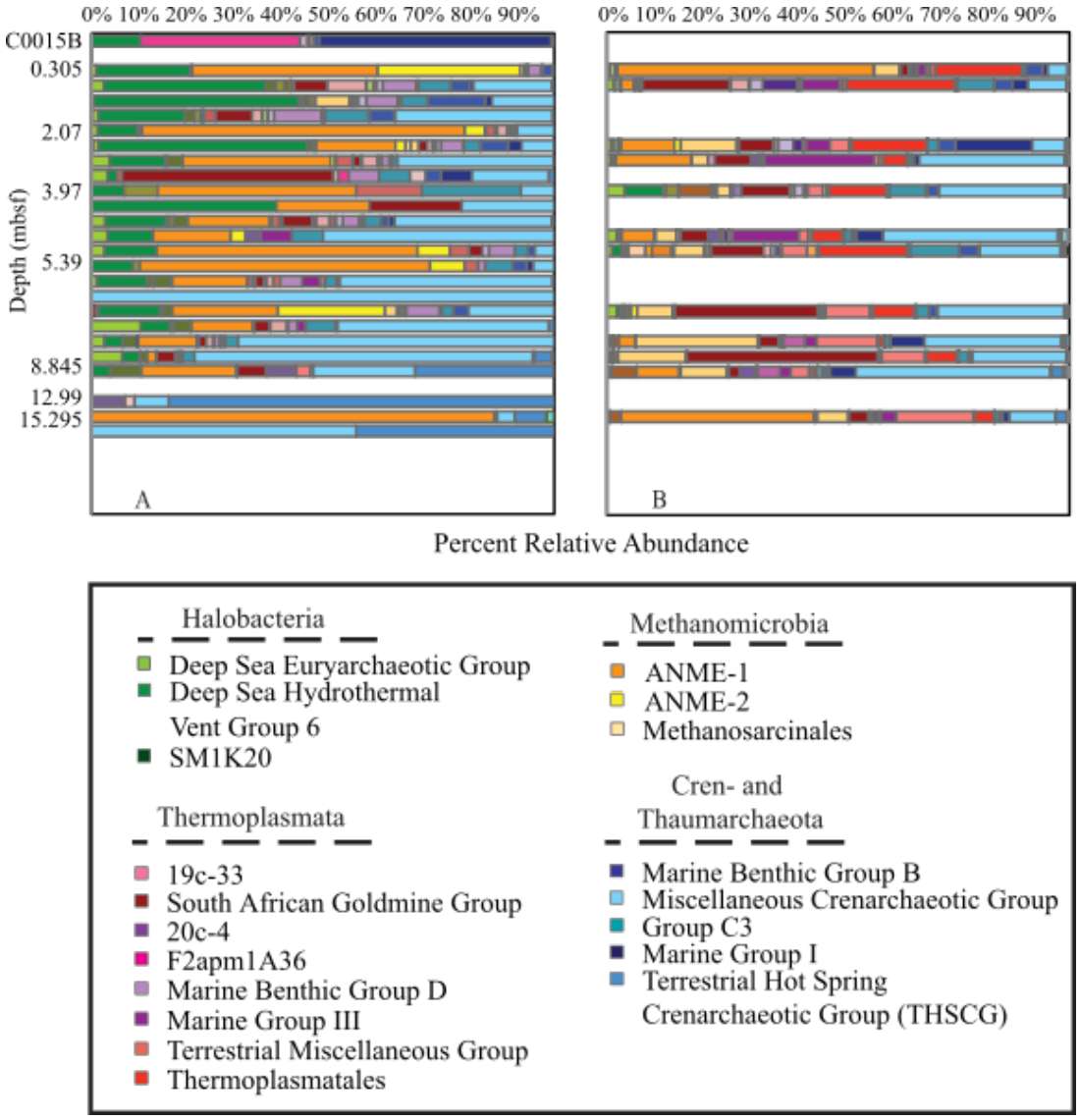

Figure **A-2**: Relative sequence abundances of archaeal 16S rRNA amplicons interpreted at the class level from 454 and Illumina sequencing efforts. Sample horizons are listed by increasing depth below seafloor. Site C0015, 600 m northwest and upslope of the hydrothermal vent, (shown separately as the topmost sample) showed no current hydrothermal activity and is being compared to represent non-hydrothermal conditions within the Iheya North Field. (A) Taxonomic dataset from 454 sequencing, using V6-V9 universal primers. Sequencing data were classified through the SILVA NGS pipeline. (B) Taxonomic dataset from Illumina sequencing, using V6 archaeal primers. Sequencing data were classified through the VAMPS pipeline.

Select samples were amplified using an archaeal primer set targeting the V6 hypervariable region and sequenced with Illumina technology (Table A-1 and Figure A-2). In order to demonstrate changes and trends in microbial diversity throughout the sediment column,

Figure A-2 shows archaeal sequences resolved to a deeper taxonomic level using two sequencing efforts. Between both datasets, there is good correspondence with respect to the observed proportion of Miscellaneous Crenarchaeotic Group (MCG) increasing with depth. Members of the highly diverse MCG are globally distributed in various marine and continental environments and are likely heterotrophic, using organic carbon derived from degradation of recalcitrant, fossil organic matter (Biddle et al., 2006; Kubo et al., 2012). Since uncultured representatives of the MCG defined only by 16S rRNA sequences are distinct from cultured Crenarchaeota, their ecological role in the subsurface is unclear (Inagaki et al., 2003; Kubo et al., 2012). Studies indicate that the MCG community is not active in methane or sulfur cycling (Biddle et al., 2006; Kubo et al., 2012), which agrees with the observed tradeoff in relative abundances between MCG and Methanomicrobia.

