CARBONATE-ASSOCIATED MICROBIAL ECOLOGY AT METHANE SEEPS

ASSEMBLAGE COMPOSITION, RESPONSE TO CHANGING ENVIRONMENTAL CONDITIONS, AND IMPLICATIONS FOR BIOMARKER LONGEVITY

Thesis by

David Hamilton Case

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2016

(Defended May 12, 2016)

David Hamilton Case

ORCID: 0000-0002-1023-0040

DEDICATION

Dedicated with love to,

C. Randy Case & Beth Ann Hamilton,

Charles, Monica, & Madeleine Case,

Peter, Aneta, Michael, & Alexander Case,

V. Edward & D. Jane Hamilton

Charles & Ilyeene Case

ACKNOWLEGEMENTS

I am indebted to my thesis advisor, Victoria Orphan, for providing an open and supportive environment in which to explore science. Perhaps Victoria's greatest strength as a mentor is the smile she imparts on every student as they walk into her office – without exception, I never once left a meeting with her feeling anything but better about my science and myself as a scientist. Such support is a great gift from Victoria to her students – thank you!

In addition, many other thoughtful, helpful, and friendly scientists have rotated through the Orphan lab and provided mentorship and friendship to me during the past six years. In thanking them all, I would like to especially acknowledge Derek Smith, Ally Pasulka, Connor Skennerton, Stephanie Connon, Elizabeth Trembath-Reichert, Roh Bhartia, Hank Yu, Jeff Marlow, Kat Dawson, Patty Tavormina, Roland Hatzenpichler, Silvan Scheller, Anne Dekas, Greg Wanger, and Sean Mullin for playing important parts in my graduate career.

Other mentors and peers at Caltech have also indelibly influenced my experience and shaped the scholar I have become. I have received invaluable advice from my committee members: Woody Fischer, Lisa Levin, Jess Adkins, and Jared Leadbetter. Members of my original graduate class – officemates in the "Pit" – have made the last six years more meaningful than merely an academic experience: Jeff Prancevic, Jena Johnson, Stephen Cox, Sarah Slotznick, and Elizabeth Trembath-Reichert. Paul Asimow, Cassandra Horii, and Daniel Thomas shaped my extracurricular interest in high-quality teaching, which I explored extensively while at Caltech.

My completion of this thesis would not have been possible without the support and influence of countless others. It is impossible to list the names of every person who has touched my life, leading to this accomlishment, and I cannot put into words how grateful I am to the cast of family, friends, colleagues, and peers who have shaped my journey. Acknowledging that a mere list does not sufficiently connote the importance of the following people, and that many important names are not given here, I nonetheless want to note the following people who mean so much to me: Brian Kohan, Chris Marotta, the Kahn family, Joel and Elaine Ziskind, Sophie Hines, Greg Adamson, Kevin Okura, Jason Lopez, Lee Coleman, Susan Chadwick & family, Divina Bautista, Alice Sogomonian, Charles Niesen, Linnea Larson, Ben Willis & the DWA team, Dan Giammar, Keith Troche, Steve Cory, Jaina Krueger, Michelle Pai, Kate Wilson, Gordon Johnston, Andrew Harris, JonMichael Rasmus, Bill Bloomenkranz, Sarah Scallon, Greg Emmerich, Bennett Javenkoski, Brian Stoveken, Jessica Pellegrino, Sharon Stahl, Andrew Harris, Julie Jensen, Joe Spolar, Larry Henderson, and Mary Schmidt.

To everyone, named here and otherwise – *Thank You*

ABSTRACT

Methane seeps are globally distributed geologic features in which reduced fluid from below the seafloor is advected upward and meets the oxidized bottom waters of Earth's oceans. This redox gradient fuels chemosynthetic communities anchored by the microbially-mediated anaerobic oxidation of methane (AOM). Both today and in Earth's past, methane seeps have supported diverse biological communities extending from microorgansisms to macrofauna and adding to the diversity of life on Earth. Simultaneously, the carbon cycling associated with methane seeps may have played a significant role in modulating ancient Earth's climate, particularly by acting as a control on methane emissions.

The AOM metabolism generates alkalinity and dissolved inorganic carbon (DIC) and at a 2:1 ratio, promoting the abiogenic, or authigenic, precipitation of carbonate minerals. Over time, these precipitates can grow into pavements covering hundreds of square meters on the seafloor and dominating the volumetric habitat space available in seep ecosystems. Importantly, carbonates are incorporated into the geologic record and therefore preserve an inorganic (i.e., $\delta^{13}C$) and organic (i.e., lipid biomarker) history of methane seepage. However, the extent to which preserved biomarkers represent a snapshot of microorganisms present at the time of primary precipitation, a time-integrated history of microbial assemblages across the life cycle of a methane seep, or a view of the final microorganisms inhabiting a carbonate prior to incorporation in the sedimentary record is unresolved.

This thesis addresses the ecology of carbonate-associated seep microorganisms. Chapters One and Two contextualize the extant microbial diversity on seep carbonates versus within seep sediments, as determined through 16S rRNA gene biomarkers. Small, protolithic carbonate "nodules" recovered from within seep sediments are observed to be capable of capturing surrounding sediment-hosted microbial diversity, but in some cases also diverge from sediments. Meanwhile, lithified carbonate blocks recovered from the seafloor host microbial assemblages demonstrably distinct from seep sediments (and seep nodules). Microbial 16S rRNA gene diversity within carbonate samples is well-differentiated by the extent of contemporary seepage. *In situ* seafloor transplantation experiments further demonstrated the microbial assemblages associated with seep carbonates to be sensitive to seep quiescence and activation on short (13-month) timescales. This was particularly true for organisms whose 16S rRNA genes imply physiologies dependent on methane or sulfur oxidation. With an improved understanding of the modern ecology of carbonate-associated microorganisms, Chapter Three applies intact polar lipid (IPL) and core lipid analyses to begin describing whether, and to what extent, geologically relevant biomarkers mimic short-term dynamics observed in 16S rRNA gene profiles versus archive a record of historic microbial diversity. Biomarker longevity is determined to increase from 16S rRNA genes to IPLs to core lipids, with IPLs preserving microbial diversity history on timescales more similar to 16S rRNA genes than core lipids. Ultimately, individual IPL biomarkers are identified which may be robust proxies for determining whether the biomarker profile recorded in a seep carbonate represents vestiges of active seepage processes, or the profile of a microbial community persisting after seep quiescence.

PUBLISHED CONTENT AND CONTRIBUTIONS

The material presented in Chapter One is published in Mason & Case et al. (2015). In the publication, D.H.C. and O.U.M. share co-first authorship. D.H.C. performed iTag processing and analysis, beta diversity analyses, and was the principal and coordinating author of the manuscript. O.U.M. processed the samples and optimized DNA extraction, as well as terminal restriction fragment length polymorphism and clone library analyses. V.J.O. conceived of the study and collected samples at sea. T.H.N. provided X-ray diffraction data, R.W.L. performed the isotopic composition analyses, J.V.B. provided thin section images, and R.B.T. performed the pore water geochemical measurements. D.H.C., O.U.M., and V.J.O. principally wrote the manuscript.

Mason*, O.U., Case*, D.H., Naehr, T.H., Lee, R.W., Thomas, R.B., Bailey, J.V., Orphan, V.J., 2015. Comparison of Archaeal and Bacterial Diversity in Methane Seep Carbonate Nodules and Host Sediments, Eel River Basin and Hydrate Ridge, USA. Microbial Ecology 70, 766-784. doi:10.1007/s00248-015-0615-6. *indicates shared first authorship.

The material presented in Chapter Two is published in Case et al. (2015). D.H.C. performed the lab work, prepared the samples for sequencing, analyzed and interpreted all data, and was the principal and coordinating author of the manuscript. V.J.O. and L.A.L. led field sampling and coordinated the seafloor experiments. All authors, including A.L.P., J.J.M., and B.M.G., provided intellectual and writing contributions.

Case, D.H., Pasulka, A.L., Marlow, J.J., Grupe, B.M., Levin, L.A., Orphan, V.J., 2015. Methane Seep Carbonates Host Distinct, Diverse, and Dynamics Microbial Assemblages. mBio 6, e01348-15. doi:10.1128/mBio.01348-15.

The material presented in Chapter Three is in preparation for publication by the listed authors. D.H.C. is the principal and coordinating author of the manuscript, acquired all iTag and geochemical data, performed all computational analyses, and generated interpretations of the data. M.Y.Y. and A.J.G. performed bench-top lipid extraction and quantification. V.J.O. led field sampling and coordinated the seafloor experiments. All authors, including A.L.P., K.S.D., and K.U.H., provided intellectual and writing contributions.

Case, D.H., Yoshinaga, M.Y., Pasulka, A.L., Dawson, K.S., Greve, A.J., Hinrichs, K.U., Orphan, V.J., in preparation. Extent and Rate of Carbonate-Associated Microbial Biomarker Turnover During *in situ* Simulated Methane Seep Quiescence.

The material presented in Chapter Four is available as a laboratory resource and will be openly provided to any researchers interested in the results. To facilitate in dissemination of this data, the authors are preparing to make the chapter and data publically available online, including as part of this thesis through thesis.library.caltech.edu. D.H.C. and S.A.C. coordinated bench-top lab work. D.H.C. performed data analyses and generated interpretations. All authors, including A.L.P., E.T.-R., K.S.D., C.T.S., and V.J.O., provided intellectual and writing contributions.

Case, D.H., Pasulka, A.L., Trembath-Reichert, E., Connon, S.A., Dawson, K.S., Skennerton, C.T., Orphan, V.J. Development of an In-House Preparation, Processing, and Analysis Workflow for Next Generation Sequencing Data. Freely available by request from the authors.

The material presented in Appendix One is in preparation for publication by the listed authors. D.H.C. is the principal and coordinating author of the manuscript, performed all microbiological bench-top laboratory work, and generated interpretations of the data. A.I. provided geochemical data. P.T. helped with clone library assays. D.H.C., Y.M., A.I., and F.I. conceived of the study. All authors, including V.J.O., provided intellectual and writing contributions.

Case, D.H., Morono, Y., Ijiri, A., Tavormina, P., Orphan, V.J., Inagaki, F., in preparation. *In Situ* Deployable Reactor Enables Examination of Microbial Communities from High Pressure Environments.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

Abbreviated as (Chapter Number)-(Figure/Table Number)

CHAPTER ONE

Tables

Figures

CHAPTER TWO

Tables

Figures

CHAPTER THREE
Tables

Figures

CHAPTER FOUR
Tables

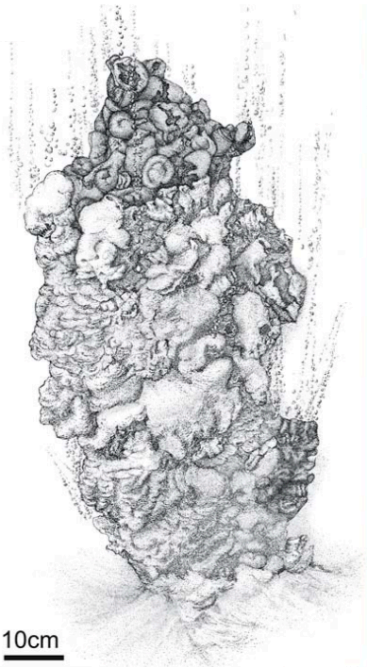
Figures

APPENDIX ONE
Tables

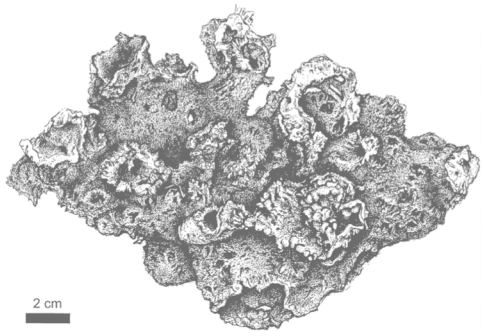
Figures

LIST OF ACRONYMS & ABBREVIATIONS

AEG	acyletherglycerol
Alk	alkalinity
ANME	anaerobic methane-oxidizing archaea
ANOSIM	analysis of similarity
AOM	anaerobic oxidation of methane
AR	archaeol
cmbsf	centimeters below seafloor
DAPI	4',6-diamidino-2-phenylindole
DEG	dietherglycerol
DAG	diacylglycerol
DHVEG	deep sea hydrothermal vent group
DIC	dissolved inorganic carbon
DSV	deep submergence vehicle
EMP	Earth Microbiome Project
ERB	Eel River Basin
HP	high pressure
HR	Hydrate Ridge
GDGT	glyceroldibipytanylglyceroltetraether
IPL	intact polar lipid
iTAG	massively parallel high-depth DNA sequencing (also, "iTag")
MBGB	marine benthic group B
MBGD	marine benthic group D
mbsf	meters below seafloor
mbsl	meters below sea level
NGS	next generation sequencing (see also, "iTAG")
NMDS	nonmetric multidimensional scaling
OH-AR	hydroxyarchaeol
OTU	operational taxonomic unit
PC	phosphatidylcholine (in the context of organic geochemistry)
PC	push core (in the context of seafloor sampling)
PCR	polymerase chain reaction
PG	phosphatidylglycerol
PI	phosphatidylinositol
PS	phosphatidylserine
QIIME	Quantitative Insights Into Microbial Ecology
RFLP	restriction fragment length polymorphism
ROV	remotely operated vehicle
RV	research vessel (also, "R/V")
SIMPER	similarity percentage
SMTZ	sulfate-methane transition zone
SRA	Sequence Read Archive
SRB	sulfate-reducing bacteria
TRFLP	terminal restriction fragment length polymorphism
XRD	X-ray diffraction



Sketch of a carbonate precipitate in the Black Sea from 230 meters below sea level in anoxic waters (Reitner et al., 2005. *Facies* 51: 66-79).



Sketch of a carbonate precipitate in the Black Sea from 188 meters below sea level in anoxic waters (Peckmann et al., 2001. *Marine Geology* 177: 129-150).

"It is not intuitively obvious why a square yard of meadow, say, should not be exactly like the next square yard in species-composition, yet it must have been noted many times, and from the earliest times, that it rarely or never is so."

— F.W. Preston, *Time and Space and the Variation of Species*, 1960

Introduction

~~~

For some time, two concepts have been appreciated concerning methane seeps: (i) At any single point location on the seafloor, seepage is an ephemeral process bound to eventual dormancy, and (ii) carbonate pavements precipitated at methane seeps have high potential to enter the geological record as an archive of historic seepage. For the first time, this thesis applies *in situ*, time-resolved experiments with parallel biomarkers to begin addressing the fundamental question connecting points (i) and (ii):

> To what extent does the taphonomy of microorganisms in seep carbonates represent those species present during active seepage vs those last present before incorporation into the rock record?

While each chapter contains its own standalone introduction, the following section is intended to familiarize the reader with some of the general geological, geochemical, and geobiological aspects of methane seeps, as well as to motivate the overarching research questions addressed herein. At the end of this introduction, a brief summary of the principle questions addressed in each chapter is provided.

~~~

Methane seeps, first discovered in the Gulf of Mexico in 1984 (Paull et al. 1984), are globally distributed geologic features in which reduced fluids, including methane, are advected upward from below the seafloor and meet the oxidized bottom waters of Earth's oceans. Methane seeps are located on both passive and active continental margins, with a variety of geologic processes (e.g., subduction, subseafloor salinity-driven density overturning) creating conduits for reduced fluids to move upwards (Judd 2003; Tunnicliffe et al. 2003; German et al. 2011). Worldwide, the most exhaustively studied methane seeps are those located in the Black Sea (Michaelis et al. 2002), the Mediterranean (Aloisi et al. 2002), and "Hydrate Ridge", a north-south promontory at 700-800 meters water depth off the coast of Oregon on the Cascadia margin (Boetius and Suess 2004).

Methane in seeps is generally sourced from decomposition of buried organic matter (originally produced photosynthetically). Methane production can be either thermogenic or biogenic, with the two processes yielding characteristically different $\delta^{13}C_{methane}$ values (Schoell 1980). At Hydrate Ridge, the seep most intensively studied in this thesis, the methane is highly depleted (approximately -65‰), indicating a mostly biogenic origin (Kastner et al. 1998; Suess et al. 1999; Boetius and Suess 2004). Once produced, depending on temperature and pressure conditions as well as the methane concentration itself, methane will either remain in the dissolved phase, become locked in methane hydrates, or bubble out of the seafloor as free methane gas. The dissolved component of subseafloor methane is oxidized predominantly in the subsurface by the microbially-mediated, sulfate-coupled anaerobic oxidation of methane (AOM):

$$CH_4 + SO_4^{=} \rightleftarrows HCO_3^{-} + HS^{-} + H_2O \qquad \text{(Eq. 1)}$$

It is well established that consortia of anaerobic methane oxidizing archaea (ANME) and sulfate-reducing bacteria (SRB) work symbiotically to achieve AOM (Hinrichs et al. 1999; Boetius et al. 2000; Orphan et al. 2001b), although the mechanistic details of the process remain an active

area of investigation (Milucka et al. 2012; McGlynn et al. 2015; Scheller et al. 2016). AOM can also be accomplished with other electron acceptors including iron (Beal et al. 2009), manganese (Beal et al. 2009), nitrite (Ettwig et al. 2010), and nitrate (Haroon et al. 2013), but sulfate-coupled AOM is the dominant process in most seep environments. ANME archaea are subdivided into three clades: ANME-1, ANME-2, and ANME-3 (Orphan et al. 2001a; 2002; Knittel et al. 2005; Niemann et al. 2006). These ANME often live in tight consortia with deltaprotebacterial SRB of the *Desulfosarcina*, *Desulfococcus*, and *Desulfobulbus* genera (Orphan et al. 2002; Knittel et al. 2005; Lösekann et al. 2007; Schreiber et al. 2010; Green-Saxena et al. 2014), although reports also exist of ANME-1 and ANME-2 living as single cells or monospecies aggregates (Orphan et al. 2001b; 2002).

AOM in methane seeps oxidizes ~96% to ~100% of dissolved methane before it can escape to ocean bottom waters (Sommer et al. 2006). In cases where methane concentrations are high enough to nucleate ebullition, more methane escapes and the biological AOM filter is reduced to ~66% to 83% efficiency (Sommer et al. 2006). Nonetheless, methane seeps are estimated to consume ~78 Tg CH_4/year globally (and when including AOM in non-seep continental shelf sediments, estimates suggest global consumption of >350 Tg CH_4/year, implying also a high annual rate of subseafloor methane production; Hinrichs and Boetius 2000; Reeburgh 2007). Estimates suggest Earth's current (including anthropogenic influences) gross production and consumption of methane to be on the order of 500-600 Tg CH_4/year (Kirschke et al. 2013). As such AOM, including AOM in methane seeps, is a significant contributor to global methane cycling, and a small change in cycling efficiency at methane seeps might have a large impact on global budgets. Moreover, the contribution of methane seeps is likely to be higher than estimated, considering that the occurrence of seeps, which has been largely unknown, now appears to be extremely high: a recent study located seeps every 4 km along the continental slope off Nicaragua, Costa Rica, and Panama (Sahling et al. 2008).

Seeps may have played one or more important roles during Earth history. This could include, of course, a similar role as today, with AOM in seeps cycling a large amount of methane. Such a role has been proposed based on evidence of strongly ^{13}C-depleted kerogens from as old as 2.7 Ga (Hinrichs 2002) as well as ^{13}C-depleted carbonate deposits in the Proterozoic and Phanerozoic (Schrag et al. 2013). However, other mechanisms have also been invoked to explain ^{13}C signals in the ancient sedimentary record, including alternative microbial metabolisms and/or secondary alternation of carbonate minerals following burial (Grotzinger et al. 2011; Slotznick and Fischer 2016). Other contributions of subseafloor methane to Earth history have also been proposed. Perhaps the most catastrophic is the hypothesis that rapid and self-reinforcing destabilization of methane hydrates may have led to and/or been a result of dramatic climate warming during the Paleocene-Eocene Thermal Maximum (Katz et al. 1999), and during post-glacial hot house periods following snowball Earth events in the Proterozoic (Jiang et al. 2003).

Besides ANME and SRB, methane seeps host a broad diversity of microorganisms. Studies have most intensively focused on methane seep sediments, finding that microbial assemblages in seep sediments are distinct from other marine sedimentary environments as assessed by 16S rRNA gene surveys (Pop Ristova et al. 2015; Ruff et al. 2015). Seeps are typified from other settings by high relative abundances of *Methanomicrobia* and *Deltaproteobacteria* classes (containing ANME and SRB, respectively), as well as bacterial candidate divisions Hyd24-12 and JS1 (Ruff et al. 2015). Targeted metagenomics have recently shown members of candidate division Hyd24-12 to contain genes consistent with fermentative metabolisms of simple sugars and the potential ability to reduce elemental sulfur to sulfide (Kirkegaard et al. 2016). Using single cell genomics, candidate division JS1 has recently been placed within the candidate phylum *Atribacteria*, and likely represents heterotrophic, strictly anaerobic microorganisms participating in fermentation of organic acids such as propionate and acetate (Nobu et al. 2016). Sulfur-cycling microorganisms, in particular putative sulfide-oxidizing members of the *Epsilonproteobacteria* and *Gammaproteobacteria*, are also common and abundant in seep sediments and are often manifested by

dense orange, pink, and/or white mats covering the seafloor (Nunoura et al. 2012; Niemann et al. 2013; Marlow et al. 2014b; Ruff et al. 2015).

The ANME, SRB, and other microbial species inhabiting methane seeps form the base of a complex food web of organisms which includes fish, bivalves (mussels and clams), tube worms, polychaetes, nematodes, and crabs (Levin 2005). The microbial and megafaunal inhabitants of methane seeps are one of the most recognizable seafloor features of methane seepage (in addition to a carbonate-strewn landscape and, occasionally, methane ebullition), helping researchers identify seep locales during seafloor exploration. Moreover, the distribution of biological organisms often mirrors the delivery of reduced fluids from below, effectively allowing biology to be used as a rough visual "map" of contemporary seepage (Barry et al. 1997; Treude et al. 2003; Orphan et al. 2004; Levin 2005).

This is useful and important because methane seepage is a spatially and temporally heterogeneous process. Lateral and vertical microbiological and geochemical heterogeneity is observed on the scale of meters or less (Sahling et al. 2002; Levin et al. 2003; Treude et al. 2003; Orphan et al. 2004; Pop Ristova et al. 2015). In general, microbial mats are observed to dominate high-flux areas, often ringed in a "bulls-eye" manner by bivalves on the periphery of seepage (Barry et al. 1997; Orphan et al. 2004). If methane ebullition is present, it is usually strong enough to disrupt any colonization by seep organisms in its direct (cm- to m-scale) proximity.

Temporal variation in seepage is one of the primary drivers of spatial heterogeneity of seep-dependent taxa and geochemistry. Seepage can vary on timescales from days (tidal forcing of the hydraulic head, fault blockage by hydrates or carbonates; Torres et al. 2002; Tryon et al. 2002) to hundreds of years (subduction thrust earthquake cycle in the Cascadia Margin; Tryon et al. 2002) to tens or hundreds of thousands of years (glacial/interglacial sea level forcing of the hydraulic head; Teichert et al. 2003; Watanabe et al. 2008). These timescales are reflected in age measurements of carbonate precipitates at methane seeps, which form inorganically (i.e., authigenically) within the sediment column as a result of the production of two units of alkalinity

(Alk) per one unit of dissolved inorganic carbon (DIC) from sulfate-coupled AOM (Eq. 1; Berner 1980; Ritger et al. 1987). Histograms of carbonate age measured by U-Th frequently indicate discontinuous precipitation on glacial/interglacial timescales (Aharon et al. 1997; Teichert et al. 2003; Kutterolf et al. 2008; Watanabe et al. 2008; Kiel 2009; Liebetrau et al. 2010; Feng et al. 2010). This presumably tracks discontinuous magnitudes of seepage flux and AOM driven by changes in methane hydrate stability as a function of sea level controlling the hydraulic head overlying seep systems. Such timescales would be compatible with other measurements which indicate approximately continuous precipitation, but not on the scale of greater than thousands of years (Naehr et al. 2000; Bayon et al. 2009).

Previous sampling efforts have shown seep carbonates to consist of a variety of morphologies including aragonite, calcite, dolomite, and mixtures thereof (Ritger et al. 1987; Kulm and Suess 1990; Bohrmann et al. 1998; Naehr et al. 2000; Greinert et al. 2001). The factors most often invoked to explain the precipitation of specific carbonate mineralogies have included Alk concentration (Ritger et al. 1987; Greinert et al. 2001; Luff and Wallmann 2003), Ca^{2+} and Mg^{2+} ion concentration (Ritger et al. 1987; Greinert et al. 2001), temperature (Naehr et al. 2000; Greinert et al. 2001), degree of calcium carbonate supersaturation (Naehr et al. 2000), and, more than any other factor, pore water sulfate concentration (Ritger et al. 1987; Burton 1993; Greinert et al. 2001; Peckmann et al. 2001; Aloisi et al. 2002; Reitner et al. 2005a). Based on empirical observations of the distribution of different calcium carbonate morphologies, thermodynamic and kinetic modeling, and carbon and oxygen isotope ratios of seep carbonates, the working hypothesis is that aragonites, which are more soluble than calcite and high-Mg calcite phases, precipitate at or near the sediment-water interface in a zone where alkalinity and carbonate supersaturation is quite high, but sulfate concentrations from the overlying bottom water remain high enough to inhibit calcite precipitation (Ritger et al. 1987; Aloisi et al. 2002; Gieskes et al. 2005). Calcites, accordingly, are hypothesized to form slightly deeper in the sediment column, perhaps near the zone of maximum AOM, where alkalinity is high and sulfate is depleted to zero

concentration. Dolomites, based primarily on carbon and oxygen isotope evidence, are predicted to form well below AOM zones, in horizons of microbial methanogenesis (Greinert et al. 2001; Gieskes et al. 2005).

At many methane seeps, including Hydrate Ridge, carbonate blocks and pavements are ubiquitous features of the seafloor landscape, covering hundreds of square meters on the seafloor (Kulm and Suess 1990; Boetius and Suess 2004) and extending vertically into the water column in special cases of euxinia such as the Black Sea (Michaelis et al. 2002). As such, seep carbonates provide an important hard habitat substrate for biological colonization. In addition to their lateral extent, subseafloor acoustic data shows that carbonates also extend vertically to meters or tens of meters below the sediment-water interface (Klaucke et al. 2008; 2012). Carbonates, therefore, represent a volumetrically significant (perhaps dominant) habitat substrate in methane seep benthic and subseafloor ecosystems (Marlow et al. 2014a). Despite this, investigations of carbonate-associated microbial and megafaunal ecology have been limited, in part due to unrecognized importance of seep carbonates as distinct habitats from seep sediments and in part due to their difficult sampling nature on the seafloor. With the exception of one early report (Jensen et al. 1992), studies of seep carbonate-associated megafaunal diversity have only recently begun to enter the literature (Ritt et al. 2010; 2011; Grupe et al. 2015; Levin et al. 2015).

The first detailed descriptions of microorganisms specifically associated with seep carbonates were a series of reports of microbial mats covering large carbonate chimneys extending meters above the seafloor in the anoxic bottom waters of the Black Sea (Pimenov et al. 1997; Thiel et al. 2001; Peckmann et al. 2001). These included microscopic observations of microbial filaments and lipid profile characterizations, and *in vitro* experiments showed the organisms to be capable of metabolic activity (Michaelis et al. 2002). Subsequent lipid and 16S rRNA gene analyses further characterized the chemotaxonomic and phylogenetic diversity of microorganisms inhabiting Black Sea seep carbonates (Tourova et al. 2002; Blumenberg et al. 2004; Reitner et al. 2005b). Around the same time, lipid and 16S rRNA gene surveys were

reported from several benthic seep carbonates in the Mediterranean Sea, unambiguously demonstrating the presence of ANME and SRB as well as other phylogeneticaly diverse microorganisms, including archaeal Marine Group I and bacterial alpha-, beta-, and gamma-proteobacteria, among others (Pancost et al. 2001; Aloisi et al. 2002; Heijs et al. 2006). Meanwhile, parallel analysis of four carbonate samples with five sediment samples from ~2,000 meters deep in the Black Sea yielded a higher concentration of microbial biomarkers and a distinct community of Archaea in the carbonates as compared to the sediments (Stadnitskaia et al. 2005). Not until 2014 was as systematic survey undertaken to compare microbial communities inhabiting seep carbonates vs seep sediments. Though limited in sample number, investigation at Hydrate Ridge revealed bacterial and archaeal 16S rRNA gene profiles to be differentiated by habitat substrate and seep activity, respectively (Marlow et al. 2014b). In the same year, seep carbonates were demonstrated to host metabolically active endolithic anaerobic methanotrophs, greatly expanding the potential contribution of AOM to subseafloor methane cycling (Marlow et al. 2014a).

The application of microbial biomarker analyses has not been limited to modern, actively venting methane seeps. Extraction and characterization of lipid biomarkers has been used in conjunction with isotopic and megafaunal fossil evidence to link geologic outcrops to historic methane seepage from as young as the Pleistocene to as old as the Carboniferous (Peckmann et al. 1999; Thiel et al. 1999; Peckmann et al. 2002; Goedert et al. 2003; Birgel et al. 2006b; a; Birgel et al. 2008b; a; Kiel et al. 2013; Natalicchio et al. 2015; Little et al. 2015). Besides just ancient seep identification, lipids have been used at sites wordwide to infer paleo-seepage flux (Stadnitskaia et al. 2008; Leefmann et al. 2008; Gontharet et al. 2009; Peckmann et al. 2009; Birgel et al. 2011; Hagemann et al. 2012) or ancient ANME-1 vs ANME-2 ratios (Niemann and Elvert 2008; Peckmann et al. 2009; Birgel et al. 2011; Natalicchio et al. 2015).

Paleo-seep studies rely on assumptions regarding both the fidelity with which biomarkers record methane seep microbial assemblages (chemotaxonomy) and the longevity of biomarkers

over time. Chemotaxomic understanding within methane seeps systems has grown as a result of extensive study and methodological development over the last ~15 years (Hinrichs et al. 2000; Rütters et al. 2001; Elvert et al. 2003; Sturt et al. 2004; Blumenberg et al. 2004; Rossel et al. 2008; Niemann and Elvert 2008; Schubotz et al. 2011; Yoshinaga et al. 2011; 2015), but the longevity (i.e., degradation rate) of biomarkers remains poorly constrained. Historically, intact polar lipids (IPLs) were assumed to be excellent recorders of live biomass due to rapid degradation after cell death of the bond between polar head group and glycerol backbone (White et al. 1979; 1997; Zink et al. 2003; Sturt et al. 2004). The core lipid, that which remains after loss of the polar head group, is assumed to be stable over geologic timescales and indeed core lipids have been recovered from seep carbonates in the sedimentary record as old as 300 million years (Birgel et al. 2008b). Recent evidence, however, has indicated IPLs may have high enough longevity to record more than just living microbial communities. In anoxic laboratory microcosm experiments of marine sediment, ester-based IPLs (in this case artificially spiked in from dead eukaryotic *Saccharomyces cerevisiae* cells, but also the dominant bond type found in bacterial lipids) degraded to <10% of original concentration within 100 days, but archaeal ether-based IPLs demonstrated no measurable degradation in the same time frame (Logemann et al. 2011). A more sensitive ^{14}C-labeling experiment, run for 300 days but only analyzing archaeal ether-linked IPLs, indicated such slow IPL degradation rates in anoxic marine sediment microcosms that the half-life of archaeal IPLs in subseafloor environments was extrapolated to be 10^3-10^5 years (Xie et al. 2013).

This begs the question posed on page 1 of this thesis: how do microorganisms respond to changes in seepage flux, and how is the microbial response recorded in carbonate-associated biomarkers?

~~~
~~~

This thesis concerns the ecology of microorganism associated with seep carbonates from basic questions of inter-habitat diversity to novel, *in situ*, multi-biomarker, time-resolved seafloor experiments. Some basic questions addressed are:

Chapter One

- Does microbial diversity differ between subseafloor protolithic carbonate "nodules" and their adjacent sediments?
- To what extent do subseafloor geochemical parameters inform the capture of sediment-hosted microbial assemblages into solid habitat substrates?

Chapter Two

- Do fully lithified seep carbonates host distinct microorganisms from seep sediments, nodules, and bottom waters?
- How sensitive are carbonate-associated microbial assemblages to seepage activation or quiescence on 13-month timescales?
- Which microorganisms colonize sterile hard substrates placed at methane seeps, and is colonization substrate- or seepage-dependent?

Chapter Three

- In parallel analysis of DNA and IPL biomarkers, are similar or different sensitivities observed to seepage flux (and changes in seepage flux)?
- Are microbial biomarkers preferentially associated with particular morphologies of calcium carbonate?
- Given, for the first time, observed microbial responses to 13 months of *in situ* imposed seep quiescence, can future biomarker response to continued seep quiescence be quantitatively estimated?
- Are some microbial species' associated biomarkers particularly suited to be applied as proxies for "active" or "low activity" seep conditions recorded in ancient seep carbonates?

These three chapters represent the primary contribution of this thesis to the scientific community's understanding of seep carbonate-associated microbial ecology (Chapters One and Two were published in 2015, while Chapter Three is in manuscript form). Chapter Four concerns methods development for next-generation sequencing of the 16S rRNA gene, the most commonly employed microbial diversity assay in the literature and in this thesis, including extensive post-processing tests to ensure robust data interpretation. An appendix describes microbiological and geochemical results from a high-pressure incubation of methane seep sediments recovered from the Joetsu Knoll, Japan.

~~~
~~~

REFERENCES

Aharon, P., H. P. Schwarcz, and H. H. Roberts. 1997. Radiometric dating of submarine hydrocarbon seeps in the Gulf of Mexico. Geological Society of America Bulletin **109**: 568–579.

Aloisi, G., I. Bouloubassi, S. K. Heijs, R. D. Pancost, C. Pierre, J. S. Sinninghe Damsté, J. C. Gottschal, L. J. Forney, and J.-M. Rouchy. 2002. CH4-consuming microorganisms and the formation of carbonate crusts at cold seeps. Earth and Planetary Science Letters **203**: 195–203.

Barry, J. P., R. E. Kochevar, and C. H. Baxter. 1997. The influence of pore-water chemistry and physiology on the distribution of vesicomyid clams at cold seeps in Monterey Bay: Implications for patterns of chemosynthetic community organization. Limnology and Oceanography **42**: 318–328.