In both datasets, orders within Methanomicrobia are common throughout, with high abundances of ANME-1 at 0.305 mbsf and 15.295 mbsf. However, ANME-1 is overrepresented in the 454 sequencing results (Figure A-2(A)) relative to those of Illumina Sequencing (Figure A-2(B)), while it appears that the Illumina results have enhanced discrimination between ANME-1 and Methanosarcinales. Anaerobic methanotrophic archaea (ANME) are members of a microbial consortium involved in the anaerobic oxidation of methane in anoxic marine sediments (Boetius et al., 2000). The high relative abundance of ANME-1 represented in the 15.295 mbsf (~55°C) horizon indicates a potential methane oxidizing niche in the thermophilic regime. Although the magnitudes of ANME-1 relative abundances in Figure A-2 are different between the two datasets, their consistent presence throughout the sediment profile suggests that methanotrophy is an important process in this hydrothermal environment.

The 454 dataset (Figure A-2(A)) shows an overall decreasing trend in the Halobacteria and Methanomicrobia (*e.g.* Deep Hydrothermal Vent Euryarchaeotic Group 6 (DHVEG-6) and ANME-1, respectively) through 10.245 mbsf, where neither taxonomic class appears in the 12.87

or 12.99 mbsf horizons. The apparent tradeoff between the Halobacteria and Methanomicrobia classes and the MCG through the top 10.245 mbsf suggest that the MCG is less impacted by the increasingly temperature. Beginning at the 8.845 mbsf horizon, the Terrestrial Hot Spring Crenarchaeotic Group (THSCG) increases in relative abundance. At the 12.99 mbsf horizon, Archaea represent the majority of indigenous sequences (Figure 1), where the THSCG represent ~80% of archaeal sequences.

Interestingly, the taxa in the IODP Expedition 331 Site C0014 sediments are different than those from the surface sediments of IODP Expedition 331 Site C0015 (Figure A-2(A)). The only commonality between the two sites is the presence of DHVEG-6. Unlike Site C0014, the upslope inactive Site C0015 shows virtually no taxa from Methanomicrobia. Approximately one-third of the archaeal sequences are represented by an uncultured Thermoplasmata, F2apm1A36, and nearly half of the archaeal sequences are represented by Marine Group I. Marine Group I has been found in surface layers of oxidized, organic-poor marine sediments (Teske, 2006; Teske and Sørensen, 2008) and seawater as prokaryotic picoplankton (DeLong et al., 1994; Teske and Sørensen, 2008). Additionally, culturing efforts have determined that Marine Group I represents aerobic, chemolithoautotrophic, nitrifying archaea that oxidize ammonia to nitrite (Könneke et al., 2005; Teske and Sørensen, 2008). Site C0015 exhibits an abundant occurrence of very permeable layers of pumice and volcaniclastic sediments, which has yielded porewater geochemistry profiles that are indistinguishable from seawater, suggesting recharge of seawater into the sediments (Takai et al., 2011a). The presence of Marine Group I and evidence for a locally oxic surface layer suggests that surface processes are different between Sites C0014 and C0015, which ultimately shape the microbial community.

While the Illumina dataset (Figure A-2(B)) differs in relative abundances among taxa compared with our primary sequence data from the V6-V9 amplicons (Figure A-2(A)) in other represented taxa, the additional sequencing effort did fortunately enhance discrimination in the

Thermoplasmata class (*i.e.* South African Gold Mine Group (SAGMEG), Thermoplasmatales, and the Terrestrial Miscellaneous Group). The results demonstrate the need for caution when working with taxonomic datasets from primer-amplified environmental DNA. Although an archaeal-specific primerset was used to complement the primary dataset, many archaeal lineages contain numerous mismatches compared with internal PCR primers and may be underrepresented (*e.g.* DHVEG-6, MG-I, and Ancient Archaeal Group) (Teske and Sørensen, 2008). Our work here aimed at effectively resolving key taxonomic groups in the heated hydrothermal subsurface of the Iheya North Field through the use of two different primersets. In spite of some variations between the two amplicon datasets, the subsurface sediments at IODP 331 Site C0014 in the Iheya North Field overall exhibit coherent and dramatic shifts in community with depth as temperature and hydrothermal influence increases.

Although the data in Figure A-2(B) were analyzed via the VAMPS pipeline, we also analyzed the Illumina sequencing data with the SilvaNGS 1.1 pipeline (see methods). The results are plotted in Figure S4 and do not show any significant discriminations between the relative abundances of certain taxa, with the exception of better resolution within the Thermoplasmatales (purple and red shaded regions). Therefore, we assume that differences in taxonomic abundances are the result of biases in using different primers, rather than different pipelines and databases.