Bayon, G., G. M. Henderson, and M. Bohn. 2009. U–Th stratigraphy of a cold seep carbonate crust. Chemical Geology **260**: 47–56.

Beal, E., C. House, and V. Orphan. 2009. Manganese-and iron-dependent marine methane oxidation. Science **325**: 184–187.

Berner, R. A. 1980. Earth Diagenesis, Princeton University Press.

Birgel, D., D. Feng, H. H. Roberts, and J. Peckmann. 2011. Changing redox conditions at cold seeps as revealed by authigenic carbonates from Alaminos Canyon, northern Gulf of Mexico. Chemical Geology **285**: 82–96.

Birgel, D., J. Peckmann, S. Klautzsch, V. Thiel, and J. Reitner. 2006a. Anaerobic and Aerobic Oxidation of Methane at Late Cretaceous Seeps in the Western Interior Seaway, USA. Geomicrobiology Journal **23**: 565-577.

Birgel, D., M. Elvert, X. Han, and J. Peckmann. 2008a. ^{13}C-depleted biphytanic diacids as tracers of past anaerobic oxidation of methane. Organic geochemistry **39**: 152–156.

Birgel, D., T. Himmler, A. Freiwald, and J. Peckmann. 2008b. A new constraint on the antiquity of anaerobic oxidation of methane: Late Pennsylvanian seep limestones from southern Namibia. Geology **36**: 543–546.

Birgel, D., V. Thiel, Hinrichs, M. Elvert, K. A. Campbell, J. Reitner, J. D. Farmer, and J. Peckmann. 2006b. Lipid biomarker patterns of methane-seep microbialites from the Mesozoic convergent margin of California. Organic geochemistry **37**: 1289–1302.

Blumenberg, M., R. Seifert, J. Reitner, T. Pape, and W. Michaelis. 2004. Membrane lipid patterns typify distinct anaerobic methanotrophic consortia. Proceedings of the National Academy of Sciences **101**: 11111–11116.

Boetius, A., and E. Suess. 2004. Hydrate Ridge: a natural laboratory for the study of microbial life fueled by methane from near-surface gas hydrates. Chemical Geology **205**: 291–310.

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

Bohrmann, G., J. Greinert, E. Suess, and M. Torres. 1998. Authigenic carbonates from the Cascadia subduction zone and their relation to gas hydrate stability. Geology **26**: 647.

Burton, E. A. 1993. Controls on marine carbonate cement mineralogy: review and reassessment. Chemical Geology **105**: 163–179.

Elvert, M., A. Boetius, K. Knittel, and B. B. Jørgensen. 2003. Characterization of Specific Membrane Fatty Acids as Chemotaxonomic Markers for Sulfate-Reducing Bacteria Involved in Anaerobic Oxidation of Methane. Geomicrobiology Journal **20**: 403–419.

Ettwig, K., M. Butler, D. Le Paslier, and E. Pelletier. 2010. Nitrite-driven anaerobic methane oxidation by oxygenic bacteria. Nature **464**: 543-548

Feng, D., H. H. Roberts, H. Cheng, J. Peckmann, G. Bohrmann, R. Lawrence Edwards, and D. Chen. 2010. U/Th dating of cold-seep carbonates: An initial comparison. Deep Sea Research Part II: Topical Studies in Oceanography **57**: 2055–2060.

German, C. R., E. Ramirez-Llodra, M. C. Baker, P. A. Tyler, and A. T. C. S. S. Committee. 2011. Deep-Water Chemosynthetic Ecosystem Research during the Census of Marine Life Decade and

Beyond: A Proposed Deep-Ocean Road Map. PLoS ONE **6**: e23259.

Gieskes, J., C. Mahn, S. Day, J. Martin, J. Greinert, T. Rathburn, and B. McAdoo. 2005. A study of the chemistry of pore fluids and authigenic carbonates in methane seep environments: Kodiak Trench, Hydrate Ridge, Monterey Bay, and Eel River Basin. Chemical Geology **220**: 329–345.

Goedert, J. L., V. Thiel, O. Schmale, W. W. Rau, W. Michaelis, and J. Peckmann. 2003. The Late Eocene "Whiskey Creek" methane-seep deposit (western Washington State). Facies **48**: 223–239.

Gontharet, S., A. Stadnitskaia, I. Bouloubassi, C. Pierre, and J. S. Damsté. 2009. Palaeo methane-seepage history traced by biomarker patterns in a carbonate crust, Nile deep-sea fan (Eastern Mediterranean Sea). Marine Geology **261**: 105–113.

Green-Saxena, A., A. E. Dekas, N. F. Dalleska, and V. J. Orphan. 2014. Nitrate-based niche differentiation by distinct sulfate-reducing bacteria involved in the anaerobic oxidation of methane. The ISME Journal **8**: 150–163.

Greinert, J., G. Bohrmann, and E. Suess. 2001. Gas Hydrate-Associated Carbonates and Methane-Venting at Hydrate Ridge: Classification, Distribution, and Origin of Authigenic Lithologies, p. 99–113. *In* C.K. Paull and W.P. Dillon [eds.], Natural Gas Hydrates. American Geophysical Union.

Grotzinger, J. P., D. A. Fike, and W. W. Fischer. 2011. Enigmatic origin of the largest-known carbon isotope excursion in Earth's history. Nature Geoscience **4**: 285–292.

Grupe, B. M., M. L. Krach, A. L. Pasulka, and J. M. Maloney. 2015. Methane seep ecosystem functions and services from a recently discovered southern California seep. Marine Ecology **36**: 91-108

Hagemann, A., T. Leefmann, J. Peckmann, V.-E. Hoffmann, and V. Thiel. 2012. Biomarkers from individual carbonate phases of an Oligocene cold-seep deposit, Washington State, USA. Lethaia **46**: 7–18.

Haroon, M. F., S. Hu, Y. Shi, M. Imelfort, J. Keller, P. Hugenholtz, Z. Yuan, and G. W. Tyson. 2013. Anaerobic oxidation of methane coupled to nitrate reduction in a novel archaeal lineage. Nature, doi:10.1038/nature12375

Heijs, S. K., G. Aloisi, I. Bouloubassi, R. D. Pancost, C. Pierre, J. S. Sinninghe Damsté, J. C. Gottschal, J. D. Elsas, and L. J. Forney. 2006. Microbial Community Structure in Three Deep-Sea Carbonate Crusts. Microbial Ecology **52**: 451–462.

Hinrichs. 2002. Microbial fixation of methane carbon at 2.7 Ga: Was an anaerobic mechanism possible? Geochemistry Geophysics Geosystems **3**: 1042.

Hinrichs, and A. Boetius. 2000. The Anaerobic Oxidation of Methane, p. 457–477. *In* G. Wefer, D. Billett, D. Hebbeln, B.B. Jørgensen, M. Schluter, and T. Van Weering [eds.], Ocean Margin Systems. Springer Netherlands.

Hinrichs, J. M. Hayes, S. P. Sylva, P. G. Brewer, and E. F. DeLong. 1999. Methane-consuming archaebacteria in marine sediments. Nature **398**: 802–805.

Hinrichs, R. E. Summons, V. Orphan, S. P. Sylva, and J. M. Hayes. 2000. Molecular and isotopic analysis of anaerobic methane-oxidizing communities in marine sediments. Organic Geochemistry **31**: 1685–1701.

Jensen, P., I. Aagaard, R. A. Burke Jr, P. R. Dando, N. O. Jørgensen, A. Kuijpers, T. Laier, S. OHara, and R. Schmaljohann. 1992. "Bubbling reefs" in the Kattegat: submarine landscapes of carbonate-cemented rocks support a diverse ecosystem at methane seeps. Marine Ecology Progress Series **83**: 103–112.

Jiang, G., M. J. Kennedy, and N. Christie-Blick. 2003. Stable isotopic evidence for methane seeps in Neoproterozoic postglacial cap carbonates. Nature **426**: 822–826.

Judd, A. 2003. The global importance and context of methane escape from the seabed. Geo-Marine Letters **23**: 147-154

Kastner, M., K. A. Kvenvolden, and T. D. Lorenson. 1998. Chemistry, isotopic composition, and origin of a methane-hydrogen sulfide hydrate at the Cascadia subduction zone. Earth and Planetary Science Letters **156**: 173–183.

Katz, M. E., D. K. Pak, G. R. Dickens, and K. G. Miller. 1999. The Source and Fate of Massive Carbon Input During the Latest Paleocene Thermal Maximum. Science **286**: 1531–1533.

Kiel, S. 2009. Global hydrocarbon seep-carbonate precipitation correlates with deep-water temperatures and eustatic sea-level fluctuations since the Late Jurassic. Terra Nova **21**: 279–284.

Kiel, S., D. Birgel, K. A. Campbell, J. S. Crampton, P. Schiøler, and J. Peckmann. 2013. Cretaceous methane-seep deposits from New Zealand and their fauna. Palaeogeography, Palaeoclimatology, Palaeoecology **390**: 17–34.

Kirkegaard, R. H., M. S. Dueholm, S. J. McIlroy, M. Nierychlo, S. M. Karst, M. Albertsen, and P. H. Nielsen. 2016. Genomic insights into members of the candidate phylum Hyd24-12 common in mesophilic anaerobic digesters. The ISME Journal, doi:10.1038/ismej.2016.43

Kirschke, S., P. Bousquet, P. Ciais, M. Saunois, J. G. Canadell, E. J. Dlugokencky, P. Bergamaschi, D. Bergmann, D. R. Blake, L. Bruhwiler, P. Cameron-Smith, S. Castaldi, F. Chevallier, L. Feng, A. Fraser, M. Heimann, E. L. Hodson, S. Houweling, B. Josse, P. J. Fraser, P. B. Krummel, J.-F. Lamarque, R. L. Langenfelds, C. Le Quéré, V. Naik, S. O'Doherty, P. I. Palmer, I. Pison, D. Plummer, B. Poulter, R. G. Prinn, M. Rigby, B. Ringeval, M. Santini, M. Schmidt, D. T. Shindell, I. J. Simpson, R. Spahni, L. P. Steele, S. A. Strode, K. Sudo, S. Szopa, G. R. van der Werf, A. Voulgarakis, M. van Weele, R. F. Weiss, J. E. Williams, and G. Zeng. 2013. Three decades of global methane sources and sinks. Nature Geoscience **6**: 813–823.

Klaucke, I., D. G. Masson, C. J. Petersen, W. Weinrebe, and C. R. Ranero. 2008. Multifrequency geoacoustic imaging of fluid escape structures offshore Costa Rica: Implications for the quantification of seep processes. Geochemistry Geophysics Geosystems **9**, doi:10.1029/2007GC001708

Klaucke, I., W. Weinrebe, P. Linke, D. Kläschen, and J. Bialas. 2012. Sidescan sonar imagery of widespread fossil and active cold seeps along the central Chilean continental margin. Geo-Marine Letters **32**: 489–499.

Knittel, K., T. Lösekann, A. Boetius, R. Kort, and R. Amann. 2005. Diversity and Distribution of Methanotrophic Archaea at Cold Seeps. Applied and Environmental Microbiology **71**: 467–479.

Kulm, L. D., and E. Suess. 1990. Relationship between carbonate deposits and fluid venting: Oregon Accretionary Prism. Journal of Geophysical Research: Solid Earth **95**: 8899–8915.

Kutterolf, S., V. Liebetrau, T. Mörz, A. Freundt, T. Hammerich, and D. Garbe-Schönberg. 2008. Lifetime and cyclicity of fluid venting at forearc mound structures determined by tephrostratigraphy and radiometric dating of authigenic carbonates. Geology **36**: 707–710.

Leefmann, T., J. Bauermeister, A. Kronz, V. Liebetrau, J. Reitner, and V. Thiel. 2008. Miniaturized biosignature analysis reveals implications for the formation of cold seep carbonates at Hydrate Ridge (off Oregon, USA). Biogeosciences **5**: 731–738.

Levin, L. A. 2005. Ecology of cold seep sediments: Interactions of fauna with flow, chemistry, and microbes. Oceanography and Marine Biology: An Annual Review **43**: 1–46.

Levin, L. A., G. F. Mendoza, B. M. Grupe, J. P. Gonzalez, B. Jellison, G. W. Rouse, A. R. Thurber, and A. Waren. 2015. Biodiversity on the Rocks: Macrofauna Inhabiting Authigenic Carbonate at Costa Rica Methane Seeps. PLoS ONE 1–31.

Levin, L. A., W. Ziebis, G. F. Mendoza, V. Growney, M. D. Tryon, K. M. Brown, C. Mahn, J. M. Gieskes, and A. E. Rathburn. 2003. Spatial heterogeneity of macrofauna at northern California methane seeps: influence of sulfide concentration and fluid flow. Marine Ecology Progress Series **265**: 123–139.

Liebetrau, V., A. Eisenhauer, and P. Linke. 2010. Cold seep carbonates and associated cold-water corals at the Hikurangi Margin, New Zealand: New insights into fluid pathways, growth structures and geochronology. Marine Geology **272**: 307–318.

Little, C. T. S., D. Birgel, A. J. Boyce, J. A. Crame, J. E. Francis, S. Kiel, J. Peckmann, D. Pirrie, G. K. Rollinson, and J. D. Witts. 2015. Late Cretaceous (Maastrichtian) shallow water hydrocarbon seeps from Snow Hill and Seymour Islands, James Ross Basin, Antarctica. Palaeogeography, Palaeoclimatology, Palaeoecology **418**: 213–228.

Logemann, J., J. Graue, J. Köster, B. Engelen, J. Rullkötter, and H. Cypionka. 2011. A laboratory experiment of intact polar lipid degradation in sandy sediments. Biogeosciences **8**: 2547–2560.

Lösekann, T., K. Knittel, T. Nadalig, B. Fuchs, H. Niemann, A. Boetius, and R. Amann. 2007. Diversity and Abundance of Aerobic and Anaerobic Methane Oxidizers at the Haakon Mosby

Mud Volcano, Barents Sea. Applied and Environmental Microbiology **73**: 3348-3362.

Luff, R., and K. Wallmann. 2003. Fluid flow, methane fluxes, carbonate precipitation and biogeochemical turnover in gas hydrate-bearing sediments at Hydrate Ridge, Cascadia Margin: numerical modeling and mass balances. Geochimica et Cosmochimica Acta **67**: 3403–3421.

Marlow, J. J., J. A. Steele, W. Ziebis, A. R. Thurber, L. A. Levin, and V. J. Orphan. 2014a. Carbonate-hosted methanotrophy represents an unrecognized methane sink in the deep sea. Nature Communications **5**: 1–12.

Marlow, J., J. A. Steele, D. Case, S. A. Connon, L. A. Levin, and V. J. Orphan. 2014b. Microbial abundance and diversity patterns associated with sediments and carbonates from the methane seep environments of Hydrate Ridge, OR. Frontiers in Marine Science **1**: 1–16.

McGlynn, S. E., G. L. Chadwick, C. P. Kempes, and V. J. Orphan. 2015. Single cell activity reveals direct electron transfer in methanotrophic consortia. Nature 1–14.

Michaelis, W., R. Seifert, K. Nauhaus, T. Treude, V. Thiel, M. Blumenberg, K. Knittel, A. Gieseke, K. Peterknecht, T. Pape, A. Boetius, R. Amann, B. B. Jørgensen, F. Widdel, J. Peckmann, N. V. Pimenov, and M. B. Gulin. 2002. Microbial Reefs in the Black Sea Fueled by Anaerobic Oxidation of Methane. Science **297**: 1013–1015.

Milucka, J., T. G. Ferdelman, L. Polerecky, D. Franzke, G. Wegener, M. Schmid, I. Lieberwirth, M. Wagner, F. Widdel, and M. M. M. Kuypers. 2012. Zero-valent sulphur is a key intermediate in marine methane oxidation. Nature 1–10.

Naehr, T. H., N. M. Rodriguez, G. Bohrmann, C. K. Paull, and R. Botz. 2000. Methane-derived authigenic carbonates associated with gas hydrate decomposition and fluid venting above the Blake Ridge Diapir. Proceedings of the Proceedings of the Ocean Drilling Program **164**: 285-300.

Natalicchio, M., J. Peckmann, D. Birgel, and S. Kiel. 2015. Seep deposits from northern Istria, Croatia: a first glimpse into the Eocene seep fauna of the Tethys region. Geol. Mag. **152**: 444–459.

Niemann, H., and M. Elvert. 2008. Diagnostic lipid biomarker and stable carbon isotope signatures of microbial communities mediating the anaerobic oxidation of methane with sulphate. Organic geochemistry **39**: 1668–1677.

Niemann, H., P. Linke, K. Knittel, E. MacPherson, A. Boetius, W. Brückmann, G. Larvik, K. Wallmann, U. Schacht, E. Omoregie, D. Hilton, K. Brown, and G. Rehder. 2013. Methane-Carbon Flow into the Benthic Food Web at Cold Seeps – A Case Study from the Costa Rica Subduction Zone. PLoS ONE **8**: e74894.

Niemann, H., T. Lösekann, D. de Beer, M. Elvert, T. Nadalig, K. Knittel, R. Amann, E. J. Sauter, M. Schlüter, M. Klages, J. P. Foucher, and A. Boetius. 2006. Novel microbial communities of the Haakon Mosby mud volcano and their role as a methane sink. Nature **443**: 854–858.

Nobu, M. K., J. A. Dodsworth, S. K. Murugapiran, C. Rinke, E. A. Gies, G. Webster, P. Schwientek, P. Kille, R. J. Parkes, H. Sass, B. B. Jørgensen, A. J. Weightman, W.-T. Liu, S. J. Hallam, G. Tsiamis, T. Woyke, and B. P. Hedlund. 2016. Phylogeny and physiology of candidate phylum 'Atribacteria' (OP9/JS1) inferred from cultivation-independent genomics. The ISME Journal **10**: 273–286.

Nunoura, T., Y. Takaki, H. Kazama, M. Hirai, J. Ashi, H. Imachi, and K. Takai. 2012. Microbial Diversity in Deep-sea Methane Seep Sediments Presented by SSU rRNA Gene Tag Sequencing. Microbes and Environments **27**: 382–390.

Orphan, V. J., Hinrichs, W. Ussler, C. K. Paull, L. T. Taylor, S. P. Sylva, J. M. Hayes, and E. F. Delong. 2001a. Comparative Analysis of Methane-Oxidizing Archaea and Sulfate-Reducing Bacteria in Anoxic Marine Sediments. Applied and Environmental Microbiology **67**: 1922–1934.

Orphan, V., C. House, and Hinrichs. 2002. Multiple archaeal groups mediate methane oxidation in anoxic cold seep sediments. Proceedings of the Proceedings of the National Academy of Sciences **99**: 7663–7668.

Orphan, V., C. House, Hinrichs, K. McKeegan, and E. DeLong. 2001b. Methane-Consuming Archaea Revealed by Directly Coupled Isotopic and Phylogenetic Analysis. Science **293**: 484–487.

Orphan, V., W. Ussler, T. H. Naehr, C. H. House, Hinrichs, and C. K. Paull. 2004. Geological,

geochemical, and microbiological heterogeneity of the seafloor around methane vents in the Eel River Basin, offshore California. Chemical Geology **205**: 265–289.

Pancost, R. D., I. Bouloubassi, G. Aloisi, J. S. Sinninghe Damsté, and T. M. S. Scientific Party. 2001. Three series of non-isoprenoidal dialkyl glycerol diethers in cold-seep carbonate crusts. Organic geochemistry **32**: 695–707.

Paull, C. K., B. Hecker, R. Commeau, R. P. Freeman-Lynde, C. Neumann, W. P. Corso, S. Golubic, J. E. Hook, E. Sikes, and J. Curray. 1984. Biological Communities at the Florida Escarpment Resemble Hydrothermal Vent Taxa. Science **226**: 965–967.

Peckmann, J., A. Reimer, C. Luth, C. Luth, B. T. Hansen, C. Heinicke, J. Hoefs, and J. Reitner. 2001. Methane-derived carbonates and authigenic pyrite from the northwestern Black Sea. Marine Geology **177**: 129–150.

Peckmann, J., D. Birgel, and S. Kiel. 2009. Molecular fossils reveal fluid composition and flow intensity at a Cretaceous seep. Geology **37**: 847–850.

Peckmann, J., J. L. Goedert, V. Thiel, W. Michaelis, and J. Reitner. 2002. A comprehensive approach to the study of methane-seep deposits from the Lincoln Creek Formation, western Washington State, USA. Sedimentology **49**: 855–873.

Peckmann, J., V. Thiel, W. Michaelis, P. Clari, C. Gaillard, L. Martire, and J. Reitner. 1999. Cold seep deposits of Beauvoisin (Oxfordian; southeastern France) and Marmorito (Miocene; northern Italy): microbially induced authigenic carbonates. Int J Earth Sci (Geol Rundsch) **88**: 60–75.

Pimenov, N. V., I. I. Rusanov, M. N. Poglazova, L. L. Mityushina, D. Y. Sorokin, V. N. Khmelenina, and Y. A. Trotsenko. 1997. Bacterial mats on coral-like structures at methane seeps in the Black Sea. Microbiology **66**: 354-360.

Pop Ristova, P., F. Wenzhöfer, A. Ramette, J. Felden, and A. Boetius. 2015. Spatial scales of bacterial community diversity at cold seeps (Eastern Mediterranean Sea). The ISME Journal **9**: 1306–1318.

Reeburgh, W. S. 2007. Oceanic Methane Biogeochemistry. Chemical Reviews **107**: 486–513.

Reitner, J., J. Peckmann, A. Reimer, G. Schumann, and V. Thiel. 2005a. Methane-derived carbonate build-ups and associated microbial communities at cold seeps on the lower Crimean shelf (Black Sea). Facies **51**: 66–79.

Reitner, J., J. Peckmann, M. Blumenberg, W. Michaelis, A. Reimer, and V. Thiel. 2005b. Concretionary methane-seep carbonates and associated microbial communities in Black Sea sediments. Palaeogeography, Palaeoclimatology, Palaeoecology **227**: 18–30.

Ritger, S., B. Carson, and E. Suess. 1987. Methane-derived authigenic carbonates formed by subduction-induced pore-water expulsion along the Oregon/Washington margin. Geological Society of America Bulletin **98**: 147–156.

Ritt, B., C. Pierre, O. Gauthier, F. Wenzhöfer, A. Boetius, and J. Sarrazin. 2011. Diversity and distribution of cold-seep fauna associated with different geological and environmental settings at mud volcanoes and pockmarks of the Nile Deep-Sea Fan. Mar Biol **158**: 1187–1210.

Ritt, B., J. Sarrazin, J.-C. Caprais, P. Noël, O. Gauthier, C. Pierre, P. Henry, and D. Desbruyères. 2010. First insights into the structure and environmental setting of cold-seep communities in the Marmara Sea. Deep Sea Research Part I: Oceanographic Research Papers **57**: 1120–1136.

Rossel, P. E., J. S. Lipp, H. F. Fredricks, J. Arnds, A. Boetius, M. Elvert, and Hinrichs. 2008. Intact polar lipids of anaerobic methanotrophic archaea and associated bacteria. Organic geochemistry **39**: 992–999.

Ruff, S. E., J. F. Biddle, A. P. Teske, K. Knittel, A. Boetius, and A. Ramette. 2015. Global dispersion and local diversification of the methane seep microbiome. Proceedings of the National Academy of Sciences 1–6.

Rütters, H., H. Sass, H. Cypionka, and J. Rullkötter. 2001. Monoalkylether phospholipids in the sulfate-reducing bacteria Desulfosarcina variabilis and Desulforhabdus amnigenus. Archives of microbiology **176**: 435–442.

Sahling, H., D. G. Masson, C. R. Ranero, V. Hühnerbach, W. Weinrebe, I. Klaucke, D. Bürk, W. Brückmann, and E. Suess. 2008. Fluid seepage at the continental margin offshore Costa Rica and southern Nicaragua. Geochemistry Geophysics Geosystems **9**: 1–22.

Sahling, H., D. Rickert, R. W. Lee, P. Linke, and E. Suess. 2002. Macrofaunal community structure and sulfide flux at gas hydrate deposits from the Cascadia convergent margin, NE Pacific. Marine Ecology Progress Series **231**: 121–138.

Scheller, S., H. Yu, G. L. Chadwick, S. E. McGlynn, and V. J. Orphan. 2016. Artificial electron acceptors decouple archaeal methane oxidation from sulfate reduction. Science **351**: 703–707.

Schoell, M. 1980. The hydrogen and carbon isotopic composition of methane from natural gases of various origins. Geochimica et Cosmochimica Acta **44**: 649–661.

Schrag, D. P., J. A. Higgins, F. A. Macdonald, and D. T. Johnston. 2013. Authigenic Carbonate and the History of the Global Carbon Cycle. Science **339**: 540–543.

Schreiber, L., T. Holler, K. Knittel, A. Meyerdierks, and R. Amann. 2010. Identification of the dominant sulfate-reducing bacterial partner of anaerobic methanotrophs of the ANME-2 clade. Environmental microbiology **12**: 2327–2340.

Schubotz, F., J. S. Lipp, M. Elvert, and Hinrichs. 2011. Stable carbon isotopic compositions of intact polar lipids reveal complex carbon flow patterns among hydrocarbon degrading microbial communities at the Chapopote asphalt volcano. Geochimica et Cosmochimica Acta **75**: 4399–4415.

Slotznick, S.P., and W.W. Fischer. 2016. Examining Archean methanotrophy. Earth and Planetary Science Letters **344**: 52-59. doi:10.1016/j.epsl.2016.02.013.

Sommer, S., O. Pfannkuche, P. Linke, R. Luff, J. Greinert, M. Drews, S. Gubsch, M. Pieper, M. Poser, and T. Viergutz. 2006. Efficiency of the benthic filter: Biological control of the emission of dissolved methane from sediments containing shallow gas hydrates at Hydrate Ridge. Global Biogeochemical Cycles **20**: 1-14.

Stadnitskaia, A., D. Nadezhkin, B. Abbas, V. Blinova, M. K. Ivanov, and J. S. Sinninghe Damsté. 2008. Carbonate formation by anaerobic oxidation of methane: Evidence from lipid biomarker and fossil 16S rDNA. Geochimica et Cosmochimica Acta **72**: 1824–1836.

Stadnitskaia, A., G. Muyzer, B. Abbas, M. J. L. Coolen, E. C. Hopmans, M. Baas, T. C. E. van Weering, M. K. Ivanov, E. Poludetkina, and J. S. Sinninghe Damsté. 2005. Biomarker and 16S rDNA evidence for anaerobic oxidation of methane and related carbonate precipitation in deep-sea mud volcanoes of the Sorokin Trough, Black Sea. Marine Geology **217**: 67–96.

Sturt, H. F., R. E. Summons, K. Smith, M. Elvert, and Hinrichs. 2004. Intact polar membrane lipids in prokaryotes and sediments deciphered by high-performance liquid chromatography/electrospray ionization multistage mass spectrometry—new biomarkers for biogeochemistry and microbial ecology. Rapid Commun. Mass Spectrom. **18**: 617–628.

Suess, E., M. E. Torres, G. Bohrmann, R. W. Collier, J. Greinert, P. Linke, G. Rehder, A. Trehu, K. Wallmann, G. Winckler, and E. Zuleger. 1999. Gas hydrate destabilization: enhanced dewatering, benthic material turnover and large methane plumes at the Cascadia convergent margin. Earth and Planetary Science Letters **170**: 1–15.

Teichert, B., A. Eisenhauer, G. Bohrmann, A. Haase-Schramm, B. Bock, and P. Linke. 2003. U/Th systematics and ages of authigenic carbonates from Hydrate Ridge, Cascadia Margin: Recorders of fluid flow variations. Geochimica et Cosmochimica Acta **67**: 3845–3857.

Thiel, V., J. Peckmann, H. H. Richnow, U. Luth, J. Reitner, and W. Michaelis. 2001. Molecular signals for anaerobic methane oxidation in Black Sea seep carbonates and a microbial mat. Marine Chemistry **73**: 97–112.

Thiel, V., J. Peckmann, R. Seifert, P. Wehrung, J. Reitner, and W. Michaelis. 1999. Highly isotopically depleted isoprenoids: molecular markers for ancient methane venting. Geochimica et Cosmochimica Acta **63**: 3959–3966.

Torres, M., J. McManus, D. Hammond, M. De Angelis, K. U. Heeschen, S. L. Colbert, M. D. Tryon, K. M. Brown, and E. Suess. 2002. Fluid and chemical fluxes in and out of sediments hosting methane hydrate deposits on Hydrate Ridge, OR, I: Hydrological provinces. **201**: 525–540.

Tourova, T. P., T. V. Kolganova, B. B. Kuznetsov, and N. V. Pimenov. 2002. Phylogenetic Diversity of the Archaeal Component in Microbial Mats on Coral-like Structures Associated with Methane Seeps in the Black Sea. Microbiology **71**: 196–201.

Treude, T., A. Boetius, K. Knittel, K. Wallmann, and B. B. Jørgensen. 2003. Anaerobic oxidation of

methane above gas hydrates at Hydrate Ridge, NE Pacific Ocean. Marine Ecology Progress Series **264**: 1–14.

Tryon, M. D., K. M. Brown, and M. E. Torres. 2002. Fluid and chemical flux in and out of sediments hosting methane hydrate deposits on Hydrate Ridge, OR, II: Hydrological processes. Earth and Planetary Science Letters **201**: 541–557.

Tunnicliffe, V., S. K. Juniper, and M. Sibuet. 2003. Reducing environments of the deep-sea floor, p. 81–110. *In* Ecosystems of the World. Elsevier Press.

Watanabe, Y., S. Nakai, A. Hiruta, R. Matsumoto, and K. Yoshida. 2008. U–Th dating of carbonate nodules from methane seeps off Joetsu, Eastern Margin of Japan Sea. Earth and Planetary Science Letters **272**: 89–96.

White, D., D. B. Ringelberg, S. J. Macnaughton, S. Alugupalli, and D. Schram. 1997. Signature Lipid Biomarker Analysis for Quantitative Assessment In Situ of Environmental Microbial Ecology, p. 22–34. *In* R.P. Eganhouse [ed.], Molecular Markers in Environmental Geochemistry. American Chemical Society.

White, D., W. Davis, J. Nickels, J. King, and R. Bobbie. 1979. Determination of the sedimentary microbial biomass by extractible lipid phosphate. Oecologia **40**: 51–62.

Xie, S., J. S. Lipp, G. Wegener, T. G. Ferdelman, and Hinrichs. 2013. Turnover of microbial lipids in the deep biosphere and growth of benthic archaeal populations. Proceedings of the National Academy of Sciences **110**: 6010–6014.

Yoshinaga, M. Y., C. S. Lazar, M. Elvert, Y.-S. Lin, C. Zhu, V. B. Heuer, A. Teske, and Hinrichs. 2015. Possible roles of uncultured archaea in carbon cycling in methane-seep sediments. Geochimica et Cosmochimica Acta **164**: 35–52.

Yoshinaga, M. Y., M. Y. Kellermann, P. E. Rossel, F. Schubotz, J. S. Lipp, and Hinrichs. 2011. Systematic fragmentation patterns of archaeal intact polar lipids by high-performance liquid chromatography/electrospray ionization ion-trap mass spectrometry. Rapid Commun. Mass Spectrom. **25**: 3563–3574.

Zink, K.-G., H. Wilkes, U. Disko, M. Elvert, and B. Horsfield. 2003. Intact phospholipids—microbial "life markers" in marine deep subsurface sediments. Organic geochemistry **34**: 755–769.

C h a p t e r O n e

COMPARISON OF ARCHAEAL AND BACTERIAL DIVERSITY IN METHANE SEEP CARBONATE NODULES AND HOST SEDIMENTS, EEL RIVER BASIN AND HYDRATE RIDGE, USA

David H. Case[1,†]

in collaboration with,

Olivia U. Mason[1,2,†], Thomas H. Naehr[3], Raymond W. Lee[4], Randal B. Thomas[5], Jake V. Bailey[6], and Victoria J. Orphan[1]

[1]Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA 91125, USA

[2]Department of Earth, Ocean, and Atmospheric Science, Florida State University, Tallahassee, FL 32306, USA

[3]Department of Physical and Environmental Sciences, Texas A&M University-Corpus Christi, Corpus Christi, TX 78412, USA

[4]School of Biological Sciences, Washington State University, Pullman, WA 99164, USA

[5]US Geological Survey, Menlo Park, CA, 94025, USA

[6]Department of Earth Sciences, University of Minnesota, Minneapolis, MN 55455, USA

[†]In the publication of this article, first authorship is shared by D.H.C. and O.U.M. D.H.C. was the coordinating and principle author of the manuscript, performed iTAG preparation and analyses, and calculated and interpreted beta diversity analyses.

*****Published in Mason & Case et al., 2015. Comparison of Archaeal and Bacterial Diversity in Methane Seep Carbonate Nodules and Host Sediments, Eel River Basin and Hydrate Ridge, USA. Microbial Ecology 70 (3), 766-784.

1.0 ABSTRACT

Anaerobic oxidation of methane (AOM) impacts carbon cycling by acting as a methane sink and by sequestering inorganic carbon via AOM-induced carbonate precipitation. These precipitates commonly take the form of carbonate nodules that form within methane seep sediments. The timing and sequence of nodule formation within methane seep sediments are not well understood. Further, the microbial diversity associated with sediment-hosted nodules has not been well characterized and the degree to which nodules reflect the microbial assemblage in surrounding sediments is unknown. Here, we conducted a comparative study of microbial assemblages in methane-derived authigenic carbonate nodules and their host sediments using molecular, mineralogical, and geochemical methods. Analysis of 16S rRNA gene diversity from paired carbonate nodules and sediments revealed that both sample types contained methanotrophic archaea (ANME-1 and ANME-2) and syntrophic sulfate-reducing bacteria (Desulfobacteraceae and Desulfobulbaceae), as well as other microbial community members. The combination of geochemical and molecular data from Eel River Basin and Hydrate Ridge suggested that some nodules formed *in situ* and captured the local sediment-hosted microbial community, while other nodules may have been translocated or may represent a record of conditions prior to the contemporary environment. Taken together, this comparative analysis offers clues to the formation regimes and mechanisms of sediment-hosted carbonate nodules.