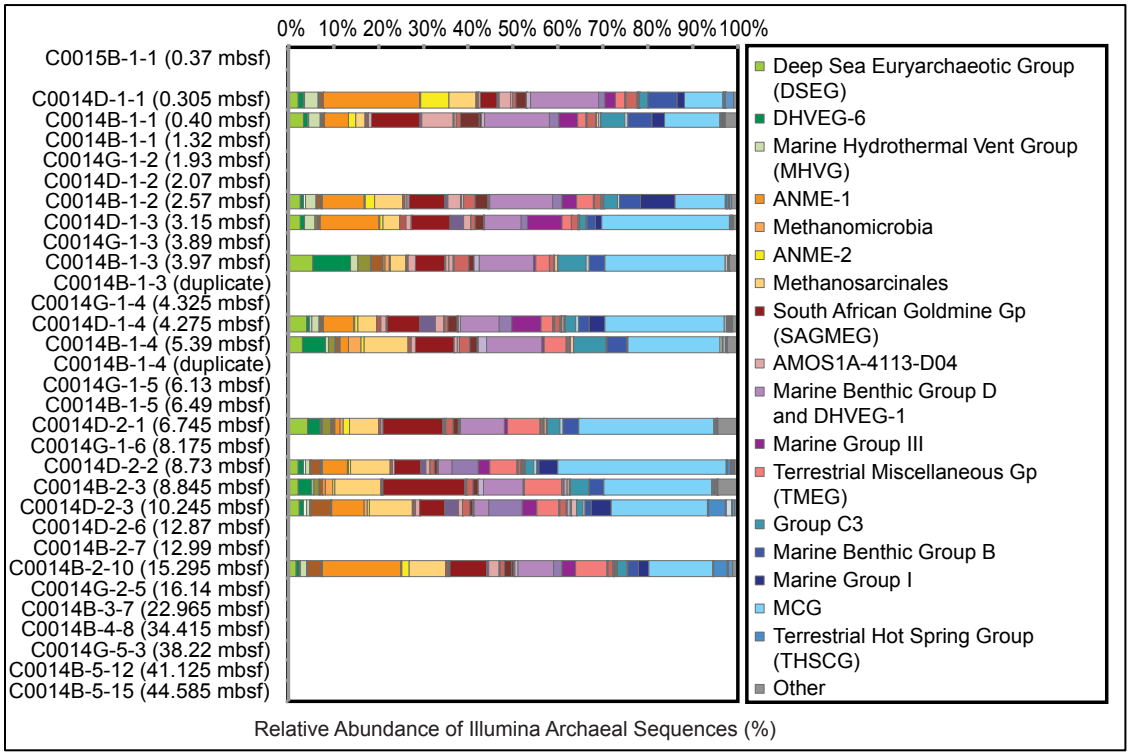

Figure **A-3**: Relative sequencing abundance of Illumina sequencing dataset that was analyzed through the SILVA NGS pipeline.

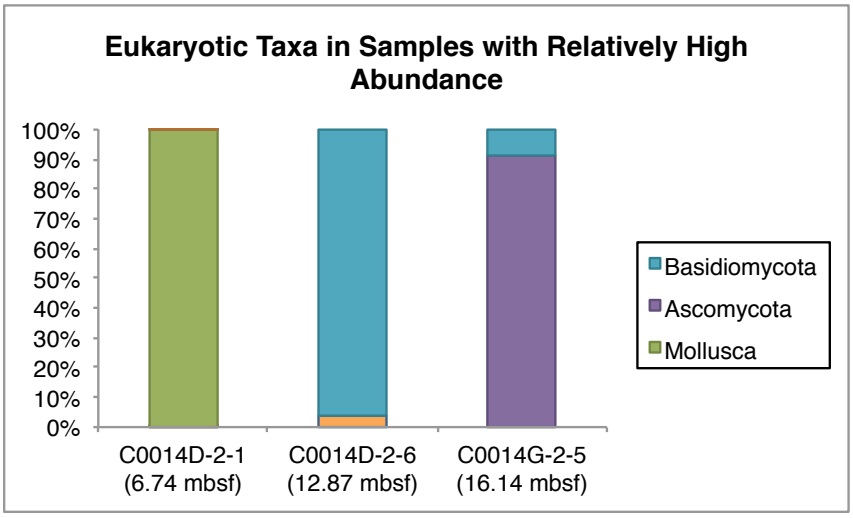

Figure **A-4**: The three horizons 6.74, 12.87, and 16.14 mbsf amplified Eukaryotic sequences (600, 58, and 11 total sequences respectively). Shown here are the relative proportions of those sequences that classified within Eukaryota. Note that C0014D-2-6 and C0014G-2-5 had significantly lower overall sequence yield. Both Basidiomycota and Ascomycota are fungal taxa, while Mollusca is an animal taxon.

Appendix B

Supplemental Information for Chapter 3

Table **B-1**: Table showing the metagenomic sequencing plan of Site C0014 sediment horizons. The table also lists metagenome amplification conditions as well as sequence and assembly yield.