1.1 INTRODUCTION

Sulfate-coupled anaerobic oxidation of methane (AOM) is a significant biogeochemical process in continental margin settings and in areas of advective seafloor methane seepage, consuming a large fraction of methane in marine sediments prior to its release to the hydrosphere (Hoehler et al. 1994; Boetius et al. 2000; Reeburgh 2007). Within the seep environment, AOM is mediated by a symbiotic partnership between uncultured anaerobic methanotrophic archaea (ANME) and sulfate-reducing deltaproteobacteria (SRB). Sulfate-driven AOM increases the saturation state of sedimentary pore waters with respect to calcium carbonate by producing two units of alkalinity (Alk) per one unit of dissolved inorganic carbon (DIC), and has therefore been hypothesized to promote the precipitation of authigenic carbonate according to the general reactions (Luff and Wallmann 2003; Lein 2004; Luff et al. 2004):

$$CH_4 + SO_4^{2-} \leftrightarrow HCO_3^- + HS^- + H_2O;\ \Delta Alk = +2;\ \Delta DIC = +1 \quad \text{(Eq. 1)}$$

$$2HCO_3^- + Ca^{2+} \leftrightarrow CaCO_{3(s)} + CO_{2(aq)} + H_2O;\ \Delta Alk = -2;\ \Delta DIC = -1 \quad \text{(Eq. 2)}$$

Indeed, authigenic carbonates are often found in association with seep environments and AOM, and vary in morphology, size, and mineralogy. Observations have included cements (Hovland et al. 1987; Jørgensen 1989; 1992), nodules (sometimes termed "concretions"; Chen et al. 2006; Ussler and Paull 2008; Watanabe et al. 2008), massive chemoherm structures extending into the water column (Griffiths et al. 1982; Michaelis et al. 2002; Gulin et al. 2003; Teichert et al. 2005), and pavements that can cover hundreds of square meters (Paull et al. 1992; Boetius and Suess 2004). Most often, authigenic carbonates and nodules are observed within the sediment column or at the sediment/water interface (Greinert et al. 2001; Gieskes et al. 2005; Naehr et al. 2007; Haas et al. 2010).

Pore water geochemical profiles of Ca^{2+} and Alk, $\delta^{13}C_{DIC}$, $\delta^{13}C_{carb}$, ^{14}C labeling experiments, and lipid and DNA biomarkers provide strong evidence to substantiate the

hypothesized link between microbially-mediated AOM activity and the precipitation of authigenic carbonates (Michaelis et al. 2002; Peckmann and Thiel 2004; Boetius and Suess 2004; Naehr et al. 2009). It was recently shown that authigenic carbonates and nodules not only entrap microbial assemblages but furthermore host metabolically active methanotrophic populations, contributing substantially to methane oxidation in seep regions (Marlow et al. 2014a). Diagnostic lipids of methanotrophic archaea and their bacterial syntrophs, often showing characteristic depletion of ^{13}C, have been recovered in carbonate slabs from extant seep habitats and paleo-methane seeps dating as far back as the Pennsylvanian (Peckmann et al. 1999; Thiel et al. 2001; Peckmann and Thiel 2004; Stadnitskaia et al. 2005; Birgel et al. 2008b; a; Stadnitskaia et al. 2008; Naehr et al. 2009). A limited number of studies have also successfully recovered AOM-associated ANME and SRB 16S rRNA gene sequences associated with exhumed carbonate slabs and chemoherms within areas of methane seepage (Aloisi et al. 2002; Heijs et al. 2006; Stadnitskaia et al. 2008; Guan et al. 2013; Marlow et al. 2014b). These lipid and DNA biomarkers indicate a persistent relationship between relatives of AOM-associated archaea and deltaproteobacteria, and carbonate precipitation.

The discovery in 2002 of massive ANME-1 dominated deep-sea carbonate "reefs" extending tens of meters above active methane vents in the euxinic waters of the Black Sea further indicates that anaerobic methanotrophs do not merely colonize preformed carbonates but are capable of inducing and shaping their formation (Michaelis et al. 2002). Accordingly, the distribution of archaeal molecular signatures recorded in modern carbonates has been used to infer past environmental conditions and/or point to zones of previous AOM activity, as well as the ecological physiology of AOM-associated archaea. For example, archaeal ANME-2 16S rRNA gene sequences recovered from the upper part of a Gulf of Cadiz carbonate slab were interpreted to reflect carbonate precipitation in sediments containing elevated methane partial pressures, while the occurrence of ANME-1 sequences in the underlying crust were assumed to be

associated with a phase of precipitation under conditions of reduced methane flux (Stadnitskaia et al. 2008).

The majority of studies to date have focused on massive authigenic carbonates, which in many cases represent tens to hundreds of thousands of years of seep activity (Teichert et al. 2003; Luff and Wallmann 2003; Kutterolf et al. 2008; Ussler and Paull 2008; Watanabe et al. 2008; Liebetrau et al. 2010; Bayon et al. 2013). However, seep sediments themselves often harbor numerous millimeter-scale carbonate nodules representing diverse shapes, mineralogies, and formation histories (Chen et al. 2006; Ussler and Paull 2008; Watanabe et al. 2008). Due to their small size, nodules may capture a shorter time interval of *in situ* carbonate precipitation, and may be more relevant than massive carbonate pavements to understanding the immediate link between sediment- and carbonate-hosted microbial assemblages. Previous comparisons of sediment-, nodule-, and carbonate-associated microbial assemblages have suggested that archaeal community structures are dependent on methane seepage flux and not physical substrate type, while bacteria appeared to be more differentiated by substrate type than seep activity (Marlow et al. 2014b). Furthermore, endolithic microbial assemblages were found to be metabolically active, suggesting the possibility that 16S rRNA signatures recovered from nodules and carbonates might be different from the surrounding sediment-based assemblages (Marlow et al. 2014a). This implied that assemblages associated with nodules and carbonates might not be passive recorders of surrounding sediment communites but rather represent an extant, active, endolithic microbial community (Marlow et al. 2014a; b).

Comparative characterization of paired nodules and host sediments would provide insight into the degree to which early stages of carbonate formation passively record the sediment assemblage during lithification, or alternatively capture (or exclude) specific microorganisms directly mediating AOM and alkalinity production. Here, we examined microbial communities within sediments and their associated carbonate nodules (hereafter "nodules") from methane seeps located on the southern summit of Hydrate Ridge, OR, USA (HR; 44° 34.20351'N, 125°

8.8409'W; 800 meters below sea level (Boetius and Suess 2004)) and the northern ridge of Eel River Basin offshore of Eureka, CA, USA (ERB; 40° 48.7024'N, 124° 36.6754'W, 517 meters below sea level; Orphan et al. 2004). The microbial communities in 18 sediment and nodule samples (i.e., nine sediment/nodule pairs) from methane-seep environments of HR (n_{pairs} = 5) and ERB (n_{pairs} = 4) were characterized using iTAG sequencing of partial-length 16S rRNA genes to characterize relationships across geography ($HR_{n=10}$ *vs* $ERB_{n=8}$) and substratum ($sediment_{n=9}$ *vs* $nodule_{n=9}$). Depth (measured in centimeters below seafloor; cmbsf) was examined as an additional variable. Terminal restriction fragment length polymorphism (TRFLP) analysis using archaeal- and bacterial-specific 16S rRNA primers was also employed to complement diversity patterns observed using iTAG sequencing, in which universal primers were used. Additionally, four paired sediment and nodule samples from HR and ERB were selected for full-length archaeal and bacterial 16S rRNA gene cloning and sequencing. Pore water and solid phase geochemical, mineralogical, and isotopic analyses were conducted to provide physicochemical context for the interpretation of observed microbiological trends.

The goals of this comparative study were twofold: first, to determine whether the nodules reflect passive capture of the local sedimentary microbial community or host a unique microbial assemblage, and second, to examine the relationship between observed seep-associated microbial assemblages and physicochemical variables.

1.2 METHODS

1.2.1 SITE DESCRIPTION AND SAMPLE COLLECTION

Sediments and carbonates were recovered in September 2006 from active methane seep areas at HR and at ERB. All push cores (PC) were collected with *DSV Alvin* during dives AD4249 (HR: PC8) and AD4256 (ERB: PC29, PC23, PC20). Sampling locations were chosen based on the presence of benthic chemosynthetic communities (sulfide-oxidizing bacterial mats and

Calyptogena clam beds) – visual seabed indicators of localized seepage with high advective flux of sulfide, itself coupled via AOM to high subsurface methane fluxes (Sahling et al. 2002; Torres et al. 2002; Treude et al. 2003; Levin 2005).

Eel River Basin lies at the southern end of the Cascadia accretionary prism where organic-rich source rocks have led to the production and sequestration of abundant methane and other hydrocarbons (Orphan et al. 2004 and references therein). Variations in advective methane flux and pore water geochemistry occur frequently within methane seep habitats associated with sulfide-oxidizing microbial mats and chemosynthetic clam beds, and are sometimes organized in "bulls eye" structures at the seafloor (Barry et al. 1997; Treude et al. 2003; Orphan et al. 2004). Three push cores at ERB were collected along a lateral seep transect, from the center to the perimeter of a "bulls eye" consisting of a white sulfide-oxidizing microbial mat radially surrounded by *Calyptogena* clams. Two cores were collected from active seep zones (PC29, under a white microbial mat, hereafter "mat core"; PC23, under *Calyptogena* clams, hereafter "clam core"), and one core was collected on the edge of the "bulls eye" to capture low seep activity (PC20; hereafter "peripheral core"). Extensive sulfide-oxidizing microbial mats tend to exclusively overlie sulfidic seep sediments with a high methane flux (Boetius and Suess 2004). Chemosynthetic clams, which also rely on reduced fluids, cause substantial bioturbation, transporting seawater sulfate and oxygen to the underlying sediment layers and deepening the sulfate-methane transition zone (Orphan et al. 2004; Gieskes et al. 2005; Fischer et al. 2012). The periphery of these chemosynthetic clam and mat communities is typically defined by lower methane flux and correspondingly deeper sulfate penetration, slower rates of AOM, and lower concentrations of sulfide (Sahling et al. 2002; Treude et al. 2003; Orphan et al. 2004; Lloyd et al. 2010).

Nodules were recovered in only a few of the sectioned depth horizons along this ERB seep transect (Table 1), including two mid-depth sections in the mat core (6–9 cmbsf and 9–12 cmbsf; Pernthaler et al. 2008), one section in the clam core (0–3 cmbsf), and a deep section of the

peripheral core (9–12 cmbsf). Geochemistry and cell counts from this seep transect have previously been reported (Green-Saxena et al. 2014).

Hydrate Ridge lies approximately 250 miles north of Eel River Basin and is well known for extensive reduced fluid seepage, gas expulsion, and the presence of methane hydrates near the seafloor (Bohrmann et al. 1998; Suess et al. 2001; Boetius and Suess 2004). The site is associated with an accretionary complex located on the Cascadia Margin approximately 50 miles offshore Newport, OR, USA, and carbonate pavements are pervasive over much of the ridge (Bohrmann et al. 1998; Gieskes et al. 2005). At this site, the focus was on a single push core, PC8, which had carbonate nodules throughout its 0–15 cmbsf penetration depth (sectioned in 3-cm increments). PC8 was collected from the southern summit of Hydrate Ridge within a thick white microbial mat and processed shipboard according to previously published protocols (Orphan et al. 2001a). Samples were immediately frozen at -80ºC for subsequent DNA extraction.

1.2.2 X-RAY DIFFRACTION AND PETROGRAPHY

Bulk mineralogy and the relative abundance of carbonate minerals in each nodule sample were determined by X-ray diffraction (XRD) at the XRD Laboratory in the Department of Chemistry at Texas A&M University using a BRUKER D8 X-ray powder diffractometer. Samples for XRD analyses were prepared following standard procedures using an internal corundum standard (cf. Naehr et al. 2000). Scans were run from 2° to 60° 2θ at a scanning speed of 0.01°2θ/s. The relative proportions of different carbonate minerals were estimated on the basis of the (104) peak heights of calcite, Mg-calcite and dolomite, and the (111) peak height of aragonite (Table 1). Thin sections of carbonate nodules were examined using a Nikon Optiphot-pol polarizing microscope equipped with a Nikon DXM1200F digital imaging system.

1.2.3 METHANE AND SULFATE MEASUREMENTS

Methane was captured by immediately collecting 3–5 g sediment plugs into 1 M NaOH at a 1:1 g:mL ratio in gas-tight 20 mL serum vials. Methane concentrations in the headspace of the vials were determined with a Shimadzu mini-2 gas chromatograph equipped with a flame ion detector and magnesium perchlorate trap, with a 2 mL injection loop. A 9.93 ppm methane standard was used for calibration as described previously (Goffredi et al. 2008).

To collect pore waters for sulfate measurements, sediment samples were centrifuged (1380*g* for 15 min) in cut-off, stoppered 10 mL syringes without a headspace (Barry et al. 1996). Separated pore fluids were collected with a gas-tight syringe by puncturing the sidewall of the 10 mL syringe with a needle. Samples were preserved immediately in 0.5 M barium chloride (1:1::mL:mL). Sulfate in the pore fluids was determined by turbidimetry using a spectrophotometer (Gieskes et al. 1991). The turbidity was measured at 420 nm.

1.2.4 CARBON ISOTOPE AND CONCENTRATION ANALYSES

Carbon stable isotope measurements were performed on four sample types: total organic carbon ($\delta^{13}C_{org}$), nodule inorganic carbon ($\delta^{13}C_{nod}$), sedimentary inorganic carbon ($\delta^{13}C_{sed}$), and, in the case of samples from ERB, pore water dissolved inorganic carbon ($\delta^{13}C_{pw}$).

Isotopic composition of total organic carbon was measured on HR and ERB sediment samples, but not nodules, due to limited nodule material (Table 1). For the $\delta^{13}C_{org}$ analysis, sedimentary solid-phase inorganic carbon was removed by repeated application of 2 N phosphoric acid. Ten mg of dry material was placed into tin capsules and combusted in a Costech elemental analyzer (Valencia, CA). The resulting gases were separated by gas chromatography and admitted to the inlet of a GV Instruments (Manchester, UK) Isoprime isotope ratio mass spectrometer (IRMS). Typical $\delta^{13}C$ precision was ±0.2‰.

The solid-phase inorganic $\delta^{13}C$ was measured for nodules (HR & ERB; $\delta^{13}C_{nod}$) and sediments (HR; $\delta^{13}C_{sed}$; Table 1). All samples were dried at 60°C for 12 h, then milled to a fine powder. Powdered sediment or carbonate (0.4–1.0 mg) was placed in a labco vial, flushed with helium, then acidified with 100% phosphoric acid at 90°C. The CO_2 gas released was sampled and admitted to an Isoprime IRMS via a GV Multiflow preparatory system. Sodium bicarbonate was used as a consistency standard.

To analyze pore water isotopic composition (ERB; $\delta^{13}C_{pw}$; Table 1), N_2-pre-flushed 20-mL stoppered serum vials were amended with 0.1 mL of saturated ammoniacal $SrCl_2$. Squeezed pore water (1.8 mL) was added via syringe to the vial and stored as a basic solution until on-shore analysis. Prior to analysis, each sample vial was acidified with phosphoric acid and briefly vented to 1 atm to relieve slight positive pressure. Subsequently, a gas-tight syringe was used to remove and compress a sample to known volume. The sample was then injected into a continuous flow irm-GCMS instrument (Finnegan MAT 252) with a carbon-PLOT column (J&W Scientific) and splitless on-column injection. Due to sample limitation, DIC was not collected from Hydrate Ridge.

1.2.5 NODULE PREPARATION AND DNA EXTRACTION

Nodules analyzed in this study were recovered directly from clay-rich sediment. Thus, exterior portions of these samples were presumed to be contaminated with sediment-associated microorganisms. To minimize sedimentary contamination, a series of tests were performed to optimize the removal of external sediment microorganisms prior to DNA extraction from carbonate nodules (Supplemental Material). Ultimately, the most effective protocol for removal of loosely associated microorganisms from carbonate nodules was used for processing samples in this study. Specifically this entailed rinsing with 0.2 µm filtered 1X PBS, followed by sonication of the intact nodule at 8 W for 45 s in fresh, sterile, 1X PBS, and finally centrifugation at 4,000*g* for 5

min. This 3-step process was repeated 3*x* for each nodule. If any sediment-associated microorganisms were entrapped within the interior of nodules, and thus not removed by our techniques, they were considered endolithic for the purposes of this study. Nodules were powdered in sterile mortar and pestle prior to DNA extraction. Genomic DNA was extracted from sediments and nodules using a modified version of the MoBio UltraSoil DNA Isolation Kit (Orphan et al. 2001a).

1.2.6 16S rRNA GENE DEEP SEQUENCING (iTAG)

Initial PCR amplification was performed based on the specifications of the Earth Microbiome Project (EMP; Gilbert et al. 2011; Caporaso et al. 2011; 2012), with two exceptions: first, the single PCR step was split into two PCR steps, in which barcode indices were added at the second step in order to minimize PCR bias by employing long primers over many cycles (Berry et al. 2011). Thus, our first PCR followed the EMP protocol for 30 cycles with primers lacking adapter, barcode, pad, or linker (515f: GTGCCAGCMGCCGCGGTAA; 806r: GGACTACHVGGGTWTCTAAT). For the second PCR step, 5 µL of the amplicon product from PCR#1 was used as template in a 5 cycle, 25 µL reconditioning reaction with the same EMP-recommended conditions and the full EMP primers (515f_barcode: AATGATACGGCGACCACCGAGATCTACACTATGGTAATTGTGTGCCAGCMGCCGCGGTAA; 806r_barcode: CAAGCAGAAGACGGCATACGAGATXXXXXXXXXXXXAGTCAGTCAGCCGGACTACHVGGGTWTCTAAT). The second modification to the EMP protocol was to perform all PCR reactions in duplicate rather than triplicate. Internal lab tests showed that sequencing results were not significantly affected by including a third, triplicate, PCR product during preparation. After all PCR reactions were completed, duplicate barcoded products were pooled and quantified. Samples were mixed together in equimolar amounts and purified in bulk through a

Qiagen PCR Purification kit (Valencia, CA). At all PCR steps, amplification success and purity was checked by gel electrophoresis.

Paired-end sequences (2*x* 250 bp) were generated from barcoded amplicon products at Laragen, Inc (Los Angeles, CA) on an Illumina MiSeq platform. At Laragen the raw data was passed through a barcode filter which demultiplexed the library into individual samples and removed any sequences which had >1 basepair (bp) mismatch on the 12-bp barcode sequence (Golay barcodes were chosen with Levenshtein Distance ≥ 3; Caporaso et al. 2012). The resulting data were passed through MiSeq Recorder software (Illumina Inc, San Diego, CA), which assigned quality scores to each basepair call on every sequence. At the same time, adapter, barcode, and primer sequences were removed. The sequence data reported are available in the Sequence Read Archive under BioProject number PRJNA265122.

The sequences were then processed in-house with QIIME1.8.0 (Caporaso et al. 2010). The paired ends were first assembled into single contigs (join_paired_ends.py; $min_{overlap}$ = 50bp, $max_{bp\_mismatch}$ = 8%; Aronesty 2011). The contigs were then quality trimmed according to the q-scores, sequences with ambiguous 'N' base calls were removed from the dataset, sample names were added to each individual sequence, and the files were converted to fasta format (split_libraries_fastq.py; $q_{max\_unacceptable}$ = 29; max_N = 0). Chimeras were removed using the UCHIME_ref algorithm in USEARCH v7.0.1090 (minh = 0.28, xn = 8.0, dn = 1.4, mindiffs = 3, mindiv = 0.8; Edgar et al. 2011). The remaining sequences were used to pick *de novo* operational taxonomic units (OTUs) at 99% similarity (pick_otus.py; s = 0.99; Edgar 2010). Next, the default algorithm in QIIME1.8.0 was used for taxonomic assignments against representative sequences from each OTU (pick_rep_set.py; m = most_abundant; Wang et al. 2007). Taxa were assigned against the Silva 115 database clustered at 99% similarity (SSURef_NR99_115_SILVA_20_07_13_opt.arb; Quast et al. 2012), which was filtered to include only sequences with pintail value >75, and appended with 1,197 high-quality, full-length, seep-related bacterial and archaeal clones from Orphan lab clone libraries (assign_taxonomy.py; -

-uclust_max_accepts = 10; --uclust_min_consensus_fraction = 0.90; --uclust_similarity = 0.9; modified database is available upon request from the corresponding authors). The same appended database was used for the UCHIME_ref command described above. Singleton OTUs were removed from the dataset (remove_otus_from_otu_table.py; n = 2), as well as OTUs which were unassigned any taxonomy or assigned to Eukarya (filter_taxa_from_otu_table.py). Known contaminants in PCR reagents from sequencing of internal lab negative controls were removed by filtering out all sequences which clustered into Pseudomonadaceae, Enterbacteraceae, or Streptococcaceae, as well as extremely poorly defined Gammaproteobacteria observed in sequencing blanks (Gammaproteobacteria;Other;Other; Salter et al. 2014). In total these contaminant taxa accounted for an average of 2% of recovered sequences (range = 0–12%). Finally, tables of relative abundance were generated at the family level (summarize_taxa.py) and for each sample, families occurring at less than 0.01% relative abundance were removed in order to reduce the influence of spurious sequences. The full table of processed iTAG sequence data can be found in the Supplemental Material, while a summary of key taxa is included as Table 2.

Alpha (Shannon-Weiner) diversity was calculated in Microsoft Excel and beta (Bray-Curtis) diversity metrics were calculated in Primer-E (Clarke and Warwick 2001) from the family-level taxa abundance tables. For non-metric multidimensional scaling (NMDS) and analysis of similarity (ANOSIM) analyses, the taxa-abundance table was transformed with the square-root function prior to generation of the Bray-Curtis similarity matrix in Primer-E. Similarity Percentage (SIMPER) analysis, which deconvolves the whole-community differences between sample groups into quantitative contributions from each taxon, was also carried out in Primer-E (Clarke and Warwick 2001).

1.2.7 TERMINAL RESTRICTION FRAGMENT LENGTH POLYMORPHISM (TRFLP)

16S rRNA genes were amplified using archaeal primers 8F (fluorescently labeled with WellRED dye D4, Sigma-Proligo, St. Louis, MO) and 958R and bacterial primers 27F (fluorescently labeled with WellRED dye D3, Sigma-Proligo) and 1492R using the same PCR conditions as described in the Supplemental Material for clone libraries. PCR products were digested with *Hae*III overnight at 37°C, cleaned, and analyzed with a CEQ 8800 Genetic Analysis System from Beckman Coulter.

Prior to analysis, TRFLP peaks less than 70 bp were removed, thus avoiding spurious peaks that fall outside of the internal standards. Data were then converted to relative abundance, and peaks with relative abundance less than 1% were removed from further analysis. Further, peaks found in less than two samples were also removed from the dataset. Shannon-Weiner diversity indices were calculated using PC-ORD (Table 2; McCune et al. 2002). NMDS analysis was completed in Primer-E after square-root transforming the dataset and calculating Bray-Curtis similarities. NMDS coordinates were then transformed against reference iTAG NMDS coordinates in a procrustes analysis with QIIME 1.8.0 (transform_coordinate_matrices.py; r=1000, d=2). The purpose of this analysis was to test whether inter-sample similarity trends were supported between iTAG and TRFLP datasets. In both the archaeal and bacterial TRFLP beta diversity analyses, carbonate sample 2693 from ERB (clam core, 0–3 cmbsf, PC23) was determined to be an outlier (the outlier analysis in PC-ORD identifies samples whose community fingerprints are more than two standard deviations from the mean of the overall sample set (McCune et al. 2002) and was excluded from inclusion in procrustes analysis because it skewed the ordination plot beyond interpretation.

1.2.8 CLONE LIBRARIES AND FULL-LENGTH 16S rRNA GENE SEQUENCING

A subset of four samples were chosen for cloning and full-length 16S rRNA gene sequencing to gain greater taxonomic resolution of representative microbial taxa from HR and ERB (Table 1). Archaeal and bacterial libraries were prepared separately, resulting in eight clone libraries. A total of 384 bacterial and 384 archaeal clones were analyzed by restriction fragment length polymorphism (RFLP). Phylogenetic analysis and tree construction was carried out in ARB (Supplemental Material; Ludwig et al. 2004). The 16S rRNA gene sequences for the archaeal and bacterial clones were submitted to the GenBank database and are accessible under the following accession numbers: JQ036237–JQ036289. Clone library sequencing results for nodule 2518 at HR have been previously published in the Supplementary Material of Marlow et al. 2014a.

1.3 RESULTS

1.3.1 METHANE CONCENTRATIONS AND $\delta^{13}C$ OF ORGANIC AND INORGANIC CARBON

Recovered methane values were higher at HR than at ERB, and all values were consistent with previous descriptions of the HR and ERB methane seep regions (Table 1; Torres et al. 2002; Orphan et al. 2004). At HR, the recovered methane concentration was always >2 mmol CH_4 per g sediment and showed a minimum at 6–9 cmbsf (Figure 1a). Within the ERB horizons, recovered methane concentrations in the mat core were ~9-fold higher than in the clam and peripheral core horizons in which nodules were recovered, but all recovered methane concentrations were <1 mmol CH_4 per g sediment (Figure 1b). It is likely that some methane degassed during core recovery; therefore, the reported values should be taken as minimum methane concentrations.

At HR, the carbon isotopic composition of nodules from the PC8 mat core was always more ^{13}C-depleted than sedimentary inorganic carbon, which was in turn always more depleted than the organic carbon (Figure 1a). The depleted $\delta^{13}C_{nod}$ was indicative of a significant contribution of methane-derived bicarbonate to the sedimentary pore water DIC pool. The ERB nodule-bearing horizons across the seep transect demonstrated less consistent carbon isotopic results. In the two horizons with highest recovered methane concentrations (mat core, 6–9 cmbsf and 9–12 cmbsf), the nodules were more enriched in ^{13}C than either the pore water DIC or TOC (Figure 1b). The same relationship was true in the horizon from the peripheral core, in which <0.2 mmol CH_4 per g sediment was recovered. Although redox state was not determined, there was a notable change in sediment coloration in the ERB peripheral core relative to parallel cores collected beneath the microbial mat and clam bed, with shallow sediments having a brown-tan coloration transitioning to dark gray in the deepest sediment layers where the carbonate nodule was recovered. The nodule found in the ERB clam core was in the shallowest depth horizon (0–3 cmbsf), which demonstrated the lowest recovered methane concentration in this study (Table 1). The $\delta^{13}C_{pw}$ in this shallow horizon is relatively near the value of seawater (assumed ~0‰), consistent with bioturbation by *Calyptogena* clams. As observed in the HR samples, this nodule was also more depleted in ^{13}C than was the organic carbon (Figure 1b).

1.3.2 MINERALOGY AND PETROGRAPHY

X-ray diffraction (XRD) analyses revealed that nodules from both seep sites were at least partly composed of calcite (Table 1, Supplementary Figure 2). HR nodules were generally similar to one another, in that they were predominately composed of calcite with some (≤50%) aragonite and no measurable dolomite (Table 1). In contrast, the mineralogy of ERB nodules was more variable (Table 1, Supplementary Figure 2). Two of the ERB nodules (mat core 9–12 cmbsf; clam core 0–3 cmbsf) were composed entirely of calcite, while the other two (mat core 6–9 cmbsf;

peripheral core 9–12 cmbsf) contained a significant amount of dolomite (≥40%). None of the ERB nodules contained aragonite. The mineralogy of these sediment-hosted nodules was similar to previous descriptions of exhumed carbonates recovered from both the ERB and HR sites (Naehr et al. 2007).

Petrographic characterization of one representative nodule from HR (3–6 cmbsf) and ERB (PC29, 6–9 cmbsf) revealed distinct lithologies. For example, the HR nodule presented as a carbonate-cemented breccia, in which large angular carbonate clasts and bivalve fragments were cemented together by an aragonitic matrix (Figure 3c–d). Void-filling acicular aragonite cements were also abundant and internal fenestrate cavities in the nodule were surrounded by iron sulfide precipitates (Figure 3c). In comparison, thin section observations of the ERB nodule revealed a carbonate-cemented, quartz-dominated silt with low internal porosity (Figure 3a-b). Iron sulfide growth was observed to surround rare iron-rich lithic grains (Figure 3b). Discrete lithoclasts resembling the phyllosilicate glauconite were also observed, but are not visible in the field of view of the thin section images.

1.3.3 ARCHAEAL COMMUNITY COMPOSITION

Microbial community composition was estimated from recovered iTAG sequences from all 18 paired sediment/nodule samples in the study. Archaeal 16S rRNA clone libraries were additionally constructed from a subset of four samples, which were principally used to explore phylogenetic relationships between recovered HR and ERB clones and previously published 16S rRNA gene sequences. As a whole, the archaeal iTAG diversity data demonstrated consistent alpha diversity (Shannon-Weiner) across all samples (H'_{avg} = 1.7±0.3; Table 2). The greatest deviation from average appeared in the shallow (0–9 cmbsf) HR nodules, which exhibited lower alpha diversity than the majority of other archaeal iTAG data in this study. The deviation was not correlated with the number of recovered archaeal sequences. TRFLP data corroborated a

consistent level of archaeal alpha diversity across samples (H'_{avg} = 2.2±0.2), but without the deviation among shallow HR nodules (Table 2).

Euryarchaeotal ANME groups accounted for the majority of archaeal 16S rRNA iTAG sequences recovered from both sediment and nodule samples in most depth horizons of HR (PC8) and the ERB seep transect (PC29, PC23, PC20), comprising >35% of the recovered archaeal sequences in all samples (Table 2). Sequences associated with ANME-1 were more abundant in HR than in ERB samples across both nodule and sediment substratum types, by a factor of 2.3±1.6 (Table 2). ERB samples exhibited the highest observed ANME-2 abundances, but were also more variable than HR samples which had low-level ANME-2 presence that increased slightly with depth (Table 2). The observed variability of ANME-2 sequence abundance in ERB samples was not correlated to substratum or depth horizon.

Other commonly observed archaeal taxa in benthic marine settings were observed in iTAG data from both sediment and nodule samples. Sequences associated with DHVEG-6 were observed at higher relative enrichment in ERB samples than in HR samples by a factor of 4.6±4.2 (Table 2), and were also abundant in sediment samples from a Nankai Trough methane seep off Japan (Nunoura et al. 2012). Thermoplasmatales-associated sequences that clustered within the Marine Benthic Group D, which overlaps with the Deep Sea Hydrothermal Vent Group 1 (Takai and Horikoshi 1999; Teske and Sorensen 2008), were observed at consistent abundance in all samples, regardless of geographic location or substratum type (Table 2). Sequences associated with the Thaumarchaeotal Marine Benthic Group B were five-fold enriched in one sample relative to all others (ERB mat core nodule 9–12 cmbsf; Table 2).

Similarity rank ordering of the archaeal dataset was well-represented on a two dimensional NMDS plot, yielding a stress value of 0.07 when computed with the square root transformed archaeal iTAG data (Figure 2a). The samples were principally differentiated by geography (i.e., HR *vs* ERB), which an ANOSIM test revealed to be statistically robust ($p = 0.002$; $R = 0.63$; $n = 18$; Figure 2a). SIMPER analysis revealed that this geographical difference

was associated with observed abundances of ANME-1a and ANME-1b (more commonly observed in HR samples), and ANME-2ab and DHVEG-6 (more commonly observed in ERB samples).

At HR, the archaeal assemblages associated with sediments and nodules were all >70% similar (Figure 2a). Within the overall highly similar HR sample set, archaeal diversity was finely differentiated by substratum, with all sediments >80% similar to one another but less than 80% similar to the HR nodules ($p = 0.008$; $R = 0.51$, $n = 10$). This is in contrast to previous data, in which substrate type was not determined to be a factor differentiating seep archaeal communities (Marlow et al. 2014b). SIMPER analysis revealed that sequences associated with the subgroups ANME-1a and ANME-1b accounted for 20% of this substratum-based difference within HR samples. Relative abundances of recovered ANME-1a sequences were higher in nodules than sediments by a factor of 1.8±1.3, while ANME-1b sequences were observed at consistent relative abundance in HR sediments (0.43±0.03) and varied according to depth in the nodules (range 0.29-0.73, higher in shallow nodules; Table 2). Overall, shallow (0–9 cmbsf, $n = 3$) nodules were >80% similar to one another and deep (9–15 cmbsf, $n = 2$) nodules were also >80% similar to one another. Deeper samples were uniformly higher in MBGD relative abundance than shallow samples at HR. Shallow sediments were enriched in DHVEG-6. Sequences associated with the ANME-2c subgroup were observed at increasing relative abundance with depth for both nodules and sediments at HR, while ANME-2ab sequences were observed at consistent relative abundance in all HR samples (Table 2).

ERB archaeal sequences were <70% similar to HR samples. Furthermore, whereas HR samples were differentiated by substratum, the ERB samples appeared to be primarily differentiated by depth and were not significantly separated by substratum ($p=0.69$), in agreement with previously published data (Marlow et al. 2014b). Deep (9–12 cmbsf) ERB samples were >70% similar to one another but not to other ERB samples. The mid-depth ERB sediment-

nodule pair from the mat core (6–9 cmbsf) was >70% similar, and the shallowest pairing (clam core; 0–3 cmbsf) was <70% similar to one another and not similar to any other samples.

Full-length ANME 16S rRNA gene sequences grouped with phylotypes recovered from other methane seep sites within the Santa Barbara and Eel River Basins (Figure 4; Orphan et al. 2001a), Hydrate Ridge (Knittel et al. 2005), and other seep sites (e.g. Heijs et al. 2005). ANME-2a and ANME-2b phylotypes have also been reported from carbonate crust samples associated with submarine mud volcanoes (Heijs et al. 2005; Stadnitskaia et al. 2008). However, none of these phylotypes were closely related to the clone library archaeal sequences recovered from ERB and HR carbonate nodules. The majority of ANME-1b clones were most closely related to phylotypes from seep sites and other reducing sediment habitats (Knittel et al. 2005; Kendall et al. 2007). Sequence representatives associated with ANME-1a were not recovered. As with ANME-2a and -2b, the carbonate-associated ANME-1b phylotypes were distinct from those reported by Stadnitskaia et al. 2008 and Heijs et al. 2006.

1.3.4 BACTERIAL COMMUNITY COMPOSITION

Unlike the archaea, the Shannon-Weiner diversity of bacterial iTAG sequences in HR and ERB sediments samples decreased with increasing depth (R^2 = 0.41, HR and ERB; R^2 = 0.52, HR only), as has been observed in other deep-sea sedimentary environments (Lloyd et al. 2010). This depth trend was even more apparent in the TRFLP data (R^2 = 0.75, HR and ERB; R^2 = 0.82, HR only). As was observed in the archaeal dataset, alpha diversity was uncorrelated to number of recovered sequences and sediments at HR were slightly more diverse than nodules, especially in shallow (0–9 cmbsf) horizons (observed in both iTAG and TRFLP data; Table 2).

Deltaproteobacteria, dominated by members of Desulfobacteraceae and Desulfobulbaceae, were observed across all samples in the iTAG data set, regardless of geography or substratum type (Table 2). Recovered Desulfobulbaceae sequences decreased with depth in

both sediment and nodule samples from the ERB mat core (PC29; 6–9 cmbsf and 9–12 cmbsf), as has previously been observed in sediments by fluorescence *in situ* hybridization in a separate study on the same core (Green-Saxena et al. 2014). Numerous Desulfobacteraceae and Desulfobulbaceae full-length 16S rRNA gene clones were recovered from sediments and nodules at both methane seep sites (Figure 5). Related clones were reported from HR sediments (Knittel et al. 2005), Santa Barbara Basin and ERB sediments (Orphan et al. 2001a), and from the previously analyzed overlying 3–6 cm interval of the ERB mat core (PC29; Pernthaler et al. 2008).

Epsilonprotebacteria were often observed at higher relative abundance in sediments than in nodules at both HR and ERB sites, and were most often associated with the genus Sulfurovum in the Helicobacteraceae family. Related Epsilonproteobacteria have been previously observed in shallow cold seep sediments (Roalkvam et al. 2011; Nunoura et al. 2012; Niemann et al. 2013) and are related to known sulfur oxidizers (Inagaki et al. 2004). Epsilonproteobacterial clones were also recovered, and the closest cultured relatives were the sulfur-oxidizers *Sulfuricurvum kujiense* (Kodama and Watanabe 2003) and *Sulfurimonas autotrophica* (Inagaki et al. 2003), both members of the Helicobacteraceae family. In six of the nine sediment-nodule pairs in this study, the Epsilonproteobacterial iTAG relative abundance in the sediment was greater than a factor of five over the corresponding nodule (n_{HR} = 4; n_{ERB} = 2; Table 2). In one case where the nodule conversely had higher abundance than the sediment (PC29, 9–12 cmbsf), the difference was so small it may be insignificant.

At HR, relative abundances of sequences associated with Gammaproteobacteria were higher in sediments than nodules, whereas at ERB the nodules were elevated in Gammaproteobacteria relative abundance as compared to the sediments. ERB samples tended to have overall higher Gammaproteobacterial relative abundances than HR samples (Table 2). Additional bacterial diversity, observed in both iTAG and clone library data, included members of the Bacteroidetes (consistent across geography and substratum type), the Nitrospirae

(principally ERB nodules), the Chloroflexi (enriched at HR over ERB by a factor of 2.0±1.6), and Candidate Division JS1, all previously described from marine methane seeps.

Bacterial communities were represented on a two dimensional Non-Metric Multidimensional Scaling (NMDS) plot with a stress value of 0.05, and the overall separation was similar as observed with archaea, where bacterial assemblages were significantly differentiated by geography ($p < 0.001$; $R = 0.35$; $n = 18$; Figure 2b). SIMPER analysis revealed that the epsilonproteobacterial Helicobacteraceae (high in HR sediments), Candidate Division JS1 (high in HR sediments and nodules), and Desulfobacteraceae (high in ERB sediments and nodules) were associated with HR *vs* ERB differences.

Within the HR samples, as was observed in the archaea, the bacterial assemblages were separated by substratum ($p = 0.008$; $R = 0.67$; $n = 10$), consistent with previous observations of bacterial community structure (Marlow et al. 2014b). The taxa most strongly associated with this separation were the epsilonproteobacterial Helicobacteraceae (enriched in sediments over nodules) and deltaproteobacterial Desulfobacteraceae (enriched in nodules over sediments). The detailed breakdown by substratum- and depth-dependent factors was more complex with bacteria than with archaea (Figure 2b). Whereas all ten HR archaeal assemblages from the mat core (PC8) were highly similar to one another (with some fine-scale differences as presented above), three of the shallow bacterial HR sediment assemblages (PC8, 0–9 cmbsf) were <70% similar to the main bacterial HR cluster of five nodules (0–15 cm) and two deeper sediment horizons (9–15 cm). Those three HR sediment bacterial assemblages were >80% similar to one another and demonstrated the highest recovery of epsilonproteobacterial Helicobacteraceae sequences among all 18 bacterial community samples. Two ERB samples that clustered near the shallow HR sediments also contained a high abundance of Helicobacteraceae sequences.

Within the cluster of seven similar HR samples, as with archaea, the nodule-associated bacterial assemblages separated into shallow (0–9 cmbsf) and deep (9–15 cmbsf) groups. The two deep HR sediment samples were most similar to the shallow nodules (Figure 2b). There was

generally more variability in bacterial iTAG data from the ERB transect than from the single HR core. To some extent, ERB bacterial assemblages appeared to be differentiated by depth. Most deep samples were highly similar to one another, as was observed in the archaeal ERB data, and the shallow samples demonstrated high biological dissimilarity (Figure 2b).

1.4 DISCUSSION

In modern and ancient settings, authigenic carbonates that precipitate as a result of AOM may provide a geological record of anaerobic methanotrophy in marine sediments (Peckmann and Thiel 2004). However, the degree to which precipitation passively captures a biological record of sediment-hosted microorganisms, or represents a distinct carbonate-hosted microbiome predicated on unique physicochemical constraints, remains unclear (Marlow et al. 2014a; b). More fundamentally, the consistency of the relationship between the microbial diversity of host sediments and carbonate nodules across geochemical regimes remains unexplored. Through parallel molecular, geochemical, and isotopic analyses of seep sediments and the carbonate nodules they host, we addressed these outstanding questions.

At HR, nodules were uniformly depleted in ^{13}C ($\delta^{13}C_{nod}$ = -45.9±3.2‰) relative to other carbon phases, including sedimentary inorganic carbon and organic carbon (Figure 1). These isotopic values are consistent with the relatively high recovered methane concentrations at HR, which could enable high AOM rates, as well previous isotopic measurements of seafloor carbonates from this site (Greinert et al. 2001). Although the carbon isotopic composition of methane in the mat core (PC8) from HR was not measured, it can reasonably be predicted to be depleted in ^{13}C, as has been consistently reported from other studies at HR (Suess et al. 1999; Boetius and Suess 2004). A high rate of AOM lowers the $\delta^{13}C$ value of the DIC pool, a signal that is then incorporated into carbonate nodules (Ussler and Paull 2008). The consistent offset between $\delta^{13}C_{nod}$ and $\delta^{13}C_{sed}$ ($offset_{avg}$ = 9.6±1.2‰) also merits consideration. It is possible that the

inorganic carbon isolated from the sediment represents more recent precipitation than the nodules, or, perhaps, the bulk sediment includes some carbonate at circa 0‰ (e.g., planktonic foraminifera tests) that is not methane-derived. This could make the bulk sediment appear less ^{13}C depleted than the nodule that is composed of all or mostly methane-derived carbon. Based on the isotopic offset, one interpretation is that $\delta^{13}C_{sed}$ may represent a contemporary snapshot of AOM activity, while the $\delta^{13}C_{nod}$ may represent a longer, time-integrated history of AOM activity at HR. Indeed, other studies have found that carbonate nodules and concretions precipitate over 10^2 to 10^4 years, and therefore represent time-integrated records of seep activity (Luff et al. 2004; Ussler and Paull 2008). The $\delta^{13}C_{org}$ values at HR were only moderately depleted in ^{13}C (-30.5±2.5‰), suggesting other contributions besides AOM-associated organisms to the total sediment-associated organic carbon pool.

Regardless of the timing of nodule formation, the geochemical data at HR was consistent across depth and concordant with conditions favoring *in situ* carbonate formation within seep sediments: relatively high methane concentrations fueling AOM and thus an increase in alkalinity and carbonate saturation (Luff et al. 2004), followed by precipitation of nodules with depleted ^{13}C content. The archaeal iTAG data are in agreement with this interpretation: all HR samples (nodules and sediments) demonstrated similar Shannon-Weiner diversity and high community similarity to one another, especially when contrasted with the diversity in ERB molecular data (see discussion below). Moreover, the overall archaeal similarity at HR was linked in SIMPER analysis to the observed abundance of ANME-1 subgroups, taxa known to be involved in AOM (Hoehler et al. 1994; Boetius et al. 2000; Orphan et al. 2001b). Thus, it appears that in geochemical regimes favorable for AOM, nodules broadly mirror the archaeal communities found in surrounding sediment. This broad finding is consistent with previous findings that substrate type was not a major differentiator of archaeal populations (Marlow et al. 2014b).

However, close examination of the dataset reveals further structure to the molecular data, which reflects subtle differences in the nodule assemblage relative to the host sediment as a

function of depth. This is most likely ultimately due to depth-dependent differences in geochemistry affecting changes in microbial assemblage composition, as has been observed previously (Lloyd et al. 2010). Within the archaeal communities, shallow (0–9 cmbsf) nodules were distinct from deep (12–15 cmbsf) nodule-associated microbial communities (Figure 2a). This suggestion of a depth-dependent factor driving microbial communities was even stronger in the bacterial iTAG data, where all nodule-associated bacterial communities at HR were highly similar to deep (12–15 cmbsf) sediment communities, but different from shallow (0–9 cmbsf) sediment-hosted communities (Figure 2b). This molecular evidence thus suggests that nodules might be formed within the deep sediment horizons, entrap the adjacent microbial communities during formation, and may be subsequently transported upward, perhaps by local uplift and sediment erosion, bioturbation, or seismic activity.

That a depth-dependent trend, and inferred translocation, was strongest in bacterial molecular data at HR is intrinsic to the fundamental differences between the geochemical and archaeal *vs* the bacterial datasets: the geochemical characteristics and archaeal sediment diversity were largely homogenous with depth, and so did not provide a framework for observing strong differences across depth within the studied sediment core (PC8). Since the bacterial sediment-associated microbial communities were well-differentiated into deep and shallow groups, the effects of an origin at depth and vertical translocation of nodules, if true, was observable. TRFLP data generally supported the iTAG molecular observations (Supplemental Material). Nonetheless, it cannot be ruled out that rather than translocation, local geochemical conditions may have shifted over time, followed by a shift in the sediment-associated microbial community but unobserved in the DNA recorded within nodule precipitates. Alternatively, macrofaunal grazing pressures in shallower horizons could have influenced the selective enrichment of native sediment microbiota (Thurber et al. 2012).

Petrographic evidence was inconclusive regarding the origination of nodules at HR. The 3–6 cmbsf nodule, the only nodule from HR that was examined petrographically, exhibited

bivalve fragments contained within an aragonitic cement (Figure 3c). Although the bacterial mat site where the PC8 core was recovered did not exhibit clam beds at the seafloor, it is possible that bivalve shell hash present in the underlying sediment was historic. The petrographic fabric of the HR nodule is consistent with one or more phases or carbonate cementation and re-precipitation. This is evidenced by the incorporation of angular clasts of previous generations of authigenic carbonate cemented into an aragonitic matrix. The angular shape of the clasts suggests localized disruption, perhaps from hydrofracturing of carbonate mudstones, involving little or no immediate subsequent transport. The absence of clasts within intraclasts suggests only one disruption event followed by cementation. A disruption event may support the hypothesis that the HR nodules formed in a deep horizon and were subsequently exhumed. However, abundant un-oxidized sulfide precipitates suggest that the nodule has not been uplifted enough for exposure to oxygenated conditions sufficient to alter those phases. This is consistent with the mat-type habitat, which is not expected to greatly bioturbate the sediment. Overall, examination of the paired sediment/nodule depth profile (0–15 cmbsf) at HR suggested that, at active seeps, carbonate nodules precipitate within a few 10s of centimeters below seafloor. These nodules can and do capture the sediment-hosted microbial community, and multiple scenarios can be invoked to explain cases where biological deviation is observed between the nodules and adjacent sediments.

Examining the geochemical data from ERB, it is clear that sediment/nodule pairs across the seep transect exhibited more complex relationships than within the single PC8 core at HR. The 0–3 cmbsf horizon from PC23 (clam core) demonstrated geochemical characteristics that do not predict a favorable environment for carbonate precipitation, despite the recovery of a nodule. The very low recovered methane concentrations indicate little contemporary geochemical driving force for alkalinity generation via AOM, and the somewhat ^{13}C-enriched $\delta^{13}C_{pw}$ value (-7.4‰; Table 1) is likely a combination of low AOM rates and mixing of ~0‰ seawater due to bioturbation by *Calyptogena* clams. Molecular data further indicate the nodule did not precipitate *in situ* in these geochemical conditions, where both archaeal and bacterial diversity, recovered

both by iTAG and TRFLP, clearly shows that the nodule is distinct from the host sediment (Figure 2). Thus, two conclusions are drawn. First, that bioturbation of shallow sediment by overlying clams generates a distinct archaeal and bacterial sediment-associated microbial community, and second, that the recovered nodule either originated in a separate location and was subsequently moved by sediment winnowing or uplift to the shallow location from which it was recovered, or precipitated *in situ* at a time when sediment geochemistry and microbial populations were different than their modern states.

The PC29 mat core (6–9 and 9–12 cmbsf) was collected from beneath a sulfide-oxidizing bacterial mat at ERB, and is thus most parallel to the HR mat core (PC8) with respect to benthic habitat type. However, the samples exhibited geochemical and microbiological variability which exemplified the potential for inhomogeneity within seep ecosystems, compared to the relatively homogenous conditions in HR core PC8. Centimeter-scale vertical variability (both geochemical and biological) has previously been reported from methane seep sediments, including Hydrate Ridge (Treude et al. 2003), Eel River Basin (a separate study of the same cores we sampled for this study; Green-Saxena et al. 2014), and the Gulf of Mexico (Lloyd et al. 2010). The depleted $\delta^{13}C_{org}$ value (-40.7‰) from the shallower (6–9 cmbsf) horizon of the ERB mat core is indicative of significant methanotrophic biomass, along with depleted $\delta^{13}C_{pw}$ (-36.1‰) suggestive of active AOM. The deeper horizon's biomass ($\delta^{13}C_{org}$ = -31.0‰) and pore water inorganic carbon ($\delta^{13}C_{pw}$ = -32.9‰) values also suggest active AOM processes, although perhaps at a more moderate rate. The deeper nodule's carbon isotope enrichment over the pore water ($\delta^{13}C_{nod}$ = -27.2‰; $\Delta^{13}C_{nod\text{-}pw}$ = 5.7‰) also suggests at most moderate AOM rates. That the nodule is slightly enriched in ^{13}C relative to surrounding pore water could be the result of time-integrated precipitation of the nodule over varying or different historic conditions.

The shallower nodule more substantially deviated from the pore water carbon isotopic composition ($\delta^{13}C_{nod}$ = -23.1‰; $\Delta^{13}C_{nod\text{-}pw}$ = 13.0‰). However, mineralogical evidence revealed that the shallower nodule was composed of 40% dolomite, and nodules that include substantial

dolomite have been documented to have much more ^{13}C-enriched carbon isotopic signatures (Greinert et al. 2001; Naehr et al. 2007). Indeed, a ^{13}C-enriched value was also observed for the other dolomite-containing nodule recovered at ERB (PC20, 9–12 cmbsf, Table 1). Microbially-mediated methanogenesis enriches the DIC pool in ^{13}C; thus, ^{13}C-enriched dolomites are often interpreted to have formed in deeper, methanogenic horizons, often below the sulfate methane transition zone (SMTZ; Greinert et al. 2001; Naehr et al. 2007). Sulfate-reducing bacteria can also mediate dolomite precipitation (Vasconcelos et al. 1995; 2005; Krause et al. 2012), but with a depletion rather than enrichment in $\delta^{13}C$ as observed in our data (Vasconcelos et al. 1995). Previous geochemical characterization of PC29 suggested the local SMTZ peaks at 6–9 cmbsf (Green-Saxena et al. 2014), making it likely that the methanogenic zone, which may have originally hosted the nodule, was deeper in the sediment column. Thus, a possible interpretation of this study's geochemical data is that both the 6–9 cmbsf and 9–12 cmbsf sediments in the mat core host contemporary, active AOM, but the nodules record one or more intervals of environmental conditions that differed from current conditions, or record conditions from precipitation elsewhere than their recovery location.

The microbiological iTAG data are mostly consistent with this hypothesis. In the deeper sediment horizon, the archaeal and bacterial sequence data demonstrated close coupling between the sediment/nodule pair (Figure 2). Both the sediment and nodule exhibited high abundances of AOM-associated taxa, most notably ANME-1 subgroups and Desulfobacteraceae (Table 1). Thus, the nodule appeared to passively mirror the adjacent sediment-associated microbial assemblage as was the case at HR. In the shallower horizon, the sediment/nodule pair similarity was notably low within the bacterial data and consistent with either nodule translocation or a nodular signal of preserved, relic genetic material combined with signatures from extant endolithic microorganisms (Figure 2). Indeed, endolithic microbial activity in massive seep carbonates and nodules has been recorded recently (Marlow et al. 2014a).

Bacteria appear to be more sensitive to nodule provenance, translocation, and/or geochemical shift than archaea, as even in the generally homogenous HR core, the bacteria demonstrated variability – shallow sediments were relatively dissimilar from other HR samples. In the shallower ERB horizon, for which bacterial and geochemical data support translocation from and/or precipitation within a methanogenic zone, the archaeal iTAG data contrasts by suggesting a tightly coupled sediment/nodule pair and little evidence of conventional methanogens such as members of the family Methanosarcinaceae (Figure 2a; Supplemental Material). Close analysis reveals that this tight archaeal coupling is due to the highest observations of ANME-2ab-affiliated sequences in this study's entire sample set. Previous measurements have found carbonate nodules to contain 100-fold more methane than surrounding sediments due to adsorption processes (Ijiri et al. 2009). Therefore, methane-consuming taxa might have a strong driving force to colonize nodules. The mechanism behind such colonization remains unknown, as ANME are thought to have doubling times on the order of several months and are not know to be motile. However, transport and/or colonization could occur via seep metazoans, some of which have been demonstrated to feed on archaea as a food source (Thurber et al. 2012). It is possible metazoans could act as a transportation mechanism for undigested microorganisms.

Whether specific mineralogy plays a role in microbial colonization is undetermined, but could be a factor contributing to some of the decoupling between sediment and nodule assemblages. The ERB nodule which was petrographically examined, from the 6-9 cmbsf horizon of the mat core, was a carbonate-cemented siliclastic sediment – a lithology that is common at seep sites throughout the world (e.g. Peckmann et al. 2001; Campbell et al. 2010). The ERB nodule did not show evidence of multiple stages of carbonate precipitation, erosion, dissolution, or exposure at the sediment/water interface. Changes in the redox regime in seep-associated sediments can commonly result in the partial or complete oxidation of sulfide mineral phases, including in the Eel River Basin (Bailey et al. 2010). The presence of unoxidized sulfides in the ERB nodule suggest that the nodule has not encountered oxygenated conditions, consistent with

the overlying presence of a bacterial mat which would not bioturbate the underlying sediment. The undisturbed, un-oxidized condition of the nodule is not conclusive but suggests it may not have been translocated, which would be consistent with the closely coupled archaeal sediment/nodule data.

Interestingly, the peripheral ERB core (PC20) exhibited geochemical data most similar to the 6–9 cmbsf horizon from the ERB mat core (PC29). That is, the nodule contained a significant fraction of dolomite (50%; Table 1) and was more ^{13}C-enriched than either the pore water or organic carbon. The bacterial iTAG data suggests the nodule is not similar to the paired sediment (Figure 2b), which is corroborated by bacterial TRFLP data. The archaeal iTAG data suggests a closer coupling (Figure 2a), although it does not appear to be due to ANME subgroup similarities. Indeed, the nodule is depleted in ANME-1 subgroups and enriched in ANME-2 subgroups relative to the host sediment. The observation of a large number of ANME-2-affiliated sequences in the dolomite-containing nodule is consistent with ANME-2-affiliated sequences recovered from the 6–9 cmbsf nodule in the ERB mat core (PC29). Thus, the bulk of the geochemical and molecular evidence from the deep horizon of the peripheral ERB core leads to a similar conclusion as for the 6–9 cmbsf sediment/nodule pair from the mat core: the nodule exhibits signatures consistent with possible original precipitation elsewhere and/or within a different geochemical regime and subsequent translocation to the current site. This hypothesis is more strongly supported by the bacterial sequence data than the more ambiguous archaeal and petrographic signals.

1.5 CONCLUSIONS

The ^{13}C-depleted carbonates, organic lipid biomarkers, and associated 16S rRNA gene signatures previously documented from carbonate pavements and chemoherm structures provide evidence for the involvement of archaeal methanotrophs and their syntrophic sulfate-reducing

bacterial partners in carbonate precipitation (Heijs et al. 2006; Stadnitskaia et al. 2008; Marlow et al. 2014b). Previous microbiological data suggested that nodule-associated microbial assemblages might not be simply passive recorders of sediment-associated microorganisms, but could host distinctive, extant, active microbial populations (Marlow et al. 2014a; b). This study directly addressed that hypothesis and confirmed that in some cases divergence was observed between sediment- and nodule-associated assemblages, while in other cases nodules most likely precipitated *in situ* and entrapped the local microbial communities. This may be due to translocation laterally and/or vertically, or shifting local geochemical conditions. Sediment/nodule disconnect appears to be a more common phenomenon among nodules recovered from shallow sites, potentially indicating that nodules form within deeper AOM horizons of the sediment column (~9–15 cmbsf), where alkalinity generation is highest, and subsequently experience exhumation from sediment winnowing. Petrographic evidence from HR indicated post-depositional fracturing of mineral phases, potentially supporting translocation of the nodules. Alternatively, bioturbation from above may rapidly change shallow sediment geochemistry, to which sediment-associated microbial assemblages may respond more quickly than nodule-associated assemblages. Studies suggesting nodules and carbonate slabs grow over 10^2 to 10^4 years support the likelihood that translocation could occur, given the geologic activity at regions such as HR and ERB. If nodules form over prolonged timescales, then the tight sediment/nodule coupling observed in some of this study's samples implies that microbial assemblages can maintain a stable composition over extended periods. Alternatively, it may be that nodules can precipitate over timescales much less than 10^2 years. Further, in petrographic thin sections the presence of reduced minerals entrapped within carbonate nodules from HR and ERB indicate that subseafloor and/or intra-nodule conditions may remain reducing for periods of time at least as long as the lifetime of the nodules. The coupling between seep flux, microbial carbon cycling, and mineralogy can be further explored with a larger sample set. Furthermore, the degree to which microbial assemblage entrapment, as demonstrated in this manuscript for

nodules, extends to massive carbonate pavements, merits further exploration. In the future, comparative tracking of carbonates and host sediments will contribute information to further constrain the many factors influencing the timing, location, and diversity of organisms linked to authigenic carbonate precipitation during sulfate-coupled AOM.

1.6 ACKNOWLEDGMENTS

VO conceived of the study and collected the samples at sea. OM processed the samples and optimized DNA extraction, as well as TRFLP and clone library analyses. DC performed iTAG processing and analyses, as well as beta diversity analyses, and was the coordinating author of the manuscript. TN provided XRD data, RL performed the isotopic composition analyses, JB provided thin section images, and RT performed the pore water geochemical measurements. DC, OM, and VO principally contributed to writing the manuscript. Three anonymous reviews provided constructive suggestions to improve the manuscript.

Elizabeth Trembath-Reichert, Stephanie Connon, and Jeff Marlow helped with customization of the Silva115_NR99 database. Alexis Pasulka provided helpful discussion regarding ordination and statistical probing of microbial communities. Josh Steele also provided discussion on ecological statistics and aided with bench-top lab work. Benjamin Harrison helped with TRFLP data interpretation. Jeff Marlow provided useful feedback on the manuscript. The crew of the R/V *Atlantis* cruise AT-15-11, as well as the pilots of DSV *Alvin* dives AD4249 and 4256, aided in sample recovery at sea.

Funding for this work was provided by a National Science Foundation grant (BIO-OCE 0825791) to VO and an early career grant by the United States Department of Energy, Office of Biological and Environmental Research (DE-SC0003940) to VO. This research was also supported by a grant from the NASA Astrobiology Institute (Award #NNA13AA92A) to VO. This is NAI-Life Underground Publication 009. DC was funded by a National Science Foundation Graduate Research Fellowship.

1.7 TABLES

Table 1. Geochemical Observations. Blank cells indicate measurement was not applicable. Cells with "–" indicate measurement was applicable but not completed, generally due to a limitation from the amount of sample collected or available

		Sample number	Core	Benthic ecosystem	Depth (cmbsf)	Methane (mmol CH_4/g sediment)	Sulfate (mM)	$\delta^{13}C_{org}$ (‰)	$\delta^{13}C_{nod}$ (‰)	$\delta^{13}C_{sed}$ (‰)	$\delta^{13}C_{pw}$ (‰)	Mineralogy (%calc/%arag/%dolo)
Hydrate Ridge	Sediments	S2517	PC8	Mat	0–3	4.33	13.99	−29.7		−35.4	–	–
		S2518	PC8	Mat	3–6	2.74	0.00	−34.5		−38.6	–	–
		S2519	PC8	Mat	6–9	2.10	10.15	−31.2		−39.5	–	–
		S2520	PC8	Mat	9–12	4.16	8.28	−29.2		−37.3	–	–
		S2521	PC8	Mat	12–15	3.41	8.72	−27.9		−32.0	–	–
	Nodules	C2517	PC8	Mat	0–3	4.33	13.99	–	−45.9			90/10/0
		C2518	PC8	Mat	3–6	2.74	0.00	–	−49.0			60/40/0
		C2519	PC8	Mat	6–9	2.10	10.15	–	−47.3			50/50/0
		C2520	PC8	Mat	9–12	4.16	8.28	–	–			–
		C2521	PC8	Mat	12–15	3.41	8.72	–	−41.6			100/0/0
Eel River Basin	Sediments	S2688	PC29	Mat	6–9	0.95	19.24	−40.7		–	−36.1	–
		S2689	PC29	Mat	9–12	0.91	5.40	−31.0		–	−32.9	–
		S2693	PC23	Clam	0–3	0.06	15.28	−24.2		–	−7.4	–
		S2703	PC20	Peripheral	9–12	0.14	25.94	−25.0		–	−31.9	–
	Nodules	C2688	PC29	Mat	6–9	0.95	19.24	–	−23.1			60/0/40
		C2689	PC29	Mat	9–12	0.91	5.40	–	−27.2			100/0/0
		C2693	PC23	Clam	0–3	0.06	15.28	–	−34.1			100/0/0
		C2703	PC20	Peripheral	9–12	0.14	25.94	–	−18.0			50/0/50

Table 2. Microbiological Observations. All microbial relative abundance data shown is calculated from iTAG sequencing results. Heat scales are calculated separately for each column. As is always the case in PCR-based analyses, relative abundance between taxa (i.e., comparing across a row) should be interpreted cautiously. A safer interpretation is the relative abundance of one taxon across multiple samples (i.e., comparing down a column). Shannon–Weiner diversity (H′) is also given for Archaea and Bacteria, calculated from both the iTAG and TRFLP data. For a complete reporting of iTAG sequencing results, see Supplemental Tables 1 and 2.

		Sample number	Core	Benthic Ecosystem	Depth (cmbsf)	16S rRNA gene deep sequencing	Clone libraries	*Archaea*											*Bacteria*						
								H' (TRFLP)	H' (iTAG)	ANME (total)	ANME-1a	ANME-1b	ANME-1_{other}	ANME-2ab	ANME-2c	MBGB	MBGD	DHVEG-6	H' (TRFLP)	H' (iTAG)	*Delta-proteobacteria* (total)	*Desulfo-bulbaceae*	*Desulfo-bacteraceae*	*Epsilon-proteobacteria* (total)	*Gamma-proteobacteria* (total)
Hydrate Ridge	Sediments	S2517	PC8	Mat	0-3	✔	--	2.5	1.8	0.66	0.20	0.41	0.01	0.02	0.02	0.01	0.15	0.11	2.8	3.2	0.21	0.08	0.08	0.29	0.08
		S2518	PC8	Mat	3-6	✔	✔	2.4	1.8	0.71	0.24	0.41	0.02	0.02	0.02	0.01	0.11	0.10	2.8	3.4	0.23	0.08	0.10	0.21	0.07
		S2519	PC8	Mat	6-9	✔	--	2.5	1.6	0.78	0.22	0.48	0.04	0.02	0.02	0.01	0.11	0.05	2.2	3.2	0.31	0.10	0.15	0.18	0.04
		S2520	PC8	Mat	9-12	✔	--	2.1	1.9	0.67	0.12	0.39	0.04	0.04	0.07	0.01	0.19	0.07	2.4	3.0	0.33	0.09	0.19	0.11	0.01
		S2521	PC8	Mat	12-15	✔	--	2.2	1.8	0.72	0.16	0.44	0.04	0.02	0.05	0.01	0.17	0.05	1.7	3.0	0.37	0.12	0.20	0.10	0.01
	Nodules	C2517	PC8	Mat	0-3	✔	--	2.2	1.2	0.90	0.17	0.65	0.04	0.02	0.02	0.01	0.07	0.01	1.7	2.9	0.42	0.19	0.15	0.02	0.01
		C2518	PC8	Mat	3-6	✔	✔	2.2	1.1	0.91	0.08	0.73	0.07	0.01	0.02	0.01	0.06	0.01	1.9	2.8	0.45	0.14	0.23	0.03	0.00
		C2519	PC8	Mat	6-9	✔	--	2.4	1.4	0.87	0.18	0.58	0.04	0.03	0.03	0.01	0.07	0.02	2.0	2.8	0.38	0.10	0.23	0.07	0.00
		C2520	PC8	Mat	9-12	✔	--	2.3	1.8	0.68	0.04	0.44	0.08	0.04	0.08	0.01	0.22	0.05	1.8	2.7	0.40	0.06	0.28	0.01	0.00
		C2521	PC8	Mat	12-15	✔	--	1.7	1.8	0.62	0.04	0.29	0.11	0.03	0.15	0.01	0.31	0.02	1.9	2.7	0.47	0.08	0.34	0.01	0.00
Eel River Basin	Sediments	S2688	PC29	Mat	6-9	✔	✔	2.1	1.7	0.68	0.04	0.06	0.02	0.49	0.07	0.00	0.04	0.20	2.5	3.2	0.38	0.15	0.17	0.14	0.04
		S2689	PC29	Mat	9-12	✔	--	2.0	1.5	0.71	0.00	0.46	0.20	0.01	0.04	0.00	0.09	0.18	1.4	2.2	0.48	0.05	0.40	0.00	0.00
		S2693	PC23	Clam	0-3	✔	--	2.3	1.6	0.35	0.02	0.01	0.01	0.30	0.01	0.02	0.01	0.46	2.6	3.6	0.31	0.10	0.11	0.23	0.14
		S2703	PC20	Peripheral	9-12	✔	--	2.1	1.8	0.51	0.00	0.24	0.18	0.05	0.03	0.00	0.13	0.32	1.5	2.2	0.50	0.02	0.44	0.01	0.00
	Nodules	C2688	PC29	Mat	6-9	✔	✔	2.1	2.0	0.66	0.01	0.06	0.09	0.38	0.13	0.05	0.12	0.09	2.7	3.6	0.36	0.18	0.10	0.02	0.11
		C2689	PC29	Mat	9-12	✔	--	1.7	1.8	0.40	0.01	0.12	0.14	0.02	0.12	0.01	0.24	0.32	1.4	1.7	0.58	0.02	0.54	0.01	0.00
		C2693	PC23	Clam	0-3	✔	--	2.3	1.4	0.49	0.00	0.03	0.46	0.00	0.00	0.00	0.26	0.07	1.5	3.2	0.15	0.03	0.03	0.01	0.51
		C2703	PC20	Peripheral	9-12	✔	--	2.1	2.1	0.49	0.01	0.17	0.09	0.19	0.04	0.00	0.21	0.20	2.2	3.2	0.39	0.19	0.15	0.19	0.03

1.8 FIGURES

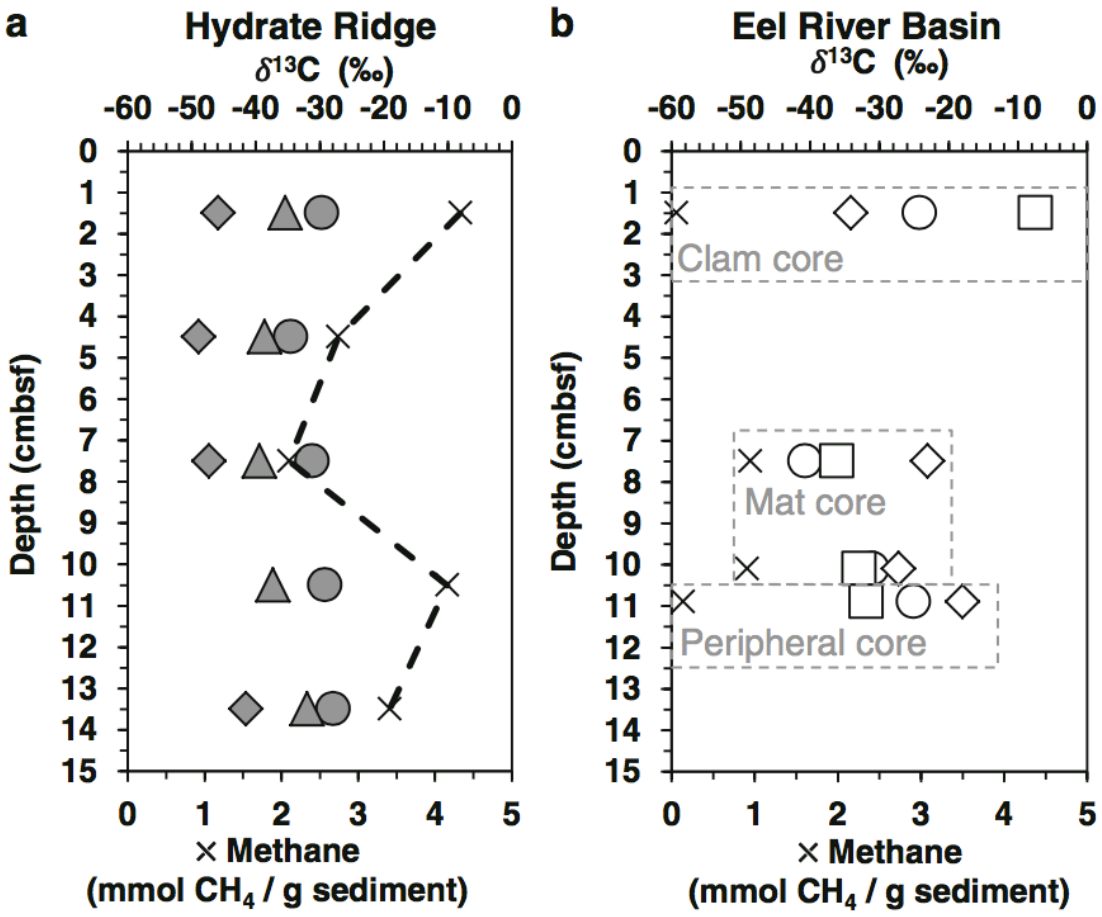


Figure 1. $\delta^{13}C$ variation at HR (a) and ERB (b), for four types of samples: nodule inorganic carbon ($\delta^{13}C_{nod}$; diamonds), organic carbon ($\delta^{13}C_{org}$; circles), sedimentary inorganic carbon ($\delta^{13}C_{sed}$; triangles), and porewater inorganic carbon ($\delta^{13}C_{pw}$; squares). Methane concentrations, shown by 'x', at HR are connected by a dashed line to emphasize that all samples originated from one core (PC8), while ERB samples were obtained from three separate cores (defined by dashed gray boxes). The deep (9–12 cmbsf) samples from the ERB mat core and ERB peripheral core were vertically offset in order to more clearly display the data