Sample Name	Depth (mbsf)	T(°C)	REPLI-g conditions	Total Number of Sequences	Number Contigs Assembled >= 100 nt
C0014B-1-1	1.32	8	5 hr	40,522,281	25,411
C0014B-1-3	3.97	17		31,259,620	20,009
C0014B-1-4	5.39	22		19,893,784	9,011
C0014B-2-3	8.84	33		27,861,782	21,869
C0014B-2-7	12.99	47		27,784,190	1,226
C0014B-2-10	15.3	55		36,615,912	5,642
C0014B-2-7*	12.99	47	8 hr	42,674,374	650
C0015B-1-1	0.22	5		could not sequence	

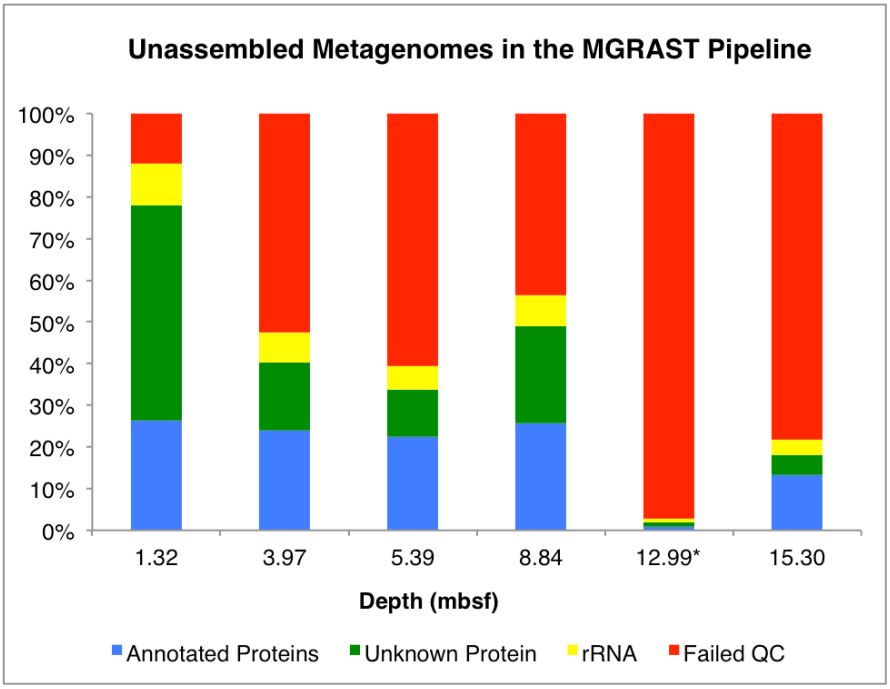

Figure **B-1**: Relative abundance of MG-RAST classified proteins of unassembled metagenomes samples. Sequences associated with "Failed QC" were removed from further sequencing.