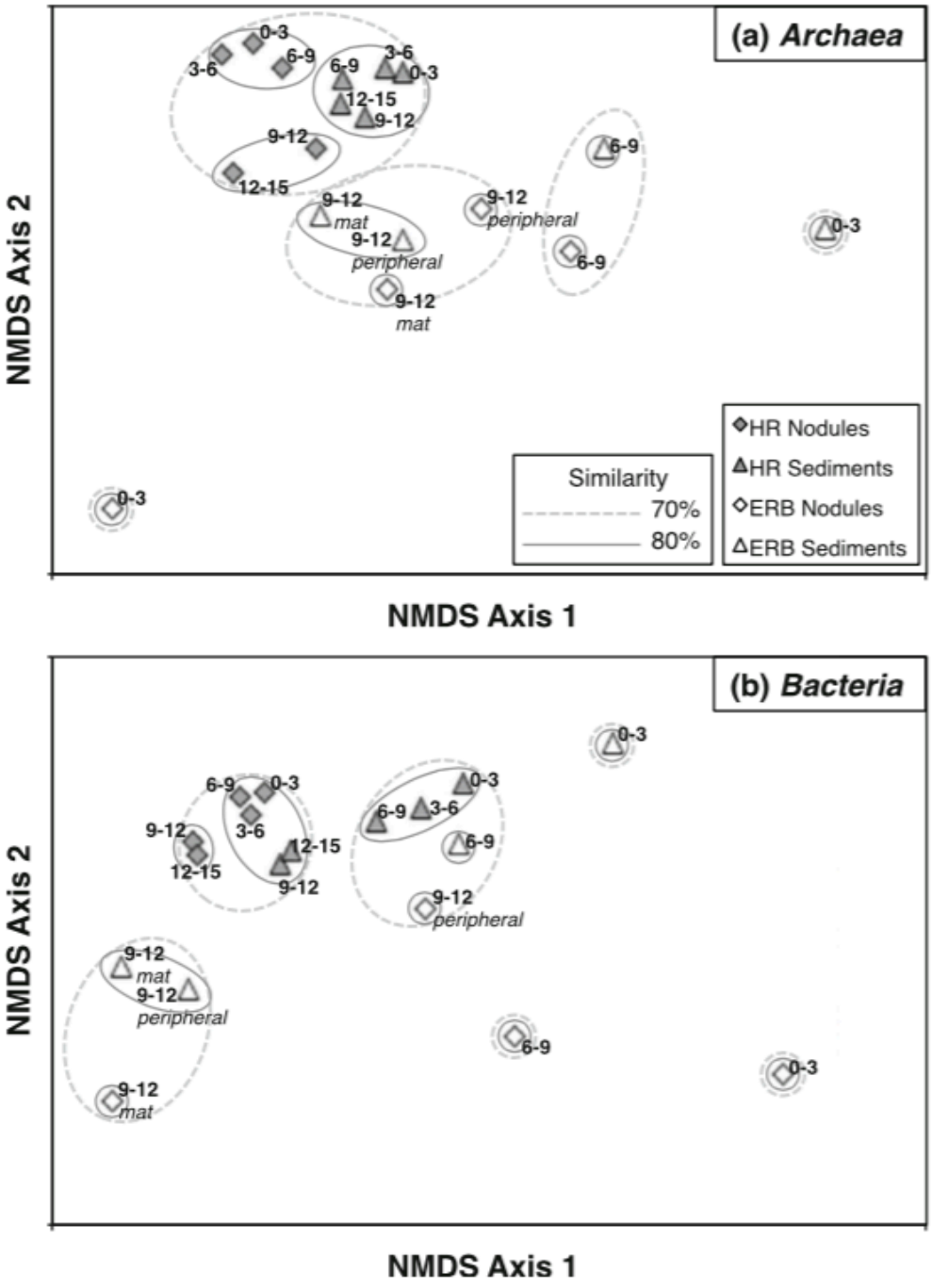


Figure 2. Nonmetric Multidimensional Scaling (NMDS) ordination of iTAG sample similarities for (a) Archaea and (b) Bacteria. Displayed data was square root transformed prior to ordination, which minimizes errors in the ordination due to PCR bias while also not sacrificing genuine differences between samples. Samples with similar microbial communities plot closer together. Archaeal plot stress is 0.07. Bacterial plot stress is 0.05. Legend in (a) applies to both panels. The depth in centimeters below seafloor is listed in bold text next to each sample point. The ERB dataset includes two sediment/nodule pairs from the 9- to 12-cmbsf horizon; for clarity, these are additionally labeled in italics with the relevant core name.

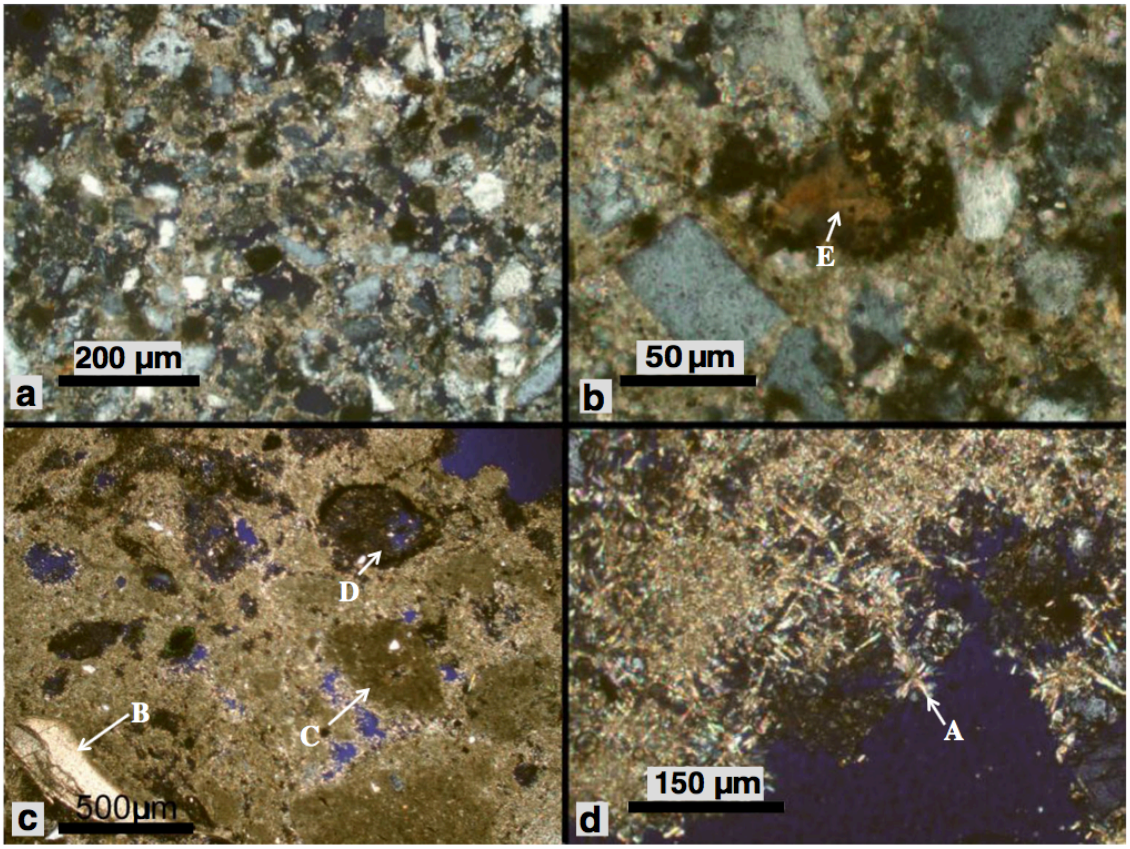


Figure 3. Petrographic thin sections of nodules from (a, b) the 6- to 9- cmbsf horizon on the ERB mat core (PC29), and (c, d) the 3- to 6- cmbsf horizon of the HR mat core (PC8). The ERB sample is quartz- dominated with low internal porosity. Carbonate phases are a mix of calcite, aragonite, and dolomite (see Table 1 in main text). Within an acicular aragonitic matrix (A), the HR sample exhibits cemented bivalve shells (B). Angular carbonate clasts are also observable in the HR sample (C), as well as iron sulfide precipitates (D). Sometimes iron sulfide precipitates surround iron-rich lithic grains (E).

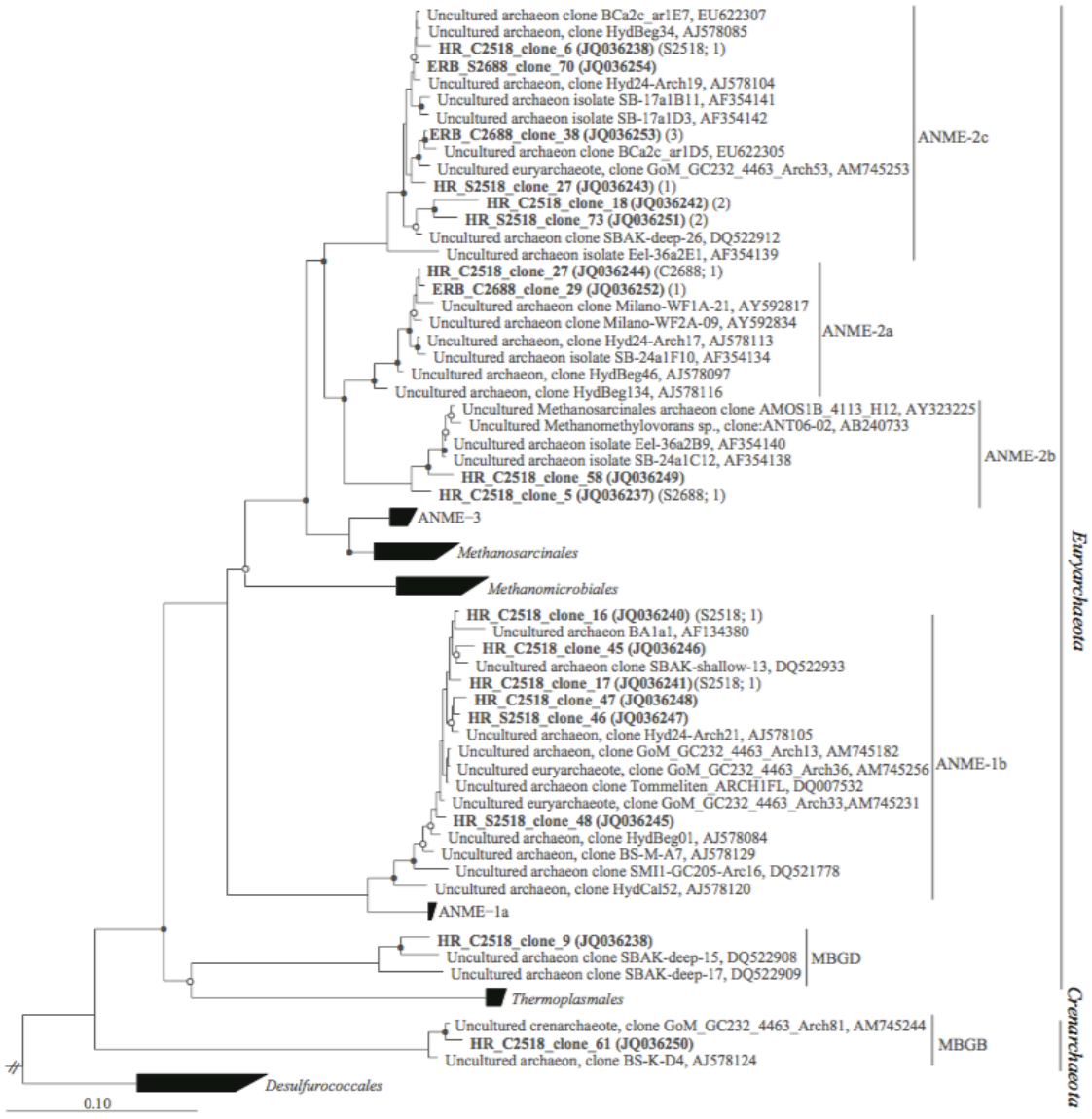


Figure 4. Phylogenetic analysis of archaeal clones. Clones were recovered representing ANME-1b, -2a, -2b, and -2c. ANME-1a clones were not recovered. Samples are named according to the following convention: *Site_Samplenumber_clone_number (accession number)*. Thus, "HR_C2518_clone_61 (JQ036250)" represents a full-length 16S sequence originating from a nodule in the 3- to 6-cmbsf horizon of the HR core.

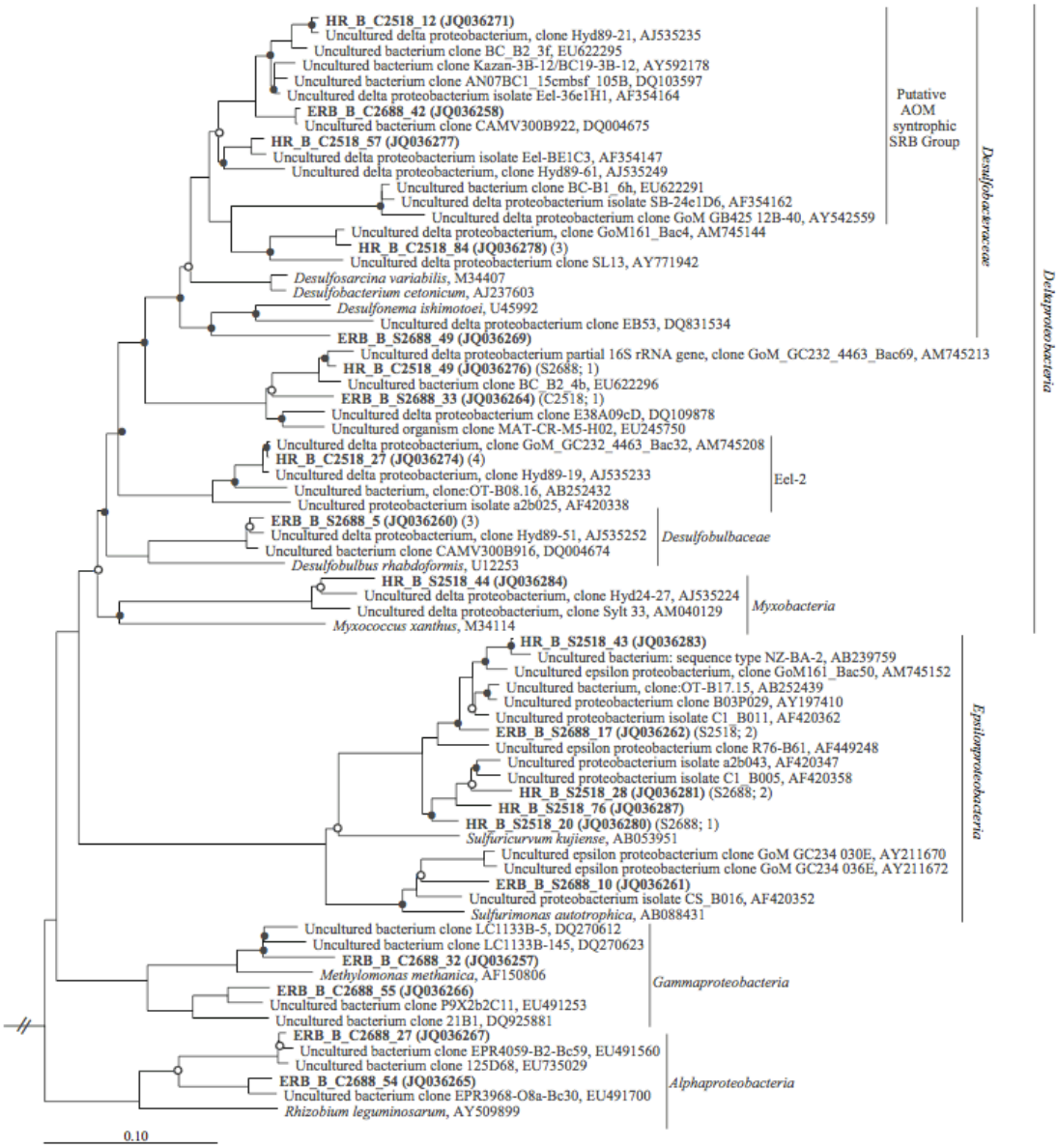


Figure 5. Phylogenetic analysis of bacterial clones. Clones were recovered representing the families *Desulfobulbaceae* and *Desulfobacteraceae*. In addition, alpha-, gamma-, and epsilonproteobacteria clones were recovered. Besides the *Proteobacteria*, clones were recovered representing the *Firmicutes*, *Nitrospirae*, *Bacteroidetes*, *Chloroflexi*, and Candidate Division JS1.

1.9 SUPPLEMENTAL MATERIAL: TEXT

1.9.1 OPTIMIZATION OF NODULE CLEANING TECHNIQUE

To optimize the removal of microbial contamination from carbonate nodules, a series of decontamination experiments were carried out using samples from a carbonate slab from the Eel River basin, broken into several ~10 cm^3 pieces and sterilized by autoclaving. For each experimental condition tested one carbonate section was aseptically maintained, while the second was placed in a 200-mL turbid *Escherichia coli* culture for several hours. Each pair of sterile and contaminated carbonate was then subjected to one of four conditions: 1) UV sterilization for 0.5 hr per side, 2) 70% ethanol rinsing and flaming, 3) rinsing with 1X PBS buffer, and 4) rinsing with 1X PBS and sonication (Branson sonifier 150, Danbury, CT). Each sample was then powdered with a mortar and pestle that was sterilized by baking overnight at 220 °C. Genomic DNA was extracted from 0.5 g of carbonate powder using an Ultraclean Soil DNA kit (MoBio Laboratories, Carlsbad, CA) following the manufacturer's protocol, with a few modifications. Specifically, following addition of the first solution, samples were incubated at 65°C for 5 min, vortexed briefly, and placed at 65°C for 5 min for a second time. After adding the MoBio IRS solution samples were placed at 4°C for 5 min. To determine which sterilization protocol removed exterior contamination, genomic DNA was amplified from both the sterile and *E. coli* contaminated carbonate following the PCR protocol discussed below.

A comparison of the four different treatment protocols indicated that the most effective treatment for removing external DNA and cell contamination (i.e. resulting in no 16S rRNA genes amplified from *E. coli* contaminated sample, or from aseptically maintained control) was to rinse the carbonates with 0.2 µm filtered 1X PBS, followed by sonication at 8 watts for 45 s in fresh, sterile 1X PBS. Samples were centrifuged at 4,000*g* for 5 min. Supernatant was removed and nodules were transferred into fresh 1X PBS between sonication treatments. A total of three

rinse and sonication steps were carried out. All subsequent carbonate nodule and sediment samples were treated according to the protocol discussed above. Genomic DNA was extracted from ERB and HR sediment and 'decontaminated' carbonate samples as described above.

1.9.2 CLONING AND SEQUENCING OF FOUR SELECTED SAMPLES

PCR mixtures (25 μl) contained 0.4 μM each of either archaeal specific primers 8F and 958R (DeLong, 1992), or the bacterial primer 27F with a general 1492R primer. Reactions also contained (final concentrations) 1X 5 Prime *HotMaster Taq Buffer* with 2.5 mM Mg^{2+} (Gaithersburg, MD), 0.2 mM each deoxynucleotide triphosphates, and 0.05 U of 5 Prime HotMaster Taq. PCR reactions were carried out according to the protocol: initial denaturation at 94°C for 3 min, followed by 35 cycles for 45 s at 94°C, 54°C, and 72°C, with a final extension of 72°C for 6 min.

PCR products of the correct length were cut out of a 1% agarose gel. Extracted bands were purified using Qiagen's QIAquick Gel Extraction Kit (Valencia, CA). Purified PCR products were then cloned into a Topo TA cloning kit (Invitrogen, Carlsbad, CA). Clones with the correct insert size were analyzed by restriction fragment length polymorphism (RFLP) using the *HaeIII* restriction enzyme. One representative from each of 35 unique archaeal OTUs was sequenced using a CEQ 8800 Genetic Analysis System from Beckman Coulter (Fullerton, CA). Of these sequenced clones a total of 18 unique, non-chimeric, near full-length archaeal 16S rRNA gene sequences were generated, all of which were 97% or less in similarity. For bacterial libraries, one clone from each of 101 unique OTUs identified by RFLP analysis was sequenced at Laragen, Inc (Los Angeles, CA). Of these sequenced near full-length clones, 35 unique, non-chimeric 16S rRNA phylotypes were recovered, all of which were 97% or less in similarity. For both archaea and bacteria, chimeric sequences were identified with Pintail (Ashelford et al. 2005) and Mallard (Ashelford et al. 2006). Non-chimeric, full-length sequences, including closely related sequences in Genbank and cultured representatives, were aligned using SINA from Silva and imported into ARB (Ludwig et al. 2004). Neighbor-joining trees were constructed using the Olsen distance correction, with 1000 replicates. Maximum likelihood trees were also generated in ARB.

1.10 SUPPLEMENTARY MATERIAL: TABLES

Supplementary Table 1. Archaeal iTAG sequence data. All data is post-processed according to the details given in the methods section.

	Core	Hydrate Ridge										Eel River Basin									
		Sediments				Nodules						Sediments				Nodules					
	Benthic Ecosystem	PC8 Mat	PC8 Rift	PC8 Rift	PC8 Mat	PC8 Mat	PC8 Mat	PC8 Rift	PC8 Rift	PC8 Mat	PC8 Mat	PC29 Mat	PC29 Rift	PC23 Clam	PC26 Peripheral	PC29 Mat	PC29 Rift	PC23 Clam	PC26 Peripheral
	Depth Cmbsf (m)	0.23	24.5	24.5	0.23	0.23	24.5	24.5	0.23	24.5	24.5	0.23	24.5	0.23	24.5	0.23	24.5	0.23	24.5
	Total Archaeal Sequences	4582	9053	8915	8817	12836	17503	9112	12449	7577	8361	7133	12059	756	12043	1829	3748	202	4446
Archaea;__Ancient_Archaeal_Group(AAG);__c;__o;__f		0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00
Archaea;__Euryarchaeota;__Halobacteria;__Halobacteriales;__Deep_Sea_Euryarchaeotic_Group(DSEG)		0.01	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.00	0.00	0.01	0.01	0.00	0.01
Archaea;__Euryarchaeota;__Halobacteria;__Halobacteriales;__Deep_Sea_Hydrothermal_Vent_Gp_6(DHVEG-6)		0.11	0.10	0.05	0.07	0.05	0.01	0.01	0.02	0.05	0.02	0.20	0.18	0.46	0.32	0.09	0.32	0.07	0.30
Archaea;__Euryarchaeota;__Halobacteria;__Halobacteriales;__Marine_Hydrothermal_Vent_Group(MHVG)		0.02	0.01	0.01	0.01	0.01	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.02	0.01	0.00	0.02
Archaea;__Euryarchaeota;__Halobacteria;__Halobacteriales;__Miscellaneous_Euryarchaeotic_Group(MEG)		0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Archaea;__Euryarchaeota;__Halobacteria;__Halobacteriales;__SM1K20		0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Archaea;__Euryarchaeota;__Methanomicrobia;__ANME-1;Other		0.01	0.02	0.04	0.04	0.04	0.07	0.04	0.08	0.11	0.02	0.20	0.01	0.18	0.09	0.14	0.46	0.09	0.09
Archaea;__Euryarchaeota;__Methanomicrobia;__ANME-1;__ANME-1a		0.20	0.24	0.22	0.12	0.16	0.17	0.08	0.18	0.04	0.04	0.04	0.00	0.02	0.00	0.01	0.01	0.01	0.01
Archaea;__Euryarchaeota;__Methanomicrobia;__ANME-1;__ANME-1b		0.41	0.41	0.48	0.39	0.44	0.65	0.73	0.58	0.44	0.28	0.06	0.46	0.01	0.24	0.06	0.12	0.03	0.17
Archaea;__Euryarchaeota;__Methanomicrobia;__Methanocellales;Other		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Archaea;__Euryarchaeota;__Methanomicrobia;__Methanocellales;__RS-K-E8		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Archaea;__Euryarchaeota;__Methanomicrobia;__Methanosarcinales;__ANME-2a-2b		0.02	0.02	0.02	0.07	0.04	0.02	0.02	0.03	0.08	0.15	0.07	0.04	0.01	0.03	0.13	0.12	0.00	0.04
Archaea;__Euryarchaeota;__Methanomicrobia;__Methanosarcinales;__ANME-2c		0.02	0.02	0.02	0.07	0.05	0.02	0.02	0.03	0.08	0.15	0.07	0.04	0.01	0.03	0.13	0.12	0.00	0.04
Archaea;__Euryarchaeota;__Methanomicrobia;__Methanosarcinales;__GOM_Arc_I		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Archaea;__Euryarchaeota;__Methanomicrobia;__Methanosarcinales;__Methanosaetaceae		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Archaea;__Euryarchaeota;__Methanomicrobia;__Methanosarcinales;__Methanosarcinaceae		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.01
Archaea;__Euryarchaeota;__Thermoplasmatales;__1dc-33;__f		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Archaea;__Euryarchaeota;__Thermoplasmatales;__20a-8;__f		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Archaea;__Euryarchaeota;__Thermoplasmatales;__Fe-Aa;__f		0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Archaea;__Euryarchaeota;__Thermoplasmatales;__Kazan-3A-21;__f		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Archaea;__Euryarchaeota;__Thermoplasmatales;__South_African_Goldmine_Gp(SAGMEG);__f		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Archaea;__Euryarchaeota;__Thermoplasmatales;__Thermoplasmatales;Other		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
Archaea;__Euryarchaeota;__Thermoplasmatales;__Thermoplasmatales;__20c-4		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Archaea;__Euryarchaeota;__Thermoplasmatales;__Thermoplasmatales;__ANT61A-H113-D04		0.01	0.01	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Archaea;__Euryarchaeota;__Thermoplasmatales;__ANT61H-05		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Archaea;__Euryarchaeota;__Thermoplasmatales;__ASC21		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00
Archaea;__Euryarchaeota;__Thermoplasmatales;__CCA47		0.02	0.02	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.04	0.00	0.01	0.01	0.00	0.02
Archaea;__Euryarchaeota;__Thermoplasmatales;__MKC5T43		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Archaea;__Euryarchaeota;__Thermoplasmatales;__Marine_Benthic_Group_D_and_DHVEG-1		0.15	0.11	0.11	0.19	0.17	0.07	0.06	0.07	0.22	0.31	0.04	0.09	0.01	0.13	0.12	0.24	0.26	0.21
Archaea;__Euryarchaeota;__Thermoplasmatales;__Marine_Group_II		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Archaea;__Euryarchaeota;__Thermoplasmatales;__Marine_Group_III		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Archaea;__Euryarchaeota;__Thermoplasmatales;__Terrestrial_Miscellaneous_Gp(TMEG)		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Archaea;__Euryarchaeota;__Thermoplasmatales;__Thermoplasmataceae		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Archaea;__Euryarchaeota;__Thermoplasmatales;__VC21_Arc6		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
Archaea;__Thaumarchaeota;__Group_C3;__o;__f		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Archaea;__Thaumarchaeota;__Marine_Benthic_Group_B;__o;__f		0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Archaea;__Thaumarchaeota;__Marine_Group_I;Other;Other		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.16	0.00
Archaea;__Thaumarchaeota;__Marine_Group_I;__Order_Incertae_Sedis;__Family_Incertae_Sedis		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Archaea;__Thaumarchaeota;__Marine_Group_I;__o;__f		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Archaea;__Thaumarchaeota;__Miscellaneous_Crenarchaeotic_Group;__o;__f		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.02	0.01	0.00	0.00
Archaea;__p;__c;__o;__f		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Supplementary Table 2-1. Bacterial iTAG sequence data, page 1 of 4. All data is post-processed according to the details given in the methods section.

	Hydrate Ridge										Eel River Basin							
	Sediments					Nodules					Sediments				Nodules			
Core	PC8	PC8	PC8	PC8	PC8	PC8	PC8	PC8	PC8	PC8	PC29	PC29	PC23	PC20	PC29	PC29	PC23	PC20
Benthic Ecosystem	Mat	Mat	Mat	Mat	Mat	Mat	Mat	Mat	Mat	Mat	Mat	Mat	Clam	Peripheral	Mat	Mat	Clam	Peripheral
Depth (cmbsf)	0-3	3-6	6-9	9-12	12-15	0-3	3-6	6-9	9-12	12-15	6-9	9-12	0-3	12-15	6-9	9-12	0-3	12-15
Total Bacterial Sequences	32773	37437	29300	23504	34074	41208	20864	33306	22835	42582	34821	24117	38928	31590	33496	36917	34752	42684
Bacteria;Other;Other;Other;Other	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Acidobacteria;Other;Other;Other	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Acidobacteria;__Acidobacteria;Other;Other	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Acidobacteria;__Acidobacteria;__AT-23-28;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Acidobacteria;__Acidobacteria;__Subgroup_13;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Acidobacteria;__Acidobacteria;__Subgroup_17;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Acidobacteria;__Acidobacteria;__Subgroup_21;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Acidobacteria;__Acidobacteria;__Subgroup_25;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Acidobacteria;__Acidobacteria;__Subgroup_3;Other	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Acidobacteria;__Acidobacteria;__Subgroup_3;__PAUC26f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Acidobacteria;__Acidobacteria;__Subgroup_9;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Acidobacteria;__Acidobacteria;__d142;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Acidobacteria;__Holophagae;__Acanthopleuribacterales;__Acanthopleuribacteraceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Acidobacteria;__Holophagae;__Subgroup_10;Other	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Acidobacteria;__Holophagae;__Subgroup_10;__CA002	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Acidobacteria;__Holophagae;__Subgroup_10;__Sva0725	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Acidobacteria;__Holophagae;__Subgroup_23;__NKB17	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.01	0.00	0.02	0.00	0.00	0.00
Bacteria;__Acidobacteria;__Holophagae;__Subgroup_7;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Acidobacteria;__Holophagae;__TPD-58;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Acidobacteria;__Subgroup_22;__o;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
Bacteria;__Acidobacteria;__Subgroup_26;__o;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Actinobacteria;Other;Other;Other	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Actinobacteria;__Acidimicrobiia;__Acidimicrobiales;Other	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Actinobacteria;__Acidimicrobiia;__Acidimicrobiales;__Acidimicrobiaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Actinobacteria;__Acidimicrobiia;__Acidimicrobiales;__Family_Incertae_Sedis	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Actinobacteria;__Acidimicrobiia;__Acidimicrobiales;__Iamiaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Actinobacteria;__Acidimicrobiia;__Acidimicrobiales;__OCS155_marine_group	0.03	0.03	0.02	0.01	0.01	0.00	0.00	0.01	0.00	0.00	0.01	0.00	0.02	0.00	0.03	0.00	0.00	0.01
Bacteria;__Actinobacteria;__Acidimicrobiia;__Acidimicrobiales;__Sva0996_marine_group	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Actinobacteria;__Acidimicrobiia;__Acidimicrobiales;__uncultured	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
Bacteria;__Actinobacteria;__Actinobacteria;Other;Other	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Actinobacteria;__Actinobacteria;__Corynebacteriales;__Corynebacteriaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Actinobacteria;__Actinobacteria;__Frankiales;__Geodermatophilaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Actinobacteria;__Actinobacteria;__Micrococcales;__Micrococcaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Actinobacteria;__Actinobacteria;__Streptomycetales;__Streptomycetaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Actinobacteria;__Coriobacteriia;__Coriobacteriales;__Coriobacteriaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Actinobacteria;__KB1-LJY01;__o;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Actinobacteria;__OPB41;__o;__f	0.00	0.00	0.00	0.01	0.00	0.01	0.01	0.01	0.02	0.01	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
Bacteria;__Actinobacteria;__Thermoleophilia;__Gaiellales;__Gaiellaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Actinobacteria;__Thermoleophilia;__Gaiellales;__uncultured	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Actinobacteria;__Thermoleophilia;__Solirubrobacterales;Other	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Actinobacteria;__Thermoleophilia;__Solirubrobacterales;__0319-6M6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Actinobacteria;__Thermoleophilia;__Solirubrobacterales;__288-2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Actinobacteria;__Thermoleophilia;__Solirubrobacterales;__480-2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Armatimonadetes;__c;__o;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__BD1-5;__c;__o;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__BHI80-139;__c;__o;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Bacteroidetes;Other;Other;Other	0.02	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.01	0.01	0.00	0.01	0.00	0.00	0.01
Bacteria;__Bacteroidetes;__BD2-2;__o;__f	0.02	0.02	0.02	0.02	0.02	0.03	0.02	0.03	0.04	0.03	0.05	0.03	0.02	0.02	0.02	0.01	0.00	0.02
Bacteria;__Bacteroidetes;__Bacteroidia;__Bacteroidales;__Bacteroidaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Bacteroidetes;__Bacteroidia;__Bacteroidales;__Marinilabiaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Bacteroidetes;__Bacteroidia;__Bacteroidales;__Prevotellaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Bacteroidetes;__Cytophagia;__Cytophagales;__Flammeovirgaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Bacteroidetes;__Cytophagia;__Order_II_Incertae_Sedis;__Rhodothermaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Bacteroidetes;__Flavobacteriia;__Flavobacteriales;__Cryomorphaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Bacteroidetes;__Flavobacteriia;__Flavobacteriales;__Flavobacteriaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.01
Bacteria;__Bacteroidetes;__SB-1;__o;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Bacteroidetes;__SB-5;__o;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Bacteroidetes;__Sphingobacteriia;__Sphingobacteriales;Other	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Bacteroidetes;__Sphingobacteriia;__Sphingobacteriales;__Chitinophagaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Bacteroidetes;__Sphingobacteriia;__Sphingobacteriales;__E6aC02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Bacteroidetes;__Sphingobacteriia;__Sphingobacteriales;__NS11-12_marine_group	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Bacteroidetes;__Sphingobacteriia;__Sphingobacteriales;__PHOS-HE51	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Bacteroidetes;__Sphingobacteriia;__Sphingobacteriales;__ST-12K33	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Bacteroidetes;__Sphingobacteriia;__Sphingobacteriales;__Saprospiraceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Bacteroidetes;__Sphingobacteriia;__Sphingobacteriales;__Sphingobacteriaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Bacteroidetes;__Sphingobacteriia;__Sphingobacteriales;__WCHB1-69	0.01	0.02	0.01	0.01	0.02	0.04	0.03	0.02	0.03	0.02	0.03	0.00	0.01	0.01	0.00	0.00	0.00	0.00
Bacteria;__Bacteroidetes;__VC2.1_Bac22;__o;__f	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Bacteroidetes;__vadinHA17;__o;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Caldiserica;__Caldisericia;__Caldisericales;__Caldisericaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Caldiserica;__Caldisericia;__Caldisericales;__LF045	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Candidate_division_BRC1;__c;__o;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Candidate_division_JS1;__c;__o;__f	0.03	0.06	0.10	0.19	0.16	0.15	0.19	0.18	0.21	0.15	0.01	0.24	0.00	0.23	0.02	0.25	0.06	0.05
Bacteria;__Candidate_division_OD1;__c;__o;__f	0.00	0.01	0.01	0.01	0.01	0.00	0.00	0.00	0.01	0.01	0.01	0.02	0.00	0.03	0.00	0.01	0.00	0.01
Bacteria;__Candidate_division_OP11;__c;__o;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Candidate_division_OP3;__c;__o;__f	0.01	0.02	0.02	0.02	0.02	0.00	0.00	0.00	0.01	0.00	0.01	0.07	0.00	0.05	0.00	0.03	0.00	0.01
Bacteria;__Candidate_division_OP8;__c;__o;__f	0.00	0.01	0.01	0.01	0.01	0.02	0.02	0.03	0.02	0.02	0.01	0.00	0.00	0.00	0.01	0.01	0.00	0.01
Bacteria;__Candidate_division_SR1;__c;__o;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Candidate_division_WS3;__c;__o;__f	0.01	0.01	0.01	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.01
Bacteria;__Chlamydiae;__Chlamydiia;__Chlamydiales;Other	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Chlamydiae;__Chlamydiae;__Chlamydiales;__Chlamydiaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Chlamydiae;__Chlamydiae;__Chlamydiales;__Family_Incertae_Sedis	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Chlamydiae;__Chlamydiae;__Chlamydiales;__Parachlamydiaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Chlamydiae;__Chlamydiae;__Chlamydiales;__Simkaniaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Chlamydiae;__Chlamydiae;__Chlamydiales;__cvE6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Chloroflexi;__Chloroflexia;__Chloroflexiales;__OPB56	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Chloroflexi;__Chloroflexia;__Chloroflexiales;__SJA-28	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Chloroflexi;__Ignavibacteria;__Ignavibacteriales;Other	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Chloroflexi;__Ignavibacteria;__Ignavibacteriales;__BSV26	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Chloroflexi;__Ignavibacteria;__Ignavibacteriales;__Ignavibacteriaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Chloroflexi;__Ignavibacteria;__Ignavibacteriales;__InsB3-7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Chloroflexi;__Ignavibacteria;__Ignavibacteriales;__PHOS-HE36	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Chloroflexi;Other;Other;Other	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Chloroflexi;__Anaerolineae;__Anaerolineales;__Anaerolineaceae	0.06	0.07	0.07	0.07	0.08	0.09	0.08	0.09	0.08	0.09	0.11	0.01	0.04	0.02	0.07	0.01	0.00	0.07
Bacteria;__Chloroflexi;__Caldilineae;__Caldilineales;__Caldilineaceae	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Chloroflexi;__Dehalococcoidia;Other;Other	0.00	0.00	0.00	0.02	0.01	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00
Bacteria;__Chloroflexi;__Dehalococcoidia;__FS117-23B-02;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Chloroflexi;__Dehalococcoidia;__FW22;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Supplementary Table 2-2. Bacterial iTAG sequence data, page 2 of 4. All data is post-processed according to the details given in the methods section.

Bacteria;__Chloroflexi;__Dehalococcoidia;__GIF3;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Chloroflexi;__Dehalococcoidia;__GIF9;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Chloroflexi;__Dehalococcoidia;__MSBL5;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Chloroflexi;__Dehalococcoidia;__Napoli-4B-65;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Chloroflexi;__Dehalococcoidia;__Sh765B-AG-111;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Chloroflexi;__Dehalococcoidia;__vadinBA26;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Chloroflexi;__JG30-KF-CM66;__o:__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Chloroflexi;__KD4-96;__o:__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
Bacteria;__Chloroflexi;__S085;__o:__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Chloroflexi;__SAR202_clade;__o:__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Chloroflexi;__TK10;__o:__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Chloroflexi;__Thermomicrobia;__AKYG1722;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Cyanobacteria;__4C0d-2;__o:__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Cyanobacteria;__Chloroplast;__o:__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Cyanobacteria;__Cyanobacteria;__Subsection0;__Family0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Cyanobacteria;__ML633J-21;__o:__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Cyanobacteria;__SHA-109;__o:__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Deferribacteres;__Deferribacteres;__Deferribacteres;__Family_Incertae_Sedis	0.01	0.01	0.01	0.01	0.01	0.03	0.01	0.01	0.01	0.01	0.01	0.00	0.01	0.00	0.01	0.00	0.00	0.01	0.00
Bacteria;__Deferribacteres;__Deferribacteres;__Deferribacteriales;__LCP-89	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Deferribacteres;__Deferribacteres;__Deferribacteriales;__PAUC34f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Deferribacteres;__Deferribacteres;__Deferribacteriales;__SAR406_clade(Marine_group_A)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
Bacteria;__Deinococcus-Thermus;__Deinococci;__KD3-62;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Elusimicrobia;__Elusimicrobia;Other;Other	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Elusimicrobia;__Elusimicrobia;__Lineage_IIa;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Elusimicrobia;__Elusimicrobia;__Lineage_IIb;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Elusimicrobia;__Elusimicrobia;__Lineage_IIc;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Elusimicrobia;__Elusimicrobia;__Lineage_IV;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Elusimicrobia;__Elusimicrobia;__Lineage_I_(Endomicrobia);__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Elusimicrobia;__Elusimicrobia;__MD2894-B20;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Elusimicrobia;__Elusimicrobia;__MVP-88;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Elusimicrobia;__Elusimicrobia;__Rs-M47;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Elusimicrobia;__Elusimicrobia;__o:__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Fibrobacteres;__Fibrobacteria;__Fibrobacterales;Other	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Fibrobacteres;__Fibrobacteria;__Fibrobacterales;__Fibrobacteraceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Fibrobacteres;__Fibrobacteria;__Fibrobacterales;__MAT-CR-H6-H10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Fibrobacteres;__Fibrobacteria;__Order_Incertae_Sedis;__Family_Incertae_Sedis	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Fibrobacteres;__Fibrobacteria;__possible_order_07;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Firmicutes;__Bacilli;__Bacillales;__Alicyclobacillaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Firmicutes;__Bacilli;__Bacillales;__Bacillaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Firmicutes;__Bacilli;__Bacillales;__Paenibacillaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Firmicutes;__Bacilli;__Bacillales;__Staphylococcaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Firmicutes;__Bacilli;__Lactobacillales;__Carnobacteriaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Firmicutes;__Bacilli;__Lactobacillales;__Enterococcaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Firmicutes;__Clostridia;__Clostridiales;Other	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Firmicutes;__Clostridia;__Clostridiales;__Christensenellaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Firmicutes;__Clostridia;__Clostridiales;__Clostridiaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Firmicutes;__Clostridia;__Clostridiales;__Eubacteriaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Firmicutes;__Clostridia;__Clostridiales;__Family_Incertae_Sedis	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Firmicutes;__Clostridia;__Clostridiales;__Family_XIII_Incertae_Sedis	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Firmicutes;__Clostridia;__Clostridiales;__Family_XII_Incertae_Sedis	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Firmicutes;__Clostridia;__Clostridiales;__Lachnospiraceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Firmicutes;__Clostridia;__Clostridiales;__Peptococcaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Firmicutes;__Clostridia;__Clostridiales;__Peptostreptococcaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Firmicutes;__Clostridia;__Clostridiales;__Ruminococcaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Firmicutes;__Clostridia;__Clostridiales;__Syntrophomonadaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Firmicutes;__Clostridia;__Clostridiales;__vadinBB60	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Firmicutes;__Clostridia;__Halanaerobiales;__SHA2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Fusobacteria;__Fusobacteriia;__Fusobacteriales;__Fusobacteriaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Fusobacteria;__Fusobacteriia;__MSBL6;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.00	0.00	0.00
Bacteria;__GOUTA4;__c:__o:__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Gemmatimonadetes;__Gemmatimonadetes;__BD2-11_terrestrial_group;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Gemmatimonadetes;__Gemmatimonadetes;__Gemmatimonadales;__Gemmatimonadaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Gemmatimonadetes;__Gemmatimonadetes;__PAUC43f_marine_benthic_group;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Hyd24-12;__c:__o:__f	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
Bacteria;__JL-ETNP-Z39;__c:__o:__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Kazan-3B-28;__c:__o:__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Lentisphaerae;__Lentisphaeria;Other;Other	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
Bacteria;__Lentisphaerae;__Lentisphaeria;__113B434;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Lentisphaerae;__Lentisphaeria;__BS5;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Lentisphaerae;__Lentisphaeria;__LD1-PB3;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Lentisphaerae;__Lentisphaeria;__Lentisphaerales;__Lentisphaeraceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Lentisphaerae;__Lentisphaeria;__MSBL3;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Lentisphaerae;__Lentisphaeria;__R76-B128;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Lentisphaerae;__Lentisphaeria;__RFP12_gut_group;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Lentisphaerae;__Lentisphaeria;__Victivallales;__Victivallaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
Bacteria;__Lentisphaerae;__Lentisphaeria;__WCHB1-41;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__MVP-21;__c:__o:__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__NPL-UPA2;__c:__o:__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Nitrospirae;__Nitrospira;__Nitrospirales;__40296	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.01	0.00	0.00
Bacteria;__Nitrospirae;__Nitrospira;__Nitrospirales;__Nitrospiraceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.03	0.00	0.00
Bacteria;__Planctomycetes;__028H05-P-BN-P5;__o:__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Planctomycetes;__BD7-11;__o:__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Planctomycetes;__MBMPE71;__o:__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Planctomycetes;__MD2896-B258;__o:__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Planctomycetes;__OM190;__o:__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
Bacteria;__Planctomycetes;__Phycisphaerae;Other;Other	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Planctomycetes;__Phycisphaerae;__062D293_hypersaline_microbial_mat_group;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Planctomycetes;__Phycisphaerae;__Amsterdam-1B-07;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Planctomycetes;__Phycisphaerae;__C86;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Planctomycetes;__Phycisphaerae;__CCM11a;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Planctomycetes;__Phycisphaerae;__ML-A-10;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Planctomycetes;__Phycisphaerae;__MSB-3A7_sediment_group;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Planctomycetes;__Phycisphaerae;__MSBL9;__f	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Planctomycetes;__Phycisphaerae;__Phycisphaerales;Other	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Planctomycetes;__Phycisphaerae;__Phycisphaerales;__0BD2Z94_hypersaline_microbial_mat_group	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Planctomycetes;__Phycisphaerae;__Phycisphaerales;__AKAU3564_sediment_group	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Planctomycetes;__Phycisphaerae;__Phycisphaerales;__ODP1230B30.02_sediment_group	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Planctomycetes;__Phycisphaerae;__Phycisphaerales;__Phycisphaeraceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Planctomycetes;__Phycisphaerae;__Phycisphaerales;__SBY2-984	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Planctomycetes;__Phycisphaerae;__Pla1_lineage;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Planctomycetes;__Phycisphaerae;__SHA-43;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Planctomycetes;__Phycisphaerae;__mle1-8;__f	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Planctomycetes;__Pla3_lineage;__o:__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

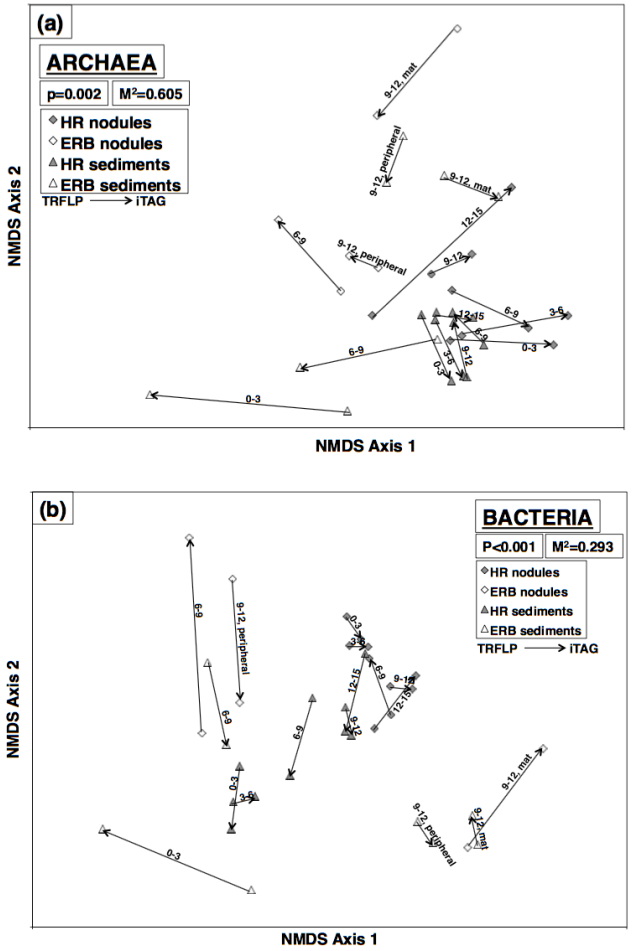
Supplementary Table 2-3. Bacterial iTAG sequence data, page 3 of 4. All data is post-processed according to the details given in the methods section.

Bacteria;__Planctomycetes;__Pla4_lineage;__o;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Planctomycetes;__Planctomycetacia;__Brocadiales;__Brocadiaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00
Bacteria;__Planctomycetes;__Planctomycetacia;__Planctomycetales;__Planctomycetaceae	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.00	0.01	0.00	0.00
Bacteria;__Planctomycetes;__SGST604;__o;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Planctomycetes;__vadinHA49;__o;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;Other;Other;Other	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.01
Bacteria;__Proteobacteria;__AEGEAN-245;__o;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Alphaproteobacteria;Other;Other	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.03	0.01
Bacteria;__Proteobacteria;__Alphaproteobacteria;__Caulobacterales;__Hyphomonadaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Alphaproteobacteria;__DB1-14;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Alphaproteobacteria;__OCS116_clade;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Alphaproteobacteria;__Parvularculales;__Parvularculaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Alphaproteobacteria;__Rhizobiales;Other	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00
Bacteria;__Proteobacteria;__Alphaproteobacteria;__Rhizobiales;__Bradyrhizobiaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Alphaproteobacteria;__Rhizobiales;__Brucellaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Alphaproteobacteria;__Rhizobiales;__Hyphomicrobiaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Alphaproteobacteria;__Rhizobiales;__Phyllobacteriaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.00	0.01	0.00
Bacteria;__Proteobacteria;__Alphaproteobacteria;__Rhizobiales;__Rhizobiaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Alphaproteobacteria;__Rhizobiales;__Rhodobiaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Alphaproteobacteria;__Rhizobiales;__Xanthobacteraceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Alphaproteobacteria;__Rhizobiales;__uncultured	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Alphaproteobacteria;__Rhodobacterales;__Rhodobacteraceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.00	0.01	0.02	0.00
Bacteria;__Proteobacteria;__Alphaproteobacteria;__Rhodospirillales;Other	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Alphaproteobacteria;__Rhodospirillales;__AT-s3-44	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Alphaproteobacteria;__Rhodospirillales;__Acetobacteraceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Alphaproteobacteria;__Rhodospirillales;__Rhodospirillaceae	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.04	0.00	0.00
Bacteria;__Proteobacteria;__Alphaproteobacteria;__Rickettsiales;Other	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Alphaproteobacteria;__Rickettsiales;__Family_Incertae_Sedis	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Alphaproteobacteria;__Rickettsiales;__Rickettsiaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00
Bacteria;__Proteobacteria;__Alphaproteobacteria;__Rickettsiales;__S25-593	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Alphaproteobacteria;__Rickettsiales;__TK34	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.04	0.00	0.00
Bacteria;__Proteobacteria;__Alphaproteobacteria;__Rickettsiales;__mitochondria	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Alphaproteobacteria;__SB1-18;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Alphaproteobacteria;__Sphingomonadales;__Erythrobacteraceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Alphaproteobacteria;__Sphingomonadales;__Sphingomonadaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Betaproteobacteria;__Burkholderiales;__Burkholderiaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Betaproteobacteria;__Burkholderiales;__Comamonadaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Betaproteobacteria;__Burkholderiales;__Oxalobacteraceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00
Bacteria;__Proteobacteria;__Betaproteobacteria;__Hydrogenophilales;__Hydrogenophilaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Betaproteobacteria;__Nitrosomonadales;__Nitrosomonadaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Deltaproteobacteria;Other;Other	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00
Bacteria;__Proteobacteria;__Deltaproteobacteria;__43F-1404R;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Deltaproteobacteria;__Bdellovibrionales;__Bacteriovoracaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Deltaproteobacteria;__Bdellovibrionales;__Bdellovibrionaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Deltaproteobacteria;__DTB120;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.02	0.00
Bacteria;__Proteobacteria;__Deltaproteobacteria;__Desulfarculales;__Desulfarculaceae	0.03	0.03	0.03	0.03	0.04	0.06	0.05	0.03	0.03	0.03	0.03	0.01	0.02	0.01	0.03	0.00	0.00	0.02
Bacteria;__Proteobacteria;__Deltaproteobacteria;__Desulfobacterales;Other	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Deltaproteobacteria;__Desulfobacterales;__Desulfobacteraceae	0.08	0.10	0.15	0.19	0.20	0.15	0.23	0.23	0.28	0.34	0.17	0.40	0.11	0.44	0.10	0.54	0.03	0.15
Bacteria;__Proteobacteria;__Deltaproteobacteria;__Desulfobacterales;__Desulfobulbaceae	0.08	0.08	0.10	0.09	0.12	0.19	0.14	0.10	0.06	0.08	0.15	0.05	0.10	0.02	0.18	0.02	0.03	0.19
Bacteria;__Proteobacteria;__Deltaproteobacteria;__Nitrospinaceae	0.00	0.00	0.01	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Deltaproteobacteria;__Desulfovibrionales;Other	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Deltaproteobacteria;__Desulfovibrionales;__Desulfovibrionaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Deltaproteobacteria;__Desulfurellales;__Desulfurellaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Deltaproteobacteria;__Desulfuromonadales;Other	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Deltaproteobacteria;__Desulfuromonadales;__Desulfuromonadaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Deltaproteobacteria;__Desulfuromonadales;__GR-WP33-58	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Deltaproteobacteria;__Desulfuromonadales;__Geobacteraceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Deltaproteobacteria;__Desulfuromonadales;__Sva1033	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Deltaproteobacteria;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Deltaproteobacteria;__FW113;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Deltaproteobacteria;__GR-WP33-30;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Deltaproteobacteria;__Myxococcales;Other	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Deltaproteobacteria;__Myxococcales;__0319-6G20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Deltaproteobacteria;__Myxococcales;__Haliangiaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00
Bacteria;__Proteobacteria;__Deltaproteobacteria;__Myxococcales;__Nannocystaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Deltaproteobacteria;__Myxococcales;__Phaselicystidaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Deltaproteobacteria;__Myxococcales;__Polyangiaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Deltaproteobacteria;__Myxococcales;__Sandaracinaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.02	0.00	0.00	0.00	0.01	0.00
Bacteria;__Proteobacteria;__Deltaproteobacteria;__Myxococcales;__VHS-B3-70	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Deltaproteobacteria;__uncultured	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01
Bacteria;__Proteobacteria;__Deltaproteobacteria;__SAR324_clade(Marine_group_B);__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00
Bacteria;__Proteobacteria;__Deltaproteobacteria;__Sh765B-TzT-29;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.01	0.00
Bacteria;__Proteobacteria;__Deltaproteobacteria;__Sva0485;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Deltaproteobacteria;__Syntrophobacterales;Other	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Deltaproteobacteria;__Syntrophobacterales;__Syntrophaceae	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Deltaproteobacteria;__Syntrophobacterales;__Syntrophobacteraceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Epsilonproteobacteria;__Campylobacterales;Other	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Epsilonproteobacteria;__Campylobacterales;__Campylobacteraceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.02
Bacteria;__Proteobacteria;__Epsilonproteobacteria;__Campylobacterales;__Helicobacteraceae	0.29	0.21	0.18	0.10	0.10	0.02	0.03	0.07	0.01	0.01	0.13	0.03	0.22	0.01	0.02	0.01	0.01	0.17
Bacteria;__Proteobacteria;__Gammaproteobacteria;__Alteromonadales;__Alteromonadaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Gammaproteobacteria;__Alteromonadales;__Colwelliaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Gammaproteobacteria;__Alteromonadales;__Idiomarinaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Gammaproteobacteria;__Alteromonadales;__Pseudoalteromonadaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Gammaproteobacteria;__Alteromonadales;__Psychromonadaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Gammaproteobacteria;__Alteromonadales;__Shewanellaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Gammaproteobacteria;__BD7-8_marine_group;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Gammaproteobacteria;__CS-B046;__f	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Gammaproteobacteria;__Chromatiales;__Chromatiaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Gammaproteobacteria;__Chromatiales;__Chromatiaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Gammaproteobacteria;__Chromatiales;__Ectothiorhodospiraceae	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Gammaproteobacteria;__E01-9C-26_marine_group;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Gammaproteobacteria;__EC3;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Gammaproteobacteria;__HOC36;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Gammaproteobacteria;__KB5A_clade;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Gammaproteobacteria;__Legionellales;__Coxiellaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Gammaproteobacteria;__Methylococcales;Other	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.05	0.00	0.00
Bacteria;__Proteobacteria;__Gammaproteobacteria;__Methylococcales;__Marine_Methylotrophic_Group_1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Gammaproteobacteria;__Methylococcales;__Marine_Methylotrophic_Group_2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.05	0.00	0.00
Bacteria;__Proteobacteria;__Gammaproteobacteria;__Methylococcales;__Methylococcaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Gammaproteobacteria;__NKB5;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Gammaproteobacteria;__Oceanospirillales;Other	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Gammaproteobacteria;__Oceanospirillales;__Alcanivoracaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Gammaproteobacteria;__Oceanospirillales;__MBAE14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

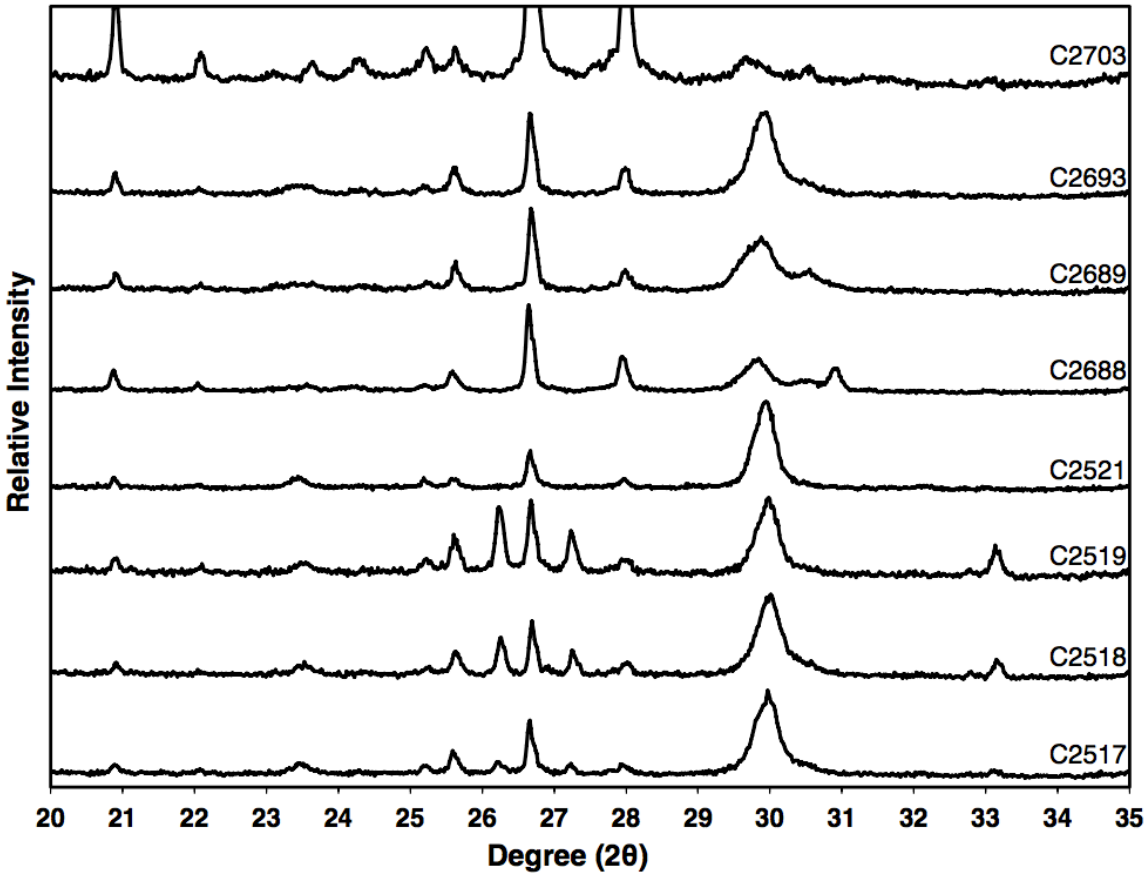
Supplementary Table 2-4. Bacterial iTAG sequence data, page 4 of 4. All data is post-processed according to the details given in the methods section.

Bacteria;__Proteobacteria;__Gammaproteobacteria;__Oceanospirillales;__OM182_clade	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Gammaproteobacteria;__Oceanospirillales;__Oceanospirillaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Gammaproteobacteria;__Oceanospirillales;__SUP05_cluster	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Gammaproteobacteria;__Order_Incertae_Sedis;__Family_Incertae_Sedis	0.01	0.01	0.01	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.01	0.00	0.03	0.00	0.24	0.00	0.28	0.01
Bacteria;__Proteobacteria;__Gammaproteobacteria;__Pseudomonadales;__Moraxellaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Gammaproteobacteria;__Sva0071;__f	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Gammaproteobacteria;__Thiotrichales;__Piscirickettsiaceae	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.00	0.01	0.01
Bacteria;__Proteobacteria;__Gammaproteobacteria;__Thiotrichales;__Thiotrichaceae	0.01	0.02	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.01	0.00	0.07	0.00
Bacteria;__Proteobacteria;__Gammaproteobacteria;__Vibrionales;__Vibrionaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Gammaproteobacteria;__Xanthomonadales;Other	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Gammaproteobacteria;__Xanthomonadales;__JTB255_marine_benthic_group	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.01	0.00	0.03	0.00
Bacteria;__Proteobacteria;__Gammaproteobacteria;__Xanthomonadales;__uncultured	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__JTB23;__o;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Milano-WF1B-44;__o;__f	0.08	0.07	0.04	0.01	0.01	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__TA18;__o;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__Zetaproteobacteria;__Mariprofundales;__Mariprofundaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Proteobacteria;__pltb-vmat-80;__o;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__RF3;__c;__o;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Spirochaetae;__Spirochaetes;__Kazan-3B-09;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Spirochaetae;__Spirochaetes;__MSBL2;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Spirochaetae;__Spirochaetes;__Spirochaetales;__Brevinemataceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Spirochaetae;__Spirochaetes;__Spirochaetales;__LNH41	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Spirochaetae;__Spirochaetes;__Spirochaetales;__Leptospiraceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Spirochaetae;__Spirochaetes;__Spirochaetales;__PL-11B10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Spirochaetae;__Spirochaetes;__Spirochaetales;__Spirochaetaceae	0.01	0.01	0.02	0.02	0.02	0.02	0.01	0.02	0.02	0.01	0.03	0.02	0.01	0.01	0.02	0.02	0.05	0.02
Bacteria;__Spirochaetae;__Spirochaetes;__Spirochaetales;__V2072-189E03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__TA06;__c;__o;__f	0.01	0.01	0.01	0.02	0.02	0.03	0.03	0.04	0.02	0.02	0.01	0.03	0.00	0.02	0.01	0.02	0.00	0.01
Bacteria;__TM6;__c;__o;__f	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.00	0.01	0.01	0.01	0.00	0.00
Bacteria;__Verrucomicrobia;__OPB35_soil_group;__o;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Verrucomicrobia;__Opitutae;Other;Other	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Verrucomicrobia;__Opitutae;__Opitutales;__Opitutaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Verrucomicrobia;__Opitutae;__Puniceicoccales;__Puniceicoccaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Verrucomicrobia;__Verrucomicrobiae;__Verrucomicrobiales;Other	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Verrucomicrobia;__Verrucomicrobiae;__Verrucomicrobiales;__DEV007	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Verrucomicrobia;__Verrucomicrobiae;__Verrucomicrobiales;__Rubritaleaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__Verrucomicrobia;__Verrucomicrobiae;__Verrucomicrobiales;__Verrucomicrobiaceae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteria;__WCHB1-60;__c;__o;__f	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

1.11 SUPPLEMENTAL MATERIAL: FIGURES



Supplementary Figure 1. Procrustes comparison of sample relationships between the iTAG and TRFLP datasets, for (a) archaea and (b) bacteria. iTAG ordination coordinates (see Figure 2 in main text) were used as the reference, against which TRFLP coordinates were transformed by translation, rotation, reflection, and/or scaling to minimize distance to the iTAG datapoints. The purpose was to test the closeness of fit between the two sets of datapoints, and therefore infer whether the biological similarity trends in iTAG and TRFLP data support one another. The analysis was completed 1,000 times and the resulting M^2 and p-values are given on the plot. Lower M^2 values indicate better closeness of fit between the datasets, and range from a possible value of 0 to 1. Lower p-values indicate stronger statistical support as determined through repeated iterations. Among the bacterial dataset, the p- and M^2-values indicate decent closeness of fit between the overall iTAG and TRFLP datasets. Among the archaeal dataset, the M^2 value is not particularly good, but NMDS biological patterns nonetheless appear to be largely maintained. As seen in the plots, some samples clearly fit more closely than others (that is, two points connected by a shorter arrow). Arrows are meant to help the reader connect iTAG and TRFLP ordination points from the same sample, to determine whether iTAG and TRFLP datasets demonstrate similar biological similarity trends. Arrows always point from the TRFLP to the iTAG point, since TRFLP coordinates were transformed to minimize distance to iTAG points. One archaeal sample (ERB clam core 0-3 cmbsf nodule) was excluded from analysis because it was substantially different than all other samples and skewed the ordination beyond possible interpretation.



Supplementary Figure 2. XRD spectra from nodules in this study. Mineralogical assignments were made based on the (104) peak heights of calcite, Mg-calcite and dolomite, and the (111) peak height of aragonite. Nodule C2520 was not measured due to lack of material. For display in this figure, data from C2703 was amplified 5*x* in order to more clearly display the carbonate peaks at circa 30˚ 2θ.

1.12 REFERENCES

Aloisi, G., I. Bouloubassi, S. K. Heijs, R. D. Pancost, C. Pierre, J. S. Sinninghe Damsté, J. C. Gottschal, L. J. Forney, and J.-M. Rouchy. 2002. CH4-consuming microorganisms and the formation of carbonate crusts at cold seeps. Earth and Planetary Science Letters **203**: 195–203.

Aronesty, E. 2011. ea-utils: Command-line tools for processing biological sequencing data.

Ashelford, K. E., N. A. Chuzhanova, J. C. Fry, A. J. Jones, and A. J. Weightman. 2005. At Least 1 in 20 16S rRNA Sequence Records Currently Held in Public Repositories Is Estimated To Contain Substantial Anomalies. Applied and Environmental Microbiology **71**: 7724–7736.

Ashelford, K. E., N. A. Chuzhanova, J. C. Fry, A. J. Jones, and A. J. Weightman. 2006. New Screening Software Shows that Most Recent Large 16S rRNA Gene Clone Libraries Contain Chimeras. Applied and Environmental Microbiology **72**: 5734–5741.

Bailey, J. V., T. D. Raub, A. N. Meckler, B. K. Harrison, T. M. D. Raub, A. M. Green, and V. J. Orphan. 2010. Pseudofossils in relict methane seep carbonates resemble endemic microbial consortia. Palaeogeography, Palaeoclimatology, Palaeoecology **285**: 131–142.

Barry, J. P., H. Gary Greene, D. L. Orange, C. H. Baxter, B. H. Robison, R. E. Kochevar, J. W. Nybakken, D. L. R, and C. M. McHugh. 1996. Biologic and geologic characteristics of cold seeps in Monterey Bay, California. Deep Sea Research Part I: Oceanographic Research Papers **43**: 1739–1762.

Barry, J. P., R. E. Kochevar, and C. H. Baxter. 1997. The influence of pore-water chemistry and physiology on the distribution of vesicomyid clams at cold seeps in Monterey Bay: Implications for patterns of chemosynthetic community organization. Limnology and Oceanography **42**: 318–328.