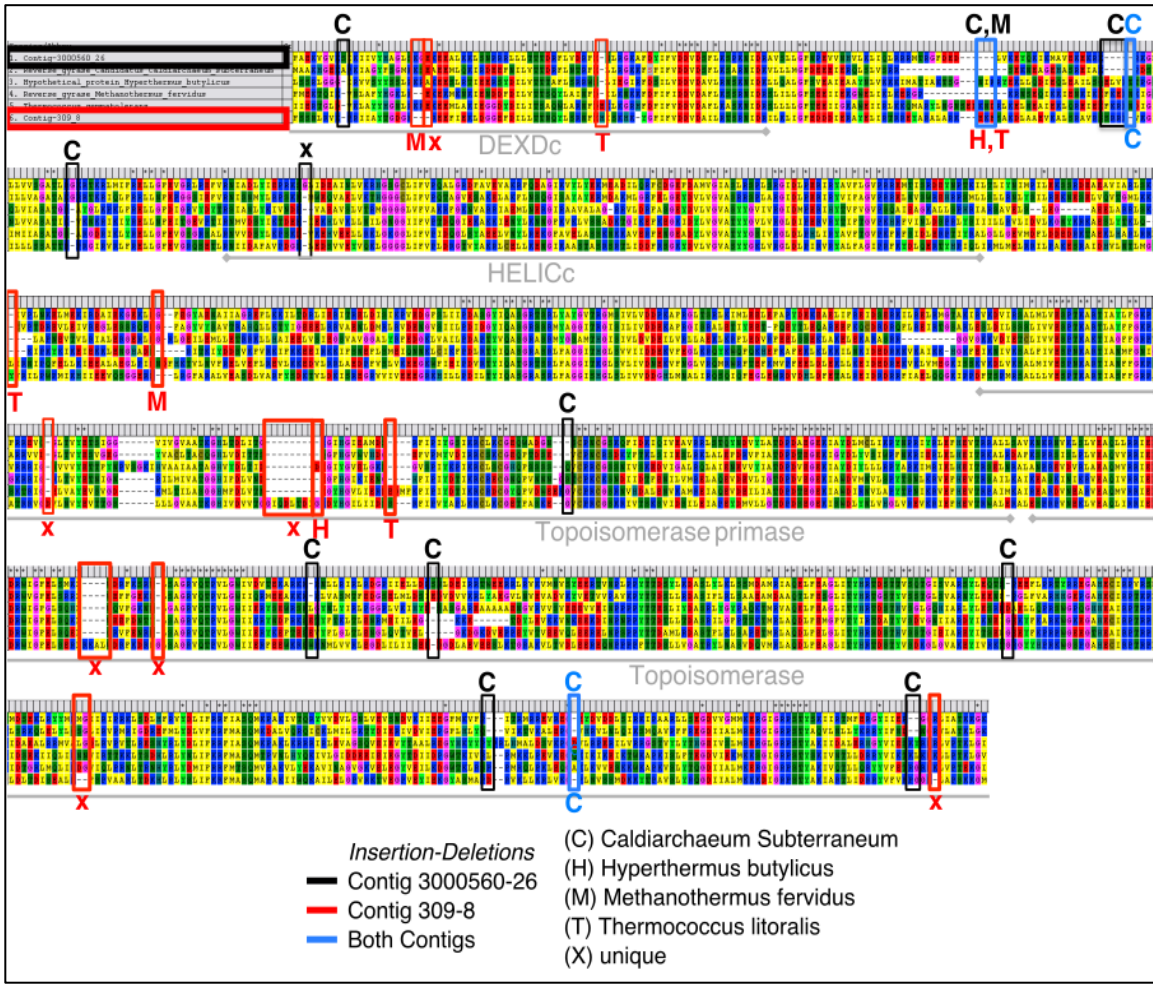

Figure **B-2**: Reverse gyrase protein sequence alignment using MEGA's ClustalW algorithm of Contig 3000560-26 (boxed in the top left in black), followed by Candidatus Caldiarchaeum subterraneum, Hyperthermus butylicus, Methanothermus fervidus, Thermococcus, and Contig 309-8 (boxed in the top left in red). Sequences are referenced from Candidatus Caldiarchaeum subterraneum amino acid positions 1-998. Amino acid insertion-deletions (indels) differentiating the contigs are boxed in black (Contig 3000560-26), red (Contig 309-8), or blue (indels common to both). The annotated sample corresponding to the indel is indicated by a single letter, referenced in the legend. Line segments below the alignment indicate the various annotated protein domains referenced from Candidatus Caldiarchaeum subterraneum.

Table **B-2**: Table showing the top BLASTP hits for C0014B-2-10 Contig 309 predicted genes. Purple: Thaumarchaeota, orange: Euryarchaeota, red: Crenarchaeota, and gray: Bacteria. NCBI protein accession numbers are listed before each protein description.

Contig 309	Top Hit 1	Top Hit 2	Top Hit 3	Top Hit 4
1	WP_052884790 hypothetical protein [Thermofilum sp. 1807-2]	AFU56974 putative anaerobic ribonucleoside-triphosphate reductase [Cand. Nitrososphaera gargensis Ga9.2]	AFK24910 putative anaerobic ribonucleoside triphosphate reductase [uncultured archaeon]	WP_010940207 anaerobic ribonucleoside triphosphate reductase [Desulfovibrio vulgaris]
2	WP_042692318 nascent polypeptide-associated complex protein [Thermococcus nautili]	WP_014737096 nascent polypeptide-associated complex protein [Thermogladius cellulolyticus]	WP_015857947 nascent polypeptide-associated complex protein [Thermococcus gammatolerans]	WP_050002138 nascent polypeptide-associated complex protein [Thermococcus eurythermalis]
3	AIF07761 PUA domain-containing protein [uncultured marine thaumarchaeote KM3_24_H04]	AIF07874 PUA domain-containing protein [uncultured marine thaumarchaeote KM3_25_B05]	AIF11034 PUA domain-containing protein [uncultured marine thaumarchaeote KM3_47_G11]	AIE97378 PUA domain-containing protein [uncultured marine thaumarchaeote AD1000_99_A07]
4	KYH39653 XRE family transcriptional regulator [Candidatus Bathyarchaeota archaeon B26-2]	WP_052885108 transcriptional regulator [Thermofilum pendens]	EKQ51745 TIGR00270 family protein [Methanobacterium sp. Maddingley MBC34]	WP_004032096 transcriptional regulator [Methanobrevibacter smithii]
5	WP_048084106 GTPase HflX [Marine Group I thaumarchaeote]	AIF11036 GTP1/OBG protein (hflX) [uncultured marine thaumarchaeote KM3_47_G11]	WP_019113978 GTPase HflX [Verrucomicrobia bacterium SCGC AAA164-O14]	ABZ09470 putative ADP-ribosylation factor family protein [uncultured marine crenarchaeote HF4000_APKG8D6]
6	AAM02017 SpoU-like RNA methylase [Methanopyrus kandleri AV19]	WP_012572127 tRNA (cytidine(56)-2'-O)-methyltransferase [Thermococcus onnurineus]	WP_062388073 tRNA (cytidine(56)-2'-O)-methyltransferase [Thermococcus peptonophilus]	BAD84249 tRNA ribose 2'-O-methyltransferase aTrm56 [Thermococcus kodakarensis KOD1]
7	AFU56958l transcription factor E [Candidatus Nitrososphaera gargensis Ga9.2]	AIF82179 transcription initiation factor IIE, alpha subunit [Candidatus Nitrososphaera evergladensis SR1]	AIC16298 transcription factor E [Nitrososphaera viennensis EN76]	ACF09613 transcription factor TFIIE alpha subunit [uncultured marine crenarchaeote]
8	KUJ99632 Reverse gyrase [Thermococcales archaeon 44_46]	WP_055283293 reverse gyrase [Thermococcus sp. EP1]	WP_042697994 reverse gyrase [Thermococcus sp. PK]	WP_058946101 reverse gyrase [Thermococcus sp. 2319x1]
9	WP_000049580 molybdenum cofactor guanylyltransferase [Bacillus thuringiensis]	WP_048082697 molybdenum cofactor guanylyltransferase [Methanobacterium]	WP_023523215 molybdenum cofactor guanylyltransferase [Bacillus thuringiensis]	WP_056835836 hypothetical protein [Paenibacillus sp. Soil787]
10	ABK78450 archaeal DNA polymerase II, subunit B [Cenarchaeum symbiosum A]	KRO31785 DNA polymerase, partial [Nitrosopumilus sp. BACL13]	KRO28511 DNA polymerase]Nitrosopumilus sp. BACL13]	WP_048114507 DNA polymerase [Candidatus Nitrosopumilus adriaticus]
11	KYH38084 reverse gyrase [Candidatus Bathyarchaeota archaeon B24]	WP_011821407 reverse gyrase[Hyperthermus butylicus]	WP_055283293 reverse gyrase [Thermococcus sp. EP1]	WP_011839408 reverse gyrase [Staphylothermus marinus]

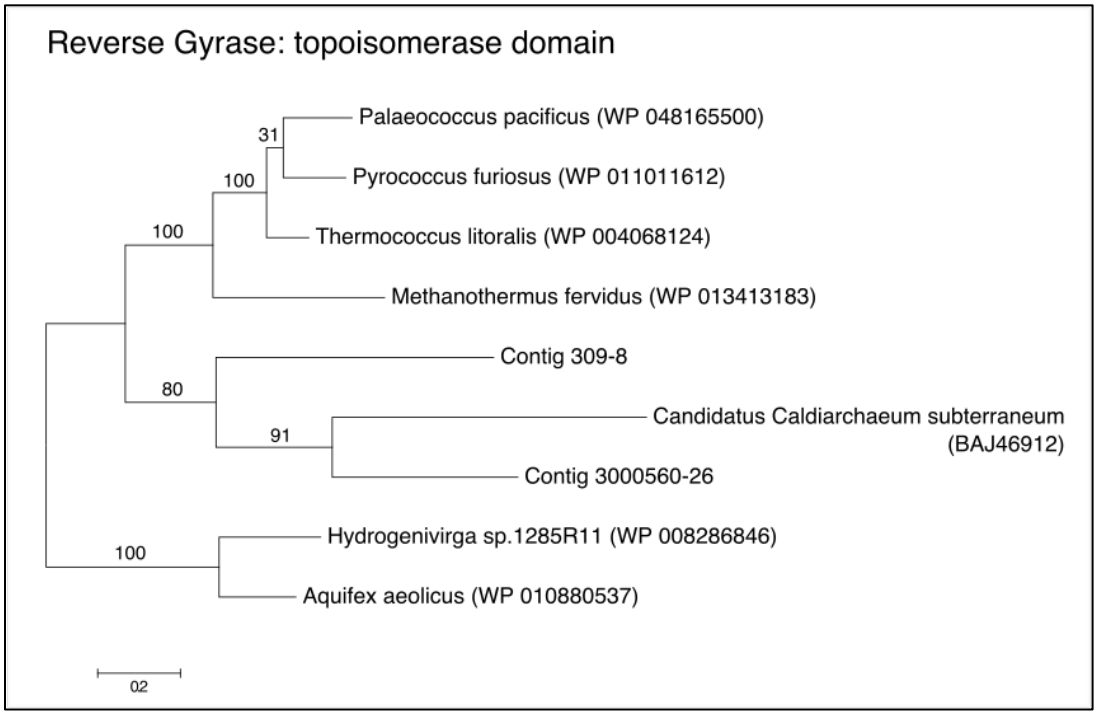

Figure **B-3**: Maximum likelihood tree of the topoisomerase domain of reverse gyrase, referenced from Candidatus Caldiarchaeum subterraneum. All default MEGA parameters were used. Numbers located on the branches refer to boostrap values. NCBI identifiers are in the parentheses next to each taxon name.

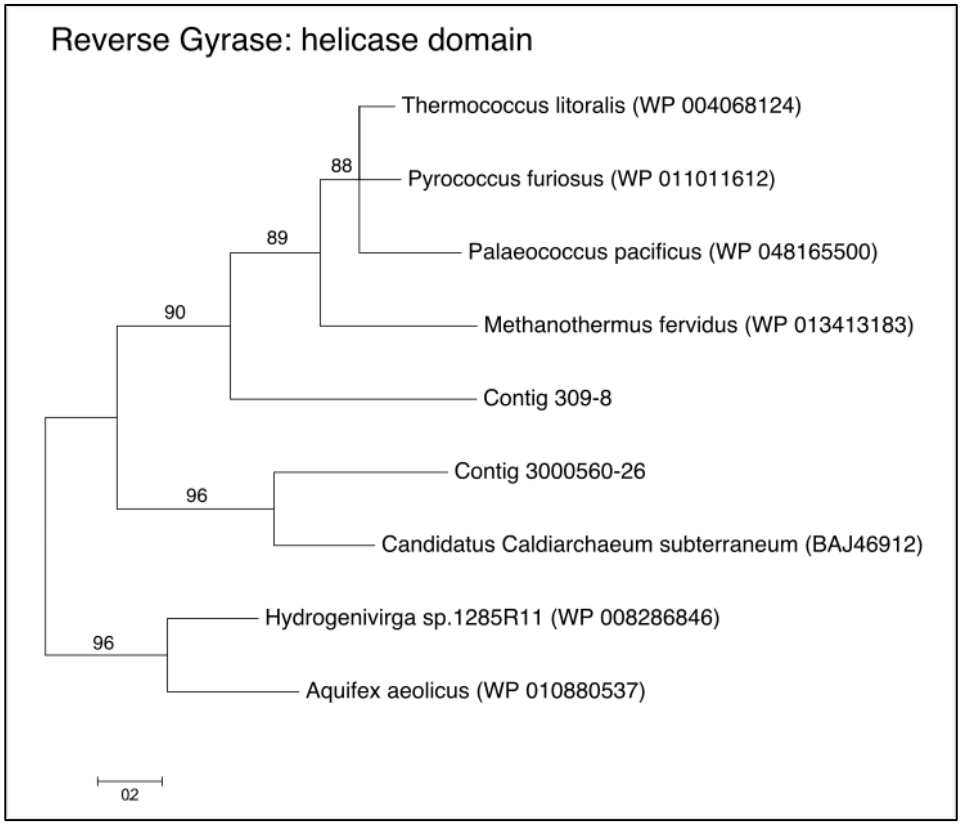

Figure **B-4**: Maximum likelihood tree of the helicase domain of reverse gyrase, referenced from Candidatus Caldiarchaeum subterraneum. All default MEGA parameters were used. Numbers located on the branches refer to boostrap values. NCBI identifiers are in the parentheses next to each taxon name.