Bayon, G., S. Dupré, E. Ponzevera, J. Etoubleau, S. Chéron, C. Pierre, J. Mascle, A. Boetius, and G. J. de Lange. 2013. Formation of carbonate chimneys in the Mediterranean Sea linked to deep-water oxygen depletion. Nature Geoscience **6**: 1–6.

Berry, D., K. Ben Mahfoudh, M. Wagner, and A. Loy. 2011. Barcoded Primers Used in Multiplex Amplicon Pyrosequencing Bias Amplification. Applied and Environmental Microbiology **77**: 7846–7849.

Birgel, D., M. Elvert, X. Han, and J. Peckmann. 2008a. ^{13}C-depleted biphytanic diacids as tracers of past anaerobic oxidation of methane. Organic geochemistry **39**: 152–156.

Birgel, D., T. Himmler, A. Freiwald, and J. Peckmann. 2008b. A new constraint on the antiquity of anaerobic oxidation of methane: Late Pennsylvanian seep limestones from southern Namibia. Geology **36**: 543–546.

Boetius, A., and E. Suess. 2004. Hydrate Ridge: a natural laboratory for the study of microbial life fueled by methane from near-surface gas hydrates. Chemical Geology **205**: 291–310.

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

Bohrmann, G., J. Greinert, E. Suess, and M. Torres. 1998. Authigenic carbonates from the Cascadia subduction zone and their relation to gas hydrate stability. Geology **26**: 647.

Campbell, K. A., C. S. Nelson, A. C. Alfaro, S. Boyd, J. Greinert, S. Nyman, E. Grosjean, G. A. Logan, M. R. Gregory, S. Cooke, P. Linke, S. Milloy, and I. Wallis. 2010. Geological imprint of methane seepage on the seabed and biota of the convergent Hikurangi Margin, New Zealand: Box core and grab carbonate results. Marine Geology **272**: 285–306.

Caporaso, J. G., C. L. Lauber, W. A. Walters, D. Berg-Lyons, C. A. Lozupone, P. J. Turnbaugh, N. Fierer, and R. Knight. 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proceedings of the National Academy of Sciences **108**: 4516–4522.

Caporaso, J. G., C. L. Lauber, W. A. Walters, D. Berg-Lyons, J. Huntley, N. Fierer, S. M.

Owens, J. Betley, L. Fraser, M. Bauer, N. Gormley, J. A. Gilbert, G. Smith, and R. Knight. 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. The ISME Journal **6**: 1621–1624.

Caporaso, J. G., J. Kuczynski, J. Stombaugh, K. Bittinger, F. D. Bushman, E. K. Costello, N. Fierer, A. G. Peña, J. K. Goodrich, J. I. Gordon, G. A. Huttley, S. T. Kelley, D. Knights, J. E. Koenig, R. E. Ley, C. A. Lozupone, D. McDonald, B. D. Muegge, M. Pirrung, J. Reeder, J. R. Sevinsky, P. J. Turnbaugh, W. A. Walters, J. Widmann, T. Yatsunenko, J. Zaneveld, and R. Knight. 2010. QIIME allows analysis of high-throughput community sequencing data. Nature Methods **7**: 335–336.

Chen, Z., W. Yan, M. Chen, S. Wang, J. Lu, F. Zhang, R. Xiang, S. Xiao, P. Yan, and S. Gu. 2006. Discovery of seep carbonate nodules as new evidence for gas venting on the northern continental slope of South China Sea. Chinese Science Bulletin **51**: 1228–1237.

Clarke, K. R., and R. M. Warwick. 2001. Change in Marine Communities, 2nd ed. PRIMER-E Ltd.

Edgar, R. C. 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics **26**: 2460–2461.

Edgar, R. C., B. J. Haas, J. C. Clemente, C. Quince, and R. Knight. 2011. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics **27**: 2194–2200.

Fischer, D., H. Sahling, K. Nöthen, G. Bohrmann, and S. Kasten. 2012. Interaction between hydrocarbon seepage, chemosynthetic communities, and bottom water redox at cold seeps of the Makran accretionary prism: insights from habitat-specific pore water sampling and modeling. Biogeosciences **9**: 2013–2031.

Gieskes, J. M., T. Gamo, and H. Brumsack. 1991. Chemical methods for interstitial water analysis aboard JOIDES Resolution. Ocean Drilling Program Texas A&M University.

Gieskes, J., C. Mahn, S. Day, J. Martin, J. Greinert, T. Rathburn, and B. McAdoo. 2005. A study of the chemistry of pore fluids and authigenic carbonates in methane seep environments: Kodiak Trench, Hydrate Ridge, Monterey Bay, and Eel River Basin. Chemical Geology **220**: 329–345.

Gilbert, J. A., F. Meyer, J. Jansson, J. Gordon, N. R. Pace, J. M. Tiedje, R. E. Ley, N. Fierer, D. Field, N. C. Kyrpides, F. O. Gloeckner, H. P. Klenk, K. E. Wommack, E. Glass, K. Docherty, R. Gallery, R. Stevens, and R. Knight. 2011. The Earth Microbiome Project: Meeting report of the "1st EMP meeting on sample selection and acquisition" at Argonne National Laboratory October 6th 2010. 1–5.

Goffredi, S. K., R. Wilpiszeski, R. Lee, and V. J. Orphan. 2008. Temporal evolution of methane cycling and phylogenetic diversity of archaea in sediments from a deep-sea whale-fall in Monterey Canyon, California. The ISME Journal **2**: 204–220.

Green-Saxena, A., A. E. Dekas, N. F. Dalleska, and V. J. Orphan. 2014. Nitrate-based niche differentiation by distinct sulfate-reducing bacteria involved in the anaerobic oxidation of methane. The ISME Journal **8**: 150–163.

Greinert, J., G. Bohrmann, and E. Suess. 2001. Gas Hydrate-Associated Carbonates and Methane-Venting at Hydrate Ridge: Classification, Distribution, and Origin of Authigenic Lithologies, p. 99–113. *In* C.K. Paull and W.P. Dillon [eds.], Natural Gas Hydrates. American Geophysical Union.

Griffiths, R. P., B. A. Caldwell, J. D. Cline, W. A. Broich, and R. Y. Morita. 1982. Field Observations of Methane Concentrations and Oxidation Rates in the Southeastern Bering Sea. Applied and Environmental Microbiology **44**: 435–446.

Guan, H., Y. Sun, X. Zhu, S. Mao, D. Feng, N. Wu, and D. Chen. 2013. Factors controlling the types of microbial consortia in cold-seep environments: A molecular and isotopic investigation of authigenic carbonates from the South China Sea. Chemical Geology **354**: 55–64.

Gulin, S. B., G. G. Polikarpov, and V. N. Egorov. 2003. The age of microbial carbonate

structures grown at methane seeps in the Black Sea with an implication of dating of the seeping methane. Marine Chemistry **84**: 67–72.

Haas, A., J. Peckmann, M. Elvert, H. Sahling, and G. Bohrmann. 2010. Patterns of carbonate authigenesis at the Kouilou pockmarks on the Congo deep-sea fan. Marine Geology **268**: 129–136.

Heijs, S. K., G. Aloisi, I. Bouloubassi, R. D. Pancost, C. Pierre, J. S. Sinninghe Damsté, J. C. Gottschal, J. D. Elsas, and L. J. Forney. 2006. Microbial Community Structure in Three Deep-Sea Carbonate Crusts. Microbial Ecology **52**: 451–462.

Heijs, S. K., J. S. Sinninghe Damsté, and L. J. Forney. 2005. Characterization of a deep-sea microbial mat from an active cold seep at the Milano mud volcano in the Eastern Mediterranean Sea. FEMS Microbiology Ecology **54**: 47–56.

Hoehler, T. M., M. J. Alperin, D. B. Albert, and C. S. Martens. 1994. Field and laboratory studies of methane oxidation in an anoxic marine sediment: Evidence for a methanogen-sulfate reducer consortium. Global Biogeochemical Cycles **8**: 451.

Hovland, M., M. R. Talbot, H. Qvale, S. Olaussen, and L. Aasberg. 1987. Methane-related Carbonate Cements in Pockmarks of the North Sea. Journal of Sedimentary Petrology **57**: 881–892.

Ijiri, A., U. Tsunogai, T. Gamo, F. Nakagawa, T. Sakamoto, and S. Saito. 2009. Enrichment of adsorbed methane in authigenic carbonate concretions of the Japan Trench. Geo-Marine Letters **29**: 301–308.

Inagaki, F., K. Takai, H. Kobayashi, K. H. Nealson, and K. Horikoshi. 2003. Sulfurimonas autotrophica gen. nov., sp. nov., a novel sulfur-oxidizing ε-proteobacterium isolated from hydrothermal sediments in the Mid-Okinawa Trough. Int J Syst Evol Microbiol **53**: 1801–1805.

Inagaki, F., K. Takai, K. H. Nealson, and K. Horikoshi. 2004. Sulfurovum lithotrophicum gen. nov., sp. nov., a novel sulfur-oxidizing chemolithoautotroph within the ε-Proteobacteria isolated from Okinawa Trough hydrothermal sediments. Int J Syst Evol Microbiol **54**: 1477–1482.

Jørgensen, N. O. 1989. Holocene methane-derived, dolomite-cemented sandstone pillars from the Kattegat, Denmark. Marine Geology **88**: 71–81.

Jørgensen, N. O. 1992. Methane-derived carbonate cementation of marine sediments from the Kattegat, Denmark: Geochemical and geological evidence. Marine Geology **103**: 1–13.

Kendall, M. M., G. D. Wardlaw, C. F. Tang, A. S. Bonin, Y. Liu, and D. L. Valentine. 2007. Diversity of Archaea in Marine Sediments from Skan Bay, Alaska, Including Cultivated Methanogens, and Description of Methanogenium boonei sp. nov. Applied and Environmental Microbiology **73**: 407–414.

Knittel, K., T. Lösekann, A. Boetius, R. Kort, and R. Amann. 2005. Diversity and Distribution of Methanotrophic Archaea at Cold Seeps. Applied and Environmental Microbiology **71**: 467–479.

Kodama, Y., and K. Watanabe. 2003. Isolation and Characterization of a Sulfur-Oxidizing Chemolithotroph Growing on Crude Oil under Anaerobic Conditions. Applied and Environmental Microbiology **69**: 107–112.

Krause, S., V. Liebetrau, S. Gorb, M. Sanchez-Roman, J. A. McKenzie, and T. Treude. 2012. Microbial nucleation of Mg-rich dolomite in exopolymeric substances under anoxic modern seawater salinity: New insight into an old enigma. Geology **40**: 587–590.

Kutterolf, S., V. Liebetrau, T. Mörz, A. Freundt, T. Hammerich, and D. Garbe-Schönberg. 2008. Lifetime and cyclicity of fluid venting at forearc mound structures determined by tephrostratigraphy and radiometric dating of authigenic carbonates. Geology **36**: 707–710.

Lein, A. 2004. Authigenic carbonate formation in the ocean. Lithology and Mineral Resources **39**: 1–30.

Levin, L. A. 2005. Ecology of cold seep sediments: Interactions of fauna with flow, chemistry, and

microbes. Oceanography and Marine Biology: An Annual Review **43**: 1–46.

Liebetrau, V., A. Eisenhauer, and P. Linke. 2010. Cold seep carbonates and associated cold-water corals at the Hikurangi Margin, New Zealand: New insights into fluid pathways, growth structures and geochronology. Marine Geology **272**: 307–318.

Lloyd, K. G., D. B. Albert, J. F. Biddle, J. P. Chanton, O. Pizarro, and A. Teske. 2010. Spatial Structure and Activity of Sedimentary Microbial Communities Underlying a Beggiatoa spp. Mat in a Gulf of Mexico Hydrocarbon Seep. PLoS ONE **5**: e8738.

Ludwig, W., O. Strunk, R. Westram, L. Richter, H. Meier, Yadhukumar, A. Buchner, T. Lai, S. Steppi, G. Jobb, W. Förster, I. Brettske, S. Gerber, A. W. Ginhart, O. Gross, S. Grumann, S. Hermann, R. Jost, A. König, T. Liss, R. Lüßmann, M. May, B. Nonhoff, B. Reichel, R. Strehlow, A. Stamatakis, N. Stuckmann, A. Vilbig, M. Lenke, T. Ludwig, A. Bode, and K.-H. Schleifer. 2004. ARB: a software environment for sequence data. Nucl. Acids Res. **32**: 1363–1371.

Luff, R., and K. Wallmann. 2003. Fluid flow, methane fluxes, carbonate precipitation and biogeochemical turnover in gas hydrate-bearing sediments at Hydrate Ridge, Cascadia Margin: numerical modeling and mass balances. Geochimica et Cosmochimica Acta **67**: 3403–3421.

Luff, R., K. Wallmann, and G. Aloisi. 2004. Numerical modeling of carbonate crust formation at cold vent sites: significance for fluid and methane budgets and chemosynthetic biological communities. Earth and Planetary Science Letters **221**: 337–353.

Marlow, J. J., J. A. Steele, W. Ziebis, A. R. Thurber, L. A. Levin, and V. J. Orphan. 2014a. Carbonate-hosted methanotrophy represents an unrecognized methane sink in the deep sea. Nature Communications **5**: 1–12.

Marlow, J., J. A. Steele, D. Case, S. A. Connon, L. A. Levin, and V. J. Orphan. 2014b. Microbial abundance and diversity patterns associated with sediments and carbonates from the methane seep environments of Hydrate Ridge, OR. Frontiers in Marine Science **1**: 1–16.

McCune, B., J. B. Grace, and D. L. Urban. 2002. Analysis of ecological communities, MjM Software Design.

Michaelis, W., R. Seifert, K. Nauhaus, T. Treude, V. Thiel, M. Blumenberg, K. Knittel, A. Gieseke, K. Peterknecht, T. Pape, A. Boetius, R. Amann, B. B. Jørgensen, F. Widdel, J. Peckmann, N. V. Pimenov, and M. B. Gulin. 2002. Microbial Reefs in the Black Sea Fueled by Anaerobic Oxidation of Methane. Science **297**: 1013–1015.

Naehr, T. H., D. Birgel, G. Bohrmann, I. R. MacDonald, and S. Kasten. 2009. Biogeochemical controls on authigenic carbonate formation at the Chapopote "asphalt volcano," Bay of Campeche. Chemical Geology **266**: 390–402.

Naehr, T. H., N. M. Rodriguez, G. Bohrmann, C. K. Paull, and R. Botz. 2000. Methane-derived authigenic carbonates associated with gas hydrate decomposition and fluid venting above the Blake Ridge Diapir. Proceedings of the Proceedings of the Ocean Drilling Program **164**: 285-300.

Naehr, T. H., P. Eichhubl, V. J. Orphan, M. Hovland, C. K. Paull, W. Ussler, T. D. Lorenson, and H. G. Greene. 2007. Authigenic carbonate formation at hydrocarbon seeps in continental margin sediments: a comparative study. Deep Sea Research Part II **54**: 1268–1291.

Niemann, H., P. Linke, K. Knittel, E. MacPherson, A. Boetius, W. Brückmann, G. Larvik, K. Wallmann, U. Schacht, E. Omoregie, D. Hilton, K. Brown, and G. Rehder. 2013. Methane-Carbon Flow into the Benthic Food Web at Cold Seeps – A Case Study from the Costa Rica Subduction Zone. PLoS ONE **8**: e74894.

Nunoura, T., Y. Takaki, H. Kazama, M. Hirai, J. Ashi, H. Imachi, and K. Takai. 2012. Microbial Diversity in Deep-sea Methane Seep Sediments Presented by SSU rRNA Gene Tag Sequencing. Microbes and Environments **27**: 382–390.

Orphan, V. J., Hinrichs, W. Ussler, C. K. Paull, L. T. Taylor, S. P. Sylva, J. M. Hayes, and E. F.

Delong. 2001a. Comparative Analysis of Methane-Oxidizing Archaea and Sulfate-Reducing Bacteria in Anoxic Marine Sediments. Applied and Environmental Microbiology **67**: 1922–1934.

Orphan, V., C. House, Hinrichs, K. McKeegan, and E. DeLong. 2001b. Methane-Consuming Archaea Revealed by Directly Coupled Isotopic and Phylogenetic Analysis. Science **293**: 484–487.

Orphan, V., W. Ussler, T. H. Naehr, C. H. House, Hinrichs, and C. K. Paull. 2004. Geological, geochemical, and microbiological heterogeneity of the seafloor around methane vents in the Eel River Basin, offshore California. Chemical Geology **205**: 265–289.

Paull, C. K., J. P. Chanton, A. C. Neumann, J. A. Coston, C. S. Martens, and W. Showers. 1992. Indicators of Methane-Derived Carbonates and Chemosynthetic Organic Carbon Deposits: Examples from the Florida Escarpment. Palaios **7**: 361.

Peckmann, J., A. Reimer, C. Luth, C. Luth, B. T. Hansen, C. Heinicke, J. Hoefs, and J. Reitner. 2001. Methane-derived carbonates and authigenic pyrite from the northwestern Black Sea. Marine Geology **177**: 129–150.

Peckmann, J., and V. Thiel. 2004. Carbon cycling at ancient methane-seeps. Chemical Geology **205**: 443–467.

Peckmann, J., V. Thiel, W. Michaelis, P. Clari, C. Gaillard, L. Martire, and J. Reitner. 1999. Cold seep deposits of Beauvoisin (Oxfordian; southeastern France) and Marmorito (Miocene; northern Italy): microbially induced authigenic carbonates. Int J Earth Sci (Geol Rundsch) **88**: 60–75.

Pernthaler, A., A. E. Dekas, C. T. Brown, S. K. Goffredi, T. Embaye, and V. J. Orphan. 2008. Diverse syntrophic partnerships from deep-sea methane vents revealed by direct cell capture and metagenomics. Proceedings of the National Academy of Sciences **105**: 7052–7057.

Quast, C., E. Pruesse, P. Yilmaz, J. Gerken, T. Schweer, P. Yarza, J. Peplies, and F. O. Glöckner. 2012. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucl. Acids Res. 1-7.

Reeburgh, W. S. 2007. Oceanic Methane Biogeochemistry. Chemical Reviews **107**: 486–513.

Roalkvam, I., S. L. Jørgensen, Y. Chen, R. Stokke, H. Dahle, W. P. Hocking, A. Lanzén, H. Haflidason, and I. H. Steen. 2011. New insight into stratification of anaerobic methanotrophs in cold seep sediments. FEMS Microbiology Ecology **78**: 233–243.

Sahling, H., D. Rickert, R. W. Lee, P. Linke, and E. Suess. 2002. Macrofaunal community structure and sulfide flux at gas hydrate deposits from the Cascadia convergent margin, NE Pacific. Marine Ecology Progress Series **231**: 121–138.

Salter, S., M. Cox, E. Turek, S. Calus, W. Cookson, M. Moffatt, P. Turner, J. Parkhill, N. Loman, and A. Walker. 2014. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biology **12**: 1–12.

Stadnitskaia, A., D. Nadezhkin, B. Abbas, V. Blinova, M. K. Ivanov, and J. S. Sinninghe Damsté. 2008. Carbonate formation by anaerobic oxidation of methane: Evidence from lipid biomarker and fossil 16S rDNA. Geochimica et Cosmochimica Acta **72**: 1824–1836.

Stadnitskaia, A., G. Muyzer, B. Abbas, M. J. L. Coolen, E. C. Hopmans, M. Baas, T. C. E. van Weering, M. K. Ivanov, E. Poludetkina, and J. S. Sinninghe Damsté. 2005. Biomarker and 16S rDNA evidence for anaerobic oxidation of methane and related carbonate precipitation in deep-sea mud volcanoes of the Sorokin Trough, Black Sea. Marine Geology **217**: 67–96.

Suess, E., M. E. Torres, G. Bohrmann, R. W. Collier, D. Rickert, C. Goldfinger, P. Linke, A. Heuser, H. Sahling, K. Heeschen, C. Jung, K. Nakamura, J. Greinert, O. Pfannkuche, A. Trehu, G. Klinkhammer, M. J. Whiticar, A. Eisenhauer, B. Teichert, and M. Elvert. 2001. Sea Floor Methane Hydrates at Hydrate Ridge, Cascadia Margin, p. 87–98. *In* C.K. Paull and W.P. Dillon [eds.], Natural Gas Hydrates. American Geophysical Union.

Suess, E., M. E. Torres, G. Bohrmann, R. W. Collier, J. Greinert, P. Linke, G. Rehder, A. Trehu, K. Wallmann, G. Winckler, and E. Zuleger. 1999. Gas hydrate destabilization: enhanced

dewatering, benthic material turnover and large methane plumes at the Cascadia convergent margin. Earth and Planetary Science Letters **170**: 1–15.

Takai, K., and K. Horikoshi. 1999. Genetic Diversity of Archaea in Deep-Sea Hydrothermal Vent Environments. Genetics **152**: 1285–1297.

Teichert, B., A. Eisenhauer, G. Bohrmann, A. Haase-Schramm, B. Bock, and P. Linke. 2003. U/Th systematics and ages of authigenic carbonates from Hydrate Ridge, Cascadia Margin: Recorders of fluid flow variations. Geochimica et Cosmochimica Acta **67**: 3845–3857.

Teichert, B., G. Bohrmann, and E. Suess. 2005. Chemoherms on Hydrate Ridge--Unique microbially-mediated carbonate build-ups growing into the water column. Palaeogeography, Palaeoclimatology, Palaeoecology **227**: 67–85.

Teske, A., and K. B. Sorensen. 2008. Uncultured archaea in deep marine subsurface sediments: have we caught them all? The ISME Journal **2**: 3–18.

Thiel, V., J. Peckmann, H. H. Richnow, U. Luth, J. Reitner, and W. Michaelis. 2001. Molecular signals for anaerobic methane oxidation in Black Sea seep carbonates and a microbial mat. Marine Chemistry **73**: 97–112.

Thurber, A. R., L. A. Levin, V. J. Orphan, and J. J. Marlow. 2012. Archaea in metazoan diets: implications for food webs and biogeochemical cycling. The ISME Journal **6**: 1602–1612.

Torres, M., J. McManus, D. Hammond, M. De Angelis, K. U. Heeschen, S. L. Colbert, M. D. Tryon, K. M. Brown, and E. Suess. 2002. Fluid and chemical fluxes in and out of sediments hosting methane hydrate deposits on Hydrate Ridge, OR, I: Hydrological provinces. **201**: 525–540.

Treude, T., A. Boetius, K. Knittel, K. Wallmann, and B. B. Jørgensen. 2003. Anaerobic oxidation of methane above gas hydrates at Hydrate Ridge, NE Pacific Ocean. Marine Ecology Progress Series **264**: 1–14.

Ussler, W., III, and C. K. Paull. 2008. Rates of anaerobic oxidation of methane and authigenic carbonate mineralization in methane-rich deep-sea sediments inferred from models and geochemical profiles. Earth and Planetary Science Letters **266**: 271–287.

Vasconcelos, C., J. A. Mckenzie, R. Warthmann, and S. M. Bernasconi. 2005. Calibration of the δ18O paleothermometer for dolomite precipitated in microbial cultures and natural environments. Geology **33**: 317–320.

Vasconcelos, C., J. A. Mckenzie, S. Bernasconi, D. Grujic, and A. J. Tiens. 1995. Microbial mediation as a possible mechanism for natural dolomite formation at low temperatures. Nature **377**: 220–222.

Wang, Q., G. M. Garrity, J. M. Tiedje, and J. R. Cole. 2007. Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. Applied and Environmental Microbiology **73**: 5261–5267.

Watanabe, Y., S. Nakai, A. Hiruta, R. Matsumoto, and K. Yoshida. 2008. U–Th dating of carbonate nodules from methane seeps off Joetsu, Eastern Margin of Japan Sea. Earth and Planetary Science Letters **272**: 89–96.

Chapter Two

METHANE SEEP CARBONATES HOST DISTINCT, DIVERSE, AND DYNAMIC MICROBIAL ASSEMBLAGES

David H. Case[1]

in collaboration with,

Alexis L. Pasulka[1], Jeffrey J. Marlow[1], Benjamin M. Grupe[2], Lisa A. Levin[2], and Victoria J. Orphan[1]

[1]Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA, USA

[2]Center for Marine Biodiversity and Conservation, Scripps Institution of Oceanography, University of California – San Diego, La Jolla, CA, USA

*Published in Case et al., 2015. Methane Seep Carbonates Host Distinct, Diverse, and Dynamic Microbial Assemblages. mBio 6 (6), e01348-15.

2.0 ABSTRACT

Marine methane seeps are globally distributed geologic features in which reduced fluids, including methane, are advected upward from the subsurface. As a result of alkalinity generation during sulfate-coupled methane oxidation, authigenic carbonates form slabs, nodules, and extensive pavements. These carbonates shape the landscape within methane seeps, persist long after methane flux is diminished, and in some cases are incorporated into the geologic record. In this study, microbial assemblages from 134 native and experimental samples across 5,500 km, representing a range of habitat substrates (carbonate nodules and slabs, sediment, bottom water, and wood) and seepage conditions (active and low-activity), are analyzed to address two fundamental questions of seep microbial ecology: (1) do carbonates host distinct microbial assemblages, and (2) how sensitive are microbial assemblages to habitat substrate type and temporal shifts in methane seepage flux? Through massively parallel 16S rRNA gene sequencing and statistical analysis, native carbonates are shown to be reservoirs of distinct and highly diverse seep microbial assemblages. Unique coupled transplantation and colonization experiments on the seafloor demonstrate that carbonate-associated microbial assemblages are resilient to seep quiescence and reactive to seep activation over 13 months. Varying rates of response to simulated seep quiescence and activation are observed among similar phylogenies (e.g. *Chloroflexi* OTUs) and similar metabolisms (e.g. putative S-oxidizers), demonstrating the wide range of microbial sensitivity to changes in seepage flux. These results imply that carbonates do not passively record a time-integrated history of seep microorganisms, but rather host distinct, diverse, and dynamic microbial assemblages.

2.1 IMPORTANCE

Since their discovery in 1984, the global distribution and importance of marine methane seeps has become increasingly clear. Much of our understanding of methane seep microorganisms – from metabolisms to community ecology – has stemmed from detailed studies of seep sediments. However, it has become apparent that carbonates represent a volumetrically significant habitat substrate at methane seeps. Through combined *in situ* characterization and incubation experiments this study demonstrates that carbonates host microbial assemblages distinct from and more diverse than other seep habitats. This emphasizes the importance of seep carbonates as biodiversity locales. Furthermore, we demonstrate that carbonate-associated microbial assemblages are well adapted to withstand fluctuations in methane seepage, and we gain novel insight into particular taxa that are responsive (or recalcitrant) to changes in seep conditions.

2.2 INTRODUCTION

Marine methane seeps serve as islands of diverse and dense deep-sea life, with food webs extending from microorganisms to varied megafauna including clams, mussels, and tube worms (Levin 2005; Thurber et al. 2012; Niemann et al. 2013). Distinct habitats associated with methane seeps include sediments, bottom water, loosely consolidated carbonate protoliths (hereafter "nodules"), fully lithified carbonate blocks and pavements (hereafter "carbonates"), and, occasionally, wood. Marine methane seep microbial communities and corresponding geochemistry within sediments have been intensively investigated, and have been found to frequently be dominated by microbial taxa performing anaerobic oxidation of methane (AOM), notably anaerobic methane-oxidizing archaea (ANME) and deltaproteobacterial sulfate-reducing bacteria (SRB; Hinrichs et al. 1999; Orphan et al. 2001; Pop Ristova et al. 2015). More broadly, seep sediments are biologically diverse locales that host microorganisms spanning many phyla, and are often rich in Epsilonproteobacteria and Gammaproteobacteria in addition to the canonical AOM-associated taxa (Nunoura et al. 2012; Ruff et al. 2013; Marlow et al. 2014b; Pop Ristova et al. 2015). A distinct "seep microbiome", rich in Deltaproteobacteria, Methanomicrobia, and Candidate Divisions Hyd24-12 and JS1, is apparent when comparing seep sediment- and nodule-associated microbial assemblages to other marine environments (Ruff et al. 2015).

Authigenic carbonates, which are believed to form as a result of increased alkalinity associated with AOM metabolism, constitute the most pervasive solid habitat substrate at methane seeps, but are historically less well sampled than sediments. Carbonates are known to host lipid (Thiel et al. 2001; Stadnitskaia et al. 2005) and ribosomal DNA (Stadnitskaia et al. 2005; Heijs et al. 2006; Marlow et al. 2014b) biomarkers, as well as record carbon isotopic compositions reflective of microbial AOM processes (Greinert et al. 2001; Gieskes et al. 2005). Seep carbonates have recently been shown to host viable autoendolithic (organisms whose

metabolism induces self-entombing mineral formation) Archaea and Bacteria capable of methane oxidation (Marlow et al. 2014a; 2015), as well a metazoan communities (Levin et al. 2015). Carbonates themselves occur in a variety of sizes, morphologies, and mineralogies. These include mm- to cm-scale poorly consolidated precipitates, termed "nodules" or "concretions", occurring within seep sediments (Chen et al. 2006; Watanabe et al. 2008; Mason et al. 2015). Seep-associated carbonates are also frequently found exposed at the seafloor in cm- to 10s of m-sized isolated blocks and continuous pavements (Hovland et al. 1987; Naehr et al. 2007), often extending both laterally and vertically from the site of contemporary methane seepage (Teichert et al. 2005; Sahling et al. 2008). Observations of carbonates at sites lacking contemporary seepage provide evidence that carbonates can outlive seepage processes on the seafloor, supported by the recovery of demonstrably seep-associated carbonates from geologic outcrops as old as 300 million years (Birgel et al. 2008). Diversity relationships between microbial assemblages associated with seep sediments, nodules, and carbonates have just recently begun to be explored (Marlow et al. 2014b; Mason et al. 2015).

Seepage flux can increase and decrease, as well as shift spatially, on the scale of days (Tryon et al. 2002) to weeks (Tryon et al. 1999) to centuries (Bekins and Dreiss 1992; Tryon et al. 2002). Microbial assemblages presumably adapt to spatial and temporal changes in seepage flux, but the extent and rate of response *in situ* remains uncharacterized. Contemporary seepage activity is often defined categorically based on the presence or absence of diagnostic seafloor chemosynthetic communities within methane seeps. "Active" sites are defined, in this study and elsewhere (Tryon et al. 2002; Orphan et al. 2004; Boetius and Suess 2004; Levin et al. 2015), as hosting sulfur-oxidizing bacterial mats, clam beds, dense snail colonies, and/or methane ebullition, while "low-activity" areas lack those diagnostic indicators of contemporary seepage. Notably, low-activity sites are often within $<10^2$ meters from active sites, frequently host carbonates, and can still exhibit microbial activity, including AOM, at reduced rates (Marlow et al. 2014a). Diversity surveys using conventional cloning and sequencing have shown that seep-

associated archaeal assemblages, of which only a fraction of the taxa were ANME subgroups, differed based on local seepage activity. The same trend was not apparent in bacterial assemblage composition, which instead was more influenced by habitat substrate (sediment vs nodule vs carbonate; Marlow et al. 2014b). Lipid biomarker profiles from seep sediment and microbial mat samples have been shown to be differentiated partially by sulfate reduction rate, which is likely in turn correlated with seep activity (Rossel et al. 2011). Off-seep sites host microbial assemblages that are distinct from both active and low-activity sites, further indicating the existence of a "seep microbiome" (Marlow et al. 2014b; Pop Ristova et al. 2015; Ruff et al. 2015).

Here a combined comparative and experimental *in situ* approach is applied to characterize the relationship between seep microbial assemblages, habitat substrata (carbonate vs sediment vs nodule vs bottom water vs wood), and varying seep activity (active vs low-activity stations). By coupling a massive sampling effort of native, unperturbed seep carbonates to *in situ* transplantation and colonization experiments, we can leverage these compatible datasets to address two fundamental microbial ecology questions: (1) do seep carbonates host distinct microbial assemblages, and (2) how sensitive are microbial assemblages to habitat substrate type and availability and temporal shifts in methane seepage flux?

2.3 MATERIALS AND METHODS

2.3.1 SAMPLE COLLECTION AND DEPLOYMENT OF EXPERIMENTS

The majority of samples in this study (114 out of 134; Table S1), including all transplantation and colonization treatments (see below), are from an extensively-studied natural laboratory of methane seepage: namely, the Northern and Southern promontories of Hydrate Ridge ("HR"), on the Cascadia margin, Oregon, USA (HR-North: 44˚40'N, 125˚6'W, ~600 meters below sea level [mbsl], HR-South; 44˚34'N, 125˚9'W, ~800 mbsl, Fig. S1A-B, D-H; Suess

et al. 1985; Boetius et al. 2000; Sahling et al. 2002; Boetius and Suess 2004; Levin et al. 2010; Guilini et al. 2012). Active and low-activity stations were identified by presence (or absence) of benthic chemosynthetic communities throughout HR and given sequential names for experimental purposes (Stations spaced 10^1-10^4 meters apart on the seafloor, see Fig. S1D-E, Table S1). Our active and low-activity station designations were confirmed by pore water sulfide concentrations from 0-3 centimeter-below-seafloor horizons of sediment cores collected within active stations (1-14 mM range, 6mM average, n=9) and low-activity stations (0-0.9 mM range, 0.2 mM average, n=5; more details in Supplemental Text). 110 out of 114 HR samples, including carbonates, sediments, nodules, bottom water, and woods, were collected from these stations, with four additional carbonate samples obtained from a seep promontory approximately 20 km SSE of HR ("Southeast Knoll"; 44˚ 27.0'N, 125˚ 7.8'W, ~620 mbsl). Of the remaining 20 samples in this study, 10 carbonates were collected from seeps off the Costa Rica coast: Mound 11, Mound 12, Quepos Mound, and Jaco Scarp (Fig. S1C; Sahling et al. 2008; Levin et al. 2012; Dekas et al. 2014). As a point of comparison to sediments and nodules collected at HR, we also included ten sediment and nodule samples from Eel River Basin (ERB; 40˚48.7'N, 124˚36.7'W, 517 mbsl; Levin et al. 2003; Orphan et al. 2004). Recently published sequencing data from 18 sediment and nodule samples (13% of our 134-sample dataset) provide valuable context for this study regarding habitat substrate and are denoted in Table S1 (Mason et al. 2015).