Table **B-3**: Table showing the taxonomic association, completeness, contamination, and incorporation of a reverse gyrase gene of C0014B-2-10 analyzed bins.

Bin ID	Marker Lineage	Completeness (%)	Contamination (%)	Reverse Gyrase
ANME Bin	Euryarchaeota (phylum)	73.79	26.62	No
Caldiarchaeum subterraneum Bin	Archaea	75	28.37	Yes
Novel High Temperature Bin	Archaea	13.46	0.96	Yes
Sulfate-Reducing Bacterium Bin	Bacteria	59.79	5.85	No

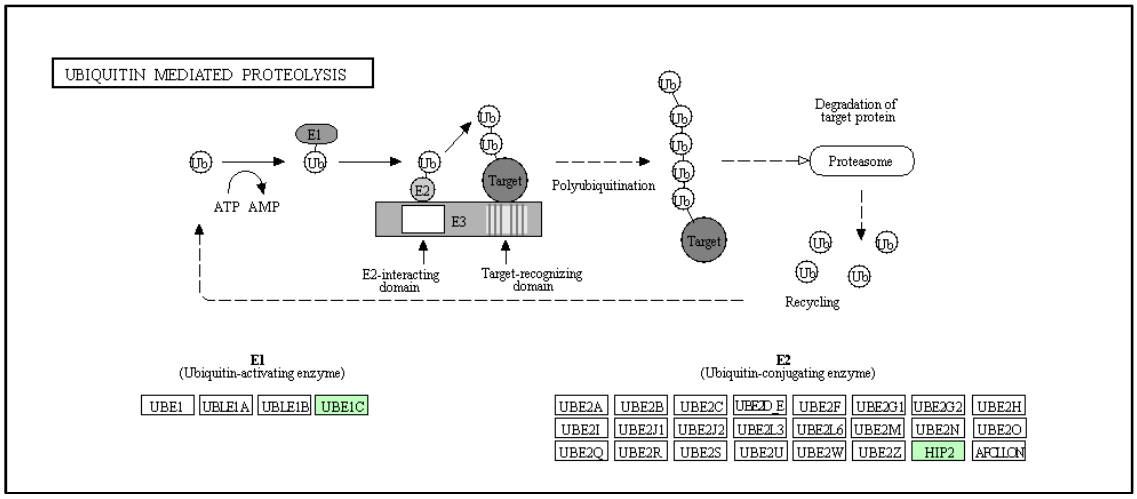

Figure **B-5**: KEGG pathway map of Ubiquitin-mediated-proteolysis. The green shaded regions indicate those genes found in Candidatus Caldiarchaeum subterraneum bin.

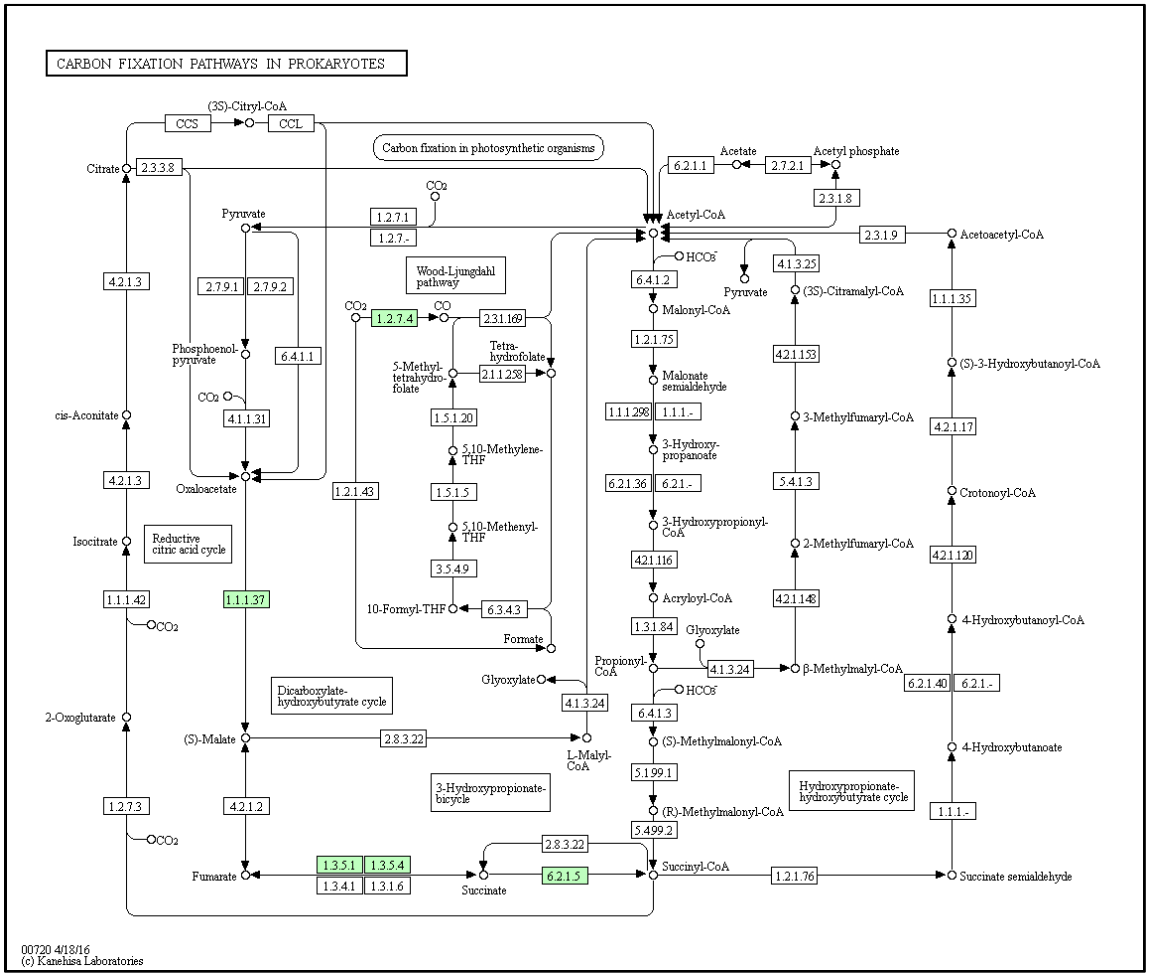

Figure **B-6**: KEGG pathway map of carbon fixation pathways. Genes highlighted in green are those found in the Novel High Temperature Bin.

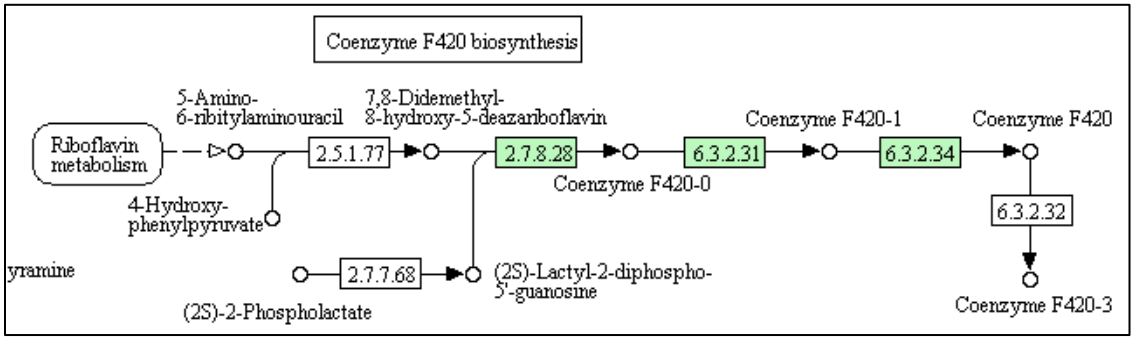

Figure **B-7**: KEGG pathway map of Coenzyme F_{420} biosynthesis. Genes highlighted in green are those found in the C0014B-2-10 Novel High Temperature Bin.

VITA

Leah D. Brandt

EDUCATION

The Pennsylvania State University — Ph.D. December 2016
University Park, PA, USA — Geosciences and Biogeochemistry

The Pennsylvania State University — B.S. August 2009
University Park, PA, USA — Meteorology
Minors: Spanish and Oceanography

SELECTED AWARDS

- Center for Dark Energy Biosphere Investigations (C-DEBI) Graduate Student Fellowship 2014-2016
- Center for Dark Energy Biosphere Investigations (C-DEBI) Small Research Grant, 2013-2016
- Donald B. and Mary E. Tait Scholarship in Microbial Biogeochemistry, 2014
- Chesapeake Energy Scholarship, 2013
- Paul D. Krynine Scholarship, 2012-2014

SERVICE

- College of Earth and Mineral Sciences Graduate Student Council, 2011-2015
- Science Outreach (Shake, Rattle and Rocks; Alien Astronomysteries; Philadelphia Girls in Science Workshop; PJAS panel judge)
- Teaching Assistant (Geosc 040: The Sea Around Us; Geosc 410: Marine Biogeochemistry)
- U7 and U9/10 Soccer Head Coach

PUBLICATIONS

- **Brandt**, L.D. and House, C.H. (2016) Marine subsurface microbial community shifts across a hydrothermal gradient in Okinawa Trough sediments. Accepted October 2016 to *Archaea*
- Martino, A.M., Rhodes, M.E., Biddle, J.F., **Brandt, L.D.**, *et al.* (2012) Novel degenerate PCR method for whole-genome amplification applied to Peru Margin (ODP 201) subsurface samples. *Frontiers in Microbiology* **3**:1-11.
- Yanagawa, K., Nunoura, T., McAllister, S.M., Hirai, M., Breuker, A., **Brandt, L**., *et al.* (2013) The first contamination assessment for microbiological study in deep-sea drilling and coring by the D/V *Chikyu* at the Iheya North hydrothermal field in the Mid-Okinawa Trough (IODP Expedition 331). *Frontiers in Microbiology* **4**: 327