Among all the collected samples, 82 out of 134 represent native, unperturbed microbial assemblages associated with a variety of habitat substrates ($n_{carbonate}$=57; n_{nodule}=10; $n_{bottom\ water}$=2; $n_{sediment}$=13) and seep activity levels (n_{active}=52, $n_{low\text{-}activity}$=28, $n_{off\ seep}$=2). Samples were collected in 2006, 2009, 2010, and 2011 during R/V Atlantis cruises AT15-11, AT15-44, AT15-68, and AT18-10, respectively. Upon shipboard retrieval, subsamples were immediately frozen at -80˚C and transferred to an onshore lab for downstream processing. Mineralogy of carbonate samples was examined by powder X-ray diffraction (see Supplemental Text).

Six transplanted carbonate and 46 introduced carbonate and wood (n_{carb}=20, n_{wood}=26) samples represent microbial assemblages after 13 months of incubation on the seafloor. Transplantation experiments were conducted using DSV Alvin in August 2010 by moving seafloor carbonates at HR-North from active to low-activity stations (n=4) and vice versa (n=2), followed by collection and freezing in September 2011 using ROV Jason II. Colonization experiments were conducted with fir and pine woods (n=26) and autoclaved, aseptically stored calcite and dolomite seep carbonates (n=20) deposited at selected seafloor stations, including those of the transplantation experiments, in August 2010 (AT15-68) and recovered in September 2011 (AT18-10). More methodological details regarding the transplantation and colonization experiments can be found in the Supplemental Text.

2.3.2 GENOMIC DNA EXTRACTION AND 16S rRNA GENE SEQUENCING AND PROCESSING

Onshore, the carbonates, sediments, and nodules were separately ground into powder with a sterile porcelain mortar and pestle. The nodules, which were only loosely consolidated and thus could have contained sediment-phase contamination, were pre-processed in order to thoroughly remove sediment as previously described (Mason et al. 2015), with the exception of nodule #5118N (Table S1). Genomic DNA was extracted following the general procedure of the MoBio PowerSoil kit (MoBio, St. Louis, MO, see Orphan et al. 2001 for variations from default protocol), using ~400 mg powder. For wood samples, a sterile razor blade was used to collect shavings from the exterior, avoiding the bark and any observed animals (e.g., shipworms) whenever possible. DNA from wood samples was extracted using the MoBio PowerPlantPro kit's recommended protocol with 40 µL of Phenolic Separation Solution and ~70 mg wood shavings. Bottom water samples from nearby station HR-9 (Fig. S1) were collected on a 0.2 µm filter and

extracted by phenol-chloroform followed by CsCl density gradient centrifugation (Tavormina et al. 2010).

Preparation for sequencing of the V4 region of the 16S rRNA gene was performed with universal primers according to the Earth Microbiome Project (EMP, "iTag" sequencing; Gilbert et al. 2011) recommended protocol (Caporaso et al. 2011; 2012), with minor modifications as previously described (Mason et al. 2015). Raw sequences were generated on an Illumina MiSeq platform at Laragen, Inc. (Los Angeles, CA) and are available in the Sequence Read Archive under accession numbers SRP055767 and SRP049675. In-house data processing was completed in QIIME1.8.0 and included joining paired ends, quality trimming, chimera checking, 97% OTU clustering, singleton removal, PCR contaminant removal, 0.01% relative abundance threshold removal, and rarefaction to 16,051 sequences per sample (Supplemental Text). Taxonomic assignments were generated according to an appended version of the Silva 115 database (details in Mason et al. 2015).

2.3.3 DIVERSITY ANALYSES

Alpha diversity calculations (Shannon Index (H'), Observed OTUs, and Chao1) were carried out in QIIME1.8.0 (alpha_diversity.py). Non-metric multidimensional scaling (NMDS) analyses were carried out in the R environment (R Core Team 2014) after applying a square root transformation to the relative abundance data. For all Analysis of Similarity (ANOSIM) tests, p-values of <0.05 were considered significant. R values are only reported in the text for tests which yielded significant results. Examples of the R commands, including options used, are given in the Supplemental Text. Distance-based Linear Modeling (distLM) was applied with Primer-E software to complement the ANOSIM analysis (Clarke and Warwick 2001). The Similarity Percentage (SIMPER) test was applied in R to identify specific OTUs which demonstrate different relative abundances between sample groups; key OTUs were selected for presentation

and usually represented the majority of sequences associated with each taxonomy (Fig. 3; Fig. 5; Table S3).

2.4 RESULTS AND DISCUSSION

2.4.1 CARBONATES HOST DISTINCT AND DIVERSE SEEP MICROBIAL ASSEMBLAGES

Ordination of the sample set reveals the microbial assemblages to be most strongly differentiated by habitat substrate (i.e. carbonate, sediments and nodules, bottom water; Fig. 1A; $R=0.49$; $p<0.001$; all ANOSIM results presented in Table S2). Habitat substrate is also the most significant factor associated with microbial assemblages as determined by distLM, accounting for 25% of the inter-sample variability. Furthermore, carbonates exhibit higher OTU richness than the other substrates included in this study (Fig. 2A; Chao1 estimates are given in the main text, raw OTU rarefactions are given in Fig. S2). These trends are also observed in the macrofauna recovered from seep carbonates (Levin et al. 2015), confirming carbonates host diverse benthic life across multiple trophic levels. Overall microbial assemblages of sediments and nodules are not statistically differentiable as determined from ANOSIM tests, indicating that sediment-hosted nodules and exhumed seafloor carbonates behave as separate, distinct habitat substrates for microbial habitation (Fig. 1A, Table S2). Sediments, nodules, and carbonates have recently been shown to host different bacterial, but not archaeal, assemblages in 16S clone library surveys (Marlow et al. 2014b), while recent examination of a subset of our iTag data demonstrated similar microbial communities inhabiting nodules and adjacent sediments, especially in active seep settings (Mason et al. 2015).

Examination the top thirty most abundant OTUs in our dataset reveals a variety of Archaea and Bacteria composing the samples (Fig. 3A), including taxonomies common to

methane seep settings (e.g., ANME subgroups and Deltaproteobacteria). The higher relative abundance of ANME-1 in sediments and nodules as compared to carbonates is in agreement with previous clone library observations at Hydrate Ridge, while the recovery of epsilonproteobacterial sequences from sediments, nodules, and carbonates is in contrast to previous findings in which they were almost exclusively recovered from sediments (Fig. 3A; Marlow et al. 2014b). Data from sequencing of mock communities suggests a slight bias for ANME-1 and stronger bias against the recovery of ANME-2 sequences by the modified EMP protocol (Trembath-Reichert et al., 2016). Thus, we note the relative abundance of these groups may in reality be slightly lower (ANME-1) or higher (ANME-2) than recovered in our iTag dataset. However, the inter-substrate trends, which are similar for ANME-1 and ANME-2, should be unaffected. Abundance patterns of ANME and other taxa are discussed in detail in the sections below, in the context of results from our experimental manipulations.

Inter-substrate differences in microbial assemblage are the cumulative result of contributions from many OTUs, with no single OTU accounting for more than 2% of the total inter-substrate differences. Nonetheless, several OTUs can be identified which are strongly associated with one habitat type (Fig. 3B). Notably, taxa previously identified as diagnostic of the "seep microbiome" (i.e., JS1 archaea and Deltaproteobacteria; Ruff et al. 2015) are observed in our dataset to be characteristic of sediments and nodules, but not carbonate habitats (Fig. 3B). In determining the "seep microbiome," Ruff et al., 2015 examined methane seep sediments and nodules exclusively; our data thus corroborate their results, but also further demonstrate that seep carbonates host distinct microbial assemblages. Carbonates, to the exclusion of other habitat substrates, are observed to host an OTU associated with the gammaproteobacterial JTB255 Marine Benthic Group (Fig. 3B). The physiology of this group remains undetermined, though uncultured members have been recovered from a variety of marine sediments (Bowman and McCuaig 2003; Schauer et al. 2010), including methane seeps (Li et al. 1999). OTUs associated with the deltaproteobacterial SAR324 clade and thaumarchaeal Marine Group 1 are particularly

abundant in the bottom water samples (Fig. 3B), although we note that a separate thaumarchaeal Marine Group 1 OTU is more abundant on carbonates than on other substrates (Fig. 3A). This exemplifies the potential for OTUs of similar phylogeny to be differentially distributed in the environment.

Shannon diversity (H'), which measures evenness in addition to richness, is higher in the carbonates than either the sediments/nodules or the bottom waters (Fig. S2D). Carbonate-associated assemblages may exhibit distinct microbial molecular signatures due to either geochemical (i.e., preferential adsorption of metabolites to the carbonate matrix; Ijiri et al. 2009), physical (i.e., a site for microbial biofilm attachment), or historic (i.e., formation within or above the sediment column; Blumenberg et al. 2015) factors. Examination of the OTU overlap among native habitat substrates (Fig. 4A) demonstrates that carbonates share more OTUs with sediments and nodules than bottom waters, supporting the hypothesis of formation within the sediments, followed by subsequent exhumation and exposure at the seafloor (Gieskes et al. 2005; Blumenberg et al. 2015). However, bottom waters share more OTUs with carbonates than sediments or nodules, revealing that a subset of bottom water microorganisms do passively or actively inhabit carbonates exposed at the seafloor (Fig. 4A).

Close overlap in assemblage composition is observed between some of the carbonates (~10 of 57, all from active seep stations) and sediments/nodules (Fig. 1A, c.f. 10 carbonates highlighted in Fig. S3A). It is possible that carbonate samples hosting microbial assemblages similar to sediments/nodules may have contained excess sediment entrained in the rock matrix upon recovery (c.f. Blumenberg et al. 2015); alternatively, nodules in the overlapping regions may have been sufficiently lithified to begin hosting "carbonate-like" microbial assemblages (e.g., nodule #C2693 in Fig. 1A), though this does not necessarily explain similarity of some sediment samples. The compositional overlap between ~10 active-station carbonate assemblages and sediment/nodule assemblages is not derived from geographic proximity, as the sediments/nodules from HR do not exclusively plot in close proximity to the carbonate samples, which are

dominantly from HR (Fig. S3A). Nor are the overlapping carbonates unified by seafloor station, mineralogy, or collection year.

2.4.2 DEMONSTRATION OF MICROBIAL VARIABILITY WITHIN SEEP CARBONATES

The high total OTU richness of carbonates (Fig. 2A), combined with OTU overlap between carbonates and other substrates (Fig. 4A) could indicate that carbonates represent a passive repository of preserved and extant microorganisms. We test this possibility by first examining in detail, in this section, the native carbonate samples (n=57), which allows inference of the environmental indicators associated with differences between carbonate-hosted microbial assemblages. In the following sections, we then couple these interpretations to the *in situ* transplantation (n=6) and colonization ($n_{carbonate}$=20; n_{wood}=26) experiments, respectively.

On their own, native carbonate-associated microbial assemblages demonstrate clear differentiation according to seep activity (R=0.45; p<0.001; Fig. 1B; Table S2), mineralogy (R=0.44; p<0.001; Fig. S4; Table S2), and seafloor station ($R_{active\ stations}$=0.31, $p_{active\ stations}$=0.002; $R_{low\text{-}activity\ stations}$=0.27, $p_{low\text{-}activity\ stations}$=0.037; Table S2). The similar parsing of native carbonate-hosted assemblages by seep activity and mineralogy is partially explained by our observation of a qualitative relationship between seep activity and carbonate mineralogy, with a higher proportion of aragonite-bearing carbonates recovered from low-activity stations (see Supplemental Text and Fig. S4). This suggests seep activity and carbonate mineralogy are not independent environmental factors in our dataset. The biogeographic differences between stations (~10^2-10^4 m) are in agreement with previous observations of within-seep microbial and geochemical heterogeneity (Treude et al. 2003; Pop Ristova et al. 2015) and recent findings that sediment-associated microorganisms in seeps exhibit "global dispersion and local diversification" (Ruff et al. 2015). A distance-decay curve demonstrated that if a biogeographic effect on microbial similarity exists

over a 10^5-10^6-m scale, it is aliased by other environmental factors (Fig. S5). We frame our further discussion in terms of seep activity because it is strongly associated with differences between carbonate-hosted assemblages and, importantly, our sample collection and *in situ* transplantation and colonization experiments were explicitly performed in order to test biological variability as a function of seep station activity. However, we emphasize that seep activity is a qualitative environmental indicator that may be correlated with other environmental factors, such as carbonate mineralogy.

With regard to standard ecological metrics of OTU richness and evenness (Fig. 2B, Fig. S2B), seep activity does not differentiate the native carbonate-associated microbial assemblages. This indicates that while carbonates at low-activity stations host distinct assemblages, they are not less diverse than microbial assemblages from carbonates at active stations. Active-station carbonates are particularly rich in OTUs associated with putative sulfur-oxidizing organisms belonging to the epsilonproteobacterial and gammaproteobacterial families Helicobacteraceae and Thiotrichaceae, respectively, as compared to carbonates from low-activity stations (Fig. 5 and Table S3). These organisms are likely supported by high sulfide concentrations produced by sulfate-coupled AOM at active seep stations. Among seep sediments in the Mediterranean Sea, epsilonproteobacterial Helicobacteraceae were found to be an indicator taxa for seepage (Pop Ristova et al. 2015), which our data corroborate. Data from hydrothermal vent systems also exhibit clear differences in abundance of putative sulfur-oxidizing Epsilonproteobacteria between active and low-activity (or inactive) sites, with increased abundance at active vent sites where delivery of reduced fluids is high (Sylvan et al. 2012b). Furthermore, Epsilonproteobacteria have been observed in time-resolved experiments to rapidly respond to geochemical heterogeneity and experimental perturbations (i.e., colonization of fresh substrate) in hydrothermal vent systems (Alain et al. 2004; Sylvan et al. 2012a; b). Physiologies of specific groups of the Gammaproteobacteria often include oxidation of either sulfur or methane (Garrity et al. 2005;

Sorokin et al. 2007), both of which are common at settings with increased delivery of reduced fluids.

ANME-1 archaea, which are the most abundantly recovered ANME in the entire iTag dataset, exhibit wide ranges of relative abundance in both active and low-activity seep stations, with higher average relative abundance at low-activity stations, in agreement with previous clone library observations (Fig. 5; Marlow et al. 2014b). Similarly, the deltaproteobacterial family Desulfobacteraceae does not exhibit a clear difference in observed relative abundance as a function of seep activity (Fig. 5). It thus appears that some ANME-1 and deltaproteobacterial OTUs may be relatively insensitive to seepage level. This was unexpected as these are key taxa involved in the AOM process and therefore hypothesized to occur at higher relative abundance in methane-replete, presumably "active", seep stations. ANME-1 may be performing methanotrophy even within carbonates at low-activity stations, consistent with recent reports of AOM associated with carbonates on the periphery of active seepage (Marlow et al. 2014a). Alternatively, relic DNA from AOM-associated organisms may be preserved within carbonate rocks, as the carbonate precipitation process causes self-entombment, potentially sealing off inhabited pores (Stadnitskaia et al. 2005; Heijs et al. 2006; Marlow et al. 2014b; 2015). Evidence for biomarker preservation within carbonates has been described for lipids, which are more recalcitrant to degradation than DNA (Thiel et al. 2001; Stadnitskaia et al. 2005; Blumenberg et al. 2015).

2.4.3 DEMONSTRATION OF SUCCESSIONAL DYNAMICS: TRANSPLANTATION EXPERIMENTS

The "snapshot" view of carbonate-associated microbial ecology is augmented by the seafloor transplantation experiments, which allow us to observe *in situ* microbial successional patterns by simulating seep quiescence and activation. *In situ* flux measurements at Hydrate Ridge

have shown that seep activity can shift on week- to month-long timescales (Tryon et al. 1999; 2002), indicating our 13-month transplantation experiments are relevant to contemporary processes at Hydrate Ridge and potentially in other methane seep regions.

The OTU composition of the four active-to-low-activity transplanted microbial assemblages are statistically differentiable from both the native, active carbonate-associated microbial assemblages ($R=0.32$, $p=0.008$) as well as the native, low-activity carbonate-associated assemblages ($R=0.88$, $p<0.001$; Table S2; Fig. 1B). The four microbial assemblages transplanted from active to low-activity stations are more similar to the native, active carbonate assemblages (i.e., from where they originated) than to the native low-activity assemblages (i.e., to where they were transplanted, Fig 2B and ANOSIM results). The four transplanted carbonates exhibit approximately 30% lower overall OTU richness as compared to native carbonates (Fig. 2B), but in-depth analysis of OTU overlap between transplanted and native carbonates reveals a level of fine structure to the microbial turnover and succession (Fig. 4B-C). At the paired HR-3/-4 and HR-7/-8 stations, 68% and 52%, respectively, of the OTUs associated with native, active control carbonates are not recovered upon simulated seep quiescence after 13 months (Table S3). The "lost" OTUs are supplanted by characteristic OTUs gained from the low-activity sites (28 and 37 OTUs, representing 18% and 17% of the recovered OTUs for HR-3/-4 and HR-7/-8 transplants, respectively) as well as OTUs unique to the transplants and not recovered from native carbonates (20 and 24 OTUs for HR-3/-4 and HR-7/-8, respectively; Table S3). Nearly half of the OTUs recovered among the HR-3/-4 and HR-7/-8 transplants were cosmopolitan OTUs that were also observed in both the native, active and native, low-activity carbonates (Fig. 4B-C; Table S3). Thus, a loss of over half the initial OTUs upon seep quiescence is masked by gain of new OTUs, both unique and shared with the low-activity controls.

Combining the observation of overall similarity to native, active assemblages, diminished overall OTU richness, and specific turnover among the carbonates transplanted to low-activity stations, we can begin to paint a picture of microbial succession upon seep quiescence. Most

major (i.e., highly abundant) constituent members of carbonate-associated microbial assemblages are resilient to one year of quiescence (or their DNA doesn't degrade), as evidenced by the fact that transplanted carbonates plot among the native, active controls in Fig. 1B. Indeed, of the four carbonates transplanted to low-activity sites, we observe that 49-90% of the recovered sequences are from resilient OTUs shared with the active-station controls (Table S3). However, over the course of a year, low-abundance assemblage members are vulnerable to cessation of seep activity: the average relative abundance of lost OTUs in the native, active controls upon simulated quiescence was <0.5% (Table S3).

Examining specific taxa of interest, we find the gammaproteobacterial Thiotrichaceae OTUs remain at a similar relative abundance as the native, active carbonates, consistent with resilience to seep quiescence (Fig. 5). In contrast, epsilonproteobacterial Helicobacteriaceae OTUs that are highly abundant in native, active carbonates had mostly disappeared after 13 months of simulated seep quiescence (Fig. 5). Thus, two putative sulfur-oxidizing groups exhibit different 16S rRNA gene distribution, highlighting the potential for variable response to environmental change, even among taxa putatively belonging to the same guild. ANME-1 OTUs were recovered at high relative abundance in the carbonates transplanted to low-activity stations, consistent with the trend observed in native, low-activity carbonates and suggesting an ability to respond over a period of time that may represent, to ANME archaea, only a few generations (Girguis et al. 2003; Orphan et al. 2009; Morono et al. 2011).

The two carbonates which experienced simulated seep activation (transplanted from low-activity to active stations) host microbial assemblages different from low-activity, native carbonates and somewhat similar to native, active assemblages (Fig. 1B), although this experimental set suffers from low sample number associated with technical difficulties in recovering two of four originally transplanted carbonates. In juxtaposition to seep quiescence, which demonstrated resilience of the bulk microbial assemblages, our simulation of seep activation indicates that assemblages are relatively quick to respond to renewed seepage

conditions. This is especially true among the epsilonproteobacterial Helicobacteraceae OTUs, which are recovered in high relative abundance in the transplant-to-active carbonates, despite low relative abundance in the low-activity carbonates (Fig. 5). Other OTUs, for example gammaproteobacterial Thiotrichaceae, clearly demonstrate a slower response to seep activation (Fig. 5). Examination of two OTUs of putatively heterotrophic Chloroflexi, the Anaerolineaceae and Caldilineaceae, also reveals slow response to seep activation, despite their relatively high recovery among native, active seep carbonates (Fig. 5). The Anaerolineaceae OTU also exhibits markedly higher tolerance to low-activity conditions than the Caldilineaceae OTU (Fig. 5), highlighting the potential for different ecological expression among groups of similar phylogeny.

The coupled transplant experiments provide strong evidence that many carbonate-associated seep microbial taxa are adapted to cycles of seep quiescence and activation. This may be ecologically advantageous in an environment where fluid flow has a tendency to fluctuate rapidly and frequently (Tryon et al. 1999; 2002). Recalcitrance to seep quiescence is consistent with low but measurable AOM from carbonates at low-activity stations (Marlow et al. 2014a), and the physical buffering provided by carbonate habitats has been proposed as a factor for maintenance of microbial assemblage viability during periods of diminished seepage (Marlow et al. 2015). Alternatively, we note that 3 of the 4 carbonates transplanted from active to low-activity stations were composed of a mix of calcite and dolomite – mineralogies more common at active stations than low-activity stations (Fig. S4). If mineralogy significantly drives microbial composition, the observed recalcitrance to community shift may be explained by the fact that the transplanted carbonates bore mineralogies qualitatively associated with "active-seep-type" microbial assemblages. In contrast, the two samples transplanted from low-activity to active stations were aragonite/calcite mixes – a mineralogical composition regularly recovered from all seep stations regardless of activity (Fig. S4). Thus, the observed shift to an "active-seep-type" community is more likely a function of the seep activity shift than of mineralogy. The rapid microbial rebound upon simulated seep activation may be analogous to previous observations of

microbial community activation from deep terrestrial and marine subsurface environments (Morono et al. 2011; Rajala et al. 2015). Species richness in carbonates transplanted to active stations is higher than the reciprocal transplants – though still lower than native carbonates – further indicating microbial assemblage responsiveness to simulated seep activation (Fig. 2B). Diminished OTU richness upon transplantation (in either direction) is also evidence against a "time-integrative" model of carbonate microbial assemblages: if carbonates were passive recorders of all historic seep microbial DNA, OTU richness would not be expected to decrease.

2.4.4 DEMONSTRATION OF SUCCESSIONAL DYNAMICS: COLONIZATION EXPERIMENTS

Though our transplantation experiments best simulate the temporal variability of seepage for established microbial assemblages, they are limited in scope. To increase the interpretative power of our dataset, we supplemented the transplant experiments with carbonate (calcite and dolomite) and wood (fir and pine) colonization experiments to address the successional patterns and responsiveness of seep microorganisms colonizing at the seabed under conditions of differing seep activity and colonization substrate type.

Results from these experiments follow similar trends observed in the survey of native microbial assemblages where both habitat substrate ($R_{\text{carbonate vs wood}}=0.63$, $p<0.001$) and seep activity ($R_{\text{active vs low-activity}}=0.38$, $p<0.001$) differentiate the recovered microbial diversity (Fig. 1C, Table S2). In contrast to the survey of native carbonates, mineralogy did not contribute significantly to differences in total colonizing assemblage diversity ($p=0.109$; Table S2), further suggesting that the relationship between mineralogy and microbial diversity in the native carbonates may be due to a qualitative link between mineralogy and seep activity (Fig. S4). Microbial assemblages colonizing carbonates exhibited higher OTU richness and evenness than

those colonizing wood (Fig. 2C, Fig. S2C-D), substantiating the role of seep carbonates, specifically, as hosts of diverse microbial populations.

While hosting comparable OTU richness to the native carbonates (Fig. 2), the microbial assemblages colonizing the sterile carbonates at the seabed were, after 13 months, significantly different from native microbial assemblages collected in this study ($R=0.65$, $p<0.001$, Table S2; Fig. 1B). This supports general trends in the transplant experiments, suggesting that more than 13 months are required to achieve a mature successional phase if it is assumed that given enough time the colonizing assemblages would eventually mimic the native assemblages. Alternatively the colonization carbonates might never host microbial assemblages completely similar to the native carbonates, considering the different history of colonization (located at the seabed) and native (believed to have formed within the sediment column and later to have been exhumed) carbonates. Notably, however, sterile carbonates incubated at the seafloor share most of their observed OTUs with the native carbonates (Fig. 4D-I). The discrepancy between colonization and native carbonates hosting quite different microbial assemblages (Fig. 1B) and yet sharing many OTUs (Fig. 4D-I) implies assemblage differences are generally a function of differential OTU relative abundance, not of presence/absence of different OTUs themselves. Indeed, an ANOSIM test on presence/absence-normalized data reveals a diminished, though still significant, strength of difference between native and colonized carbonate microbial assemblages ($R=0.53$, $p<0.001$). In further support, among the six colonization/native pairings examined in detail (Fig. 4D-I), the majority (average 63%, range 46-84%, $n_{\text{colonization samples}}=12$) of the recovered colonization sequences were from OTUs shared between the colonization and native carbonates.

In-depth analysis of OTU overlap at station HR-9, chosen because of the wide array of habitat types and experimental samples obtained there, reveals that of the various OTUs shared between native and colonized carbonate assemblages, many are also shared with sediment and nodule assemblages (Fig. S6). This suggests some transference of sediment-hosted microbes onto the colonization carbonates. The mode of transfer is currently not known but may be associated

with direct microbial motility (Bernard and Fenchel 1995; Sievert et al. 2007), macrofaunal grazing/bioturbation (Bernard and Fenchel 1995; Thurber et al. 2012), and/or advection from fluid flow or gas ebullition (Schmale et al. 2015). At station HR-9, where 376 OTUs were reproducibly recovered from both colonization carbonates, 19% (n=71), 3% (n=11), and 4% (n=14) were exclusively sourced from carbonates, sediments/nodules, and bottom waters, respectively. The bottom-water samples, associated with this station, contained 1%-2% relative abundance of an OTU associated with the gammaproteobacterial Colwelliaceae, which were also recovered at moderate relative abundances from the colonization carbonates (<1% up to 20%; Table S3; Fig. 5) despite a lack of detection on either native or transplanted carbonates. This further indicates some transference of bottom water microorganisms onto carbonates during early-phase succession, and is consistent with common ecophysiology of Colwellia as generally marine, psychrophilic, motile, chemoorganotrophic microorganisms (Garrity et al. 2005). Thus, OTU recovery from multiple nearby substrates, coupled to the observed difference between colonized carbonate and wood microbial assemblages after 13 months ($R=0.63$, $p<0.001$; Fig. 1C) could be explained by two hypotheses: either (1) OTUs are recruited from all surrounding habitats, followed by assemblage differentiation according to habitat substrate (i.e., carbonates diverge from woods), or (2) carbonate colonization is a substrate-specific process from the very first microbial succession, and then over time occasional passive capture of OTUs from other habitat substrates occurs. In either case, the colonization data support the observation from native samples that carbonates host distinct microbial assemblages. Furthermore, carbonate distinctiveness is not simply a product of time-integrated passive capture of sediment-hosted microorganisms, nor does it depend on a history of burial in sediment.

Microbial diversity within the carbonate colonization experiments is almost wholly explained by seep activity differences, in further support of observations from native carbonates (Fig. 1C, $R=0.81$, $p<0.001$). Indeed, OTUs associated with the epsilonproteobacterial Helicobacteraceae exhibit a wide range of relative abundances in the colonization carbonates at

active stations, but only a very minor amount of colonization at low-activity stations (Fig. 5; Table S3). The Thiotrichaceae OTUs also demonstrate colonization patterns reminiscent of distributions observed in the native carbonates, again indicating that putative sulfur-oxidizing OTUs are dynamic responders to carbonate substrate availability in regions of seep activity at the seabed. However, the specific Thiotrichaceae OTU observed to most strongly colonize experimental carbonates was different than the Thiotrichaceae OTU more frequently observed in the native carbonates (Table S3) – demonstrating the potential for within-group variability in ecological expression. The recovery of Helicobacteraceae or Thiotrichaceae OTUs was not obviously tied to qualitative observations of bacterial mats upon recovery of colonized carbonates from the seafloor. Previous studies of microbial colonization in shallow marine sediments and near hydrothermal vents have observed a dominance of early-stage colonization by Epsilonproteobacteria (Bernard and Fenchel 1995; Taylor et al. 1999; Alain et al. 2004; Sylvan et al. 2012a), and similar ecological behavior appears to be occurring in methane seeps. The rapid colonization by Epsilonproteobacteria in various marine settings has been attributed to both a tolerance for rapidly changing physico-chemical conditions and motility within many members of the class (Bernard and Fenchel 1995; Alain et al. 2004). We observe that our key Helicobacteraceae OTUs were recovered in high relative abundance in methane seep sediments and low relative abundance in bottom water samples (Table S3); therefore, it appears likely that the Helicobacteraceae recovered in the colonization experiments were inoculated from underlying sediments, in contrast to Colwelliaceae OTUs derived from overlying bottom waters.

Colonization by the ANME-1-associated OTUs (the same OTUs as recovered from native carbonates) on the sterile carbonates was observed at low levels at both active and low-activity stations (Fig. 5). Any level of colonization by ANME-1 is intriguing for two reasons. First, ANME-1 are believed to have doubling times on the order of several months, so the 13-month course of the colonization experiments could reasonably be expected not to have provided enough time for ANME-1 archaea to colonize and become established on the fresh carbonate

substrates (Girguis et al. 2003; Orphan et al. 2009; Morono et al. 2011). Second, ANME-1 are obligate anaerobes typically associated with highly reducing conditions located deeper within the sediment column at seeps and near the sulfate-methane transition zone, not at the sediment/water interface where the colonization experiments were located (Knittel et al. 2005). An exception to this are the Black Sea 'reefs', composed partly of ANME-1; however, these grow into permanently stratified bottom water of the euxinic Black Sea (Reitner et al. 2005). That ANME-1 OTUs are observed at significant levels in the sediment samples but at negligible levels in the aerobic bottom water samples (Fig. 3) indicates that ANME-1 are almost certainly colonizing the carbonates seeded by the underlying sediments. This highlights the complexity of potential mechanisms driving regional and global between-seep dispersion of ANME-1 archaea and perhaps other ANME sub-clades, as previously observed (Ruff et al. 2015) and perhaps accomplished through periodic sediment disturbance. In contrast to our observations, Archaea were not observed as early colonizers in hydrothermal vent colonization experiments, despite their presence within *in situ* vent communities (Alain et al. 2004). Our experiments suggest that ANME-1 archaea may exhibit phenotypes thus far undiscovered in seep settings, or may be distributed by hydrological flow or macrofaunal movements (pumping, filtering, burrowing, defecation, etc.).

Wood-colonizing microbial assemblages at methane seeps in the Mediterranean Sea have been observed to be different than surrounding, off-seep sediment-hosted microbial assemblages (Bienhold et al. 2013). Our data further demonstrate that even among active and low-activity seep stations, wood-colonizing microbial assemblages differ after 13 months (Fig. 1C). The stark difference between carbonate- and wood-colonizing assemblages in our dataset highlights the importance of habitat substrate to deep-sea microbial assemblages. The mere presence of putative sulfur-oxidizing Epsilon-, and Gammaproteobacteria in the wood colonization experiments suggests wood-falls may act as ephemeral sulfide-rich reducing habitats, possibly representing stepping stones between seeps and vents for chemosynthetic communities as has been

hypothesized for metazoans and Bacteria (Distel et al. 2000; Bienhold et al. 2013). Our results are consistent with previous characterizations of native wood-fall samples, as well as deep-sea benthic wood colonization experiments, which yielded observations of phylogenetically diverse microbial assemblages including, but not limited to, the Bacteroidetes, Firmicutes, Spirochaetes, Epsilon-, and Gammaproteobacteria (Fagervold et al. 2012; Bienhold et al. 2013; Fagervold et al. 2014), but very limited recovery of methanogenic and methanotrophic archaeal taxa (Fagervold et al. 2012). The lack of significant ANME colonization in the wood experiments (Table S3) indicates that AOM-related archaeal taxa may have more difficulty spreading geographically via wood substrates than many Bacteria. That AOM-related archaeal taxa appear to be able to colonize carbonate substrates, even on relatively short timescales, indicates a possible mode of wide geographic dispersion. Other hypotheses have included transportation in the guts of deep-sea metazoans or distribution during ocean anoxic events (Ruff et al. 2015), both of which may complement the apparent suitability of carbonate habitats for ANME.

2.5 CONCLUSIONS

In summary the deployment of *in situ* manipulation experiments, coupled to an extensive characterization of native microbial assemblages in association with varying seep habitat substrates, has enabled unique insights into the ecology of seep microorganisms. Microbial assemblages associated with carbonates at methane seeps are distinct from, and more diverse than, other habitat substrates examined in this study, i.e., sediments, nodules, and bottom waters. Further, bulk carbonate-associated microbial assemblages are adapted to resist seep quiescence and poised to respond to seep activation over 13 months. OTUs associated with the epsilonproteobacterial Helicobacteraceae are particularly sensitive to seep activity. Colonization experiments corroborate that carbonates host distinct and diverse microbial assemblages, and

recovery of ANME-1 OTUs associated with the carbonates suggests more dynamic physiologies and/or distribution processes for these organisms than previously hypothesized.

The difference in the microbial assemblages associated with native active and low-activity carbonates, coupled to the dynamics and decreased OTU richness observed in the transplant experiments, suggests that upon the final quiescence of a historic methane seep, the genomic microbial signatures recorded in carbonates could differ from those microbes which were present during active seepage. Investigation of our same research questions should be applied to lipid profiles, to investigate whether trends observed at the genomic level are likely to be preserved in the rock record; in particular, whether microbial signatures in the rock record merely reflect the final, low-activity period of seep activity rather than the biological assemblage present during the most active phases of seepage and AOM.

2.6 ACKNOWLEDGMENTS

We thank Connor Skennerton and Elizabeth Trembath-Reichert for helpful discussion on data processing. We acknowledge Stephanie Connon, Patricia Tavormina, Josh Steele, Heather Grotzinger, and shipboard teams from the Orphan, Levin, Rathburn, and Rouse labs for assistance with sample collection and processing. We are indebted to the captain, crew, and pilots of the DSV Alvin and ROV Jason II from cruises AT15-11, AT15-44, AT15-68 who made this work possible. We thank the two anonymous reviewers whose comments strengthened the study. This research was supported by a grant to VO from the NASA Astrobiology Institute (Award # NNA13AA92A). This is NAI-Life Underground Publication Number 053. This work was also support by a National Science Foundation (NSF) grant (OCE-0825791) and a Gordon and Betty Moore Foundation, Marine Microbiology Initiative grant (#3780) to VO. DC was supported by a NSF Graduate Research Fellowship. Levin lab research was supported by NSF grants OCE-0826254 and OCE-0939557.

2.7 FIGURES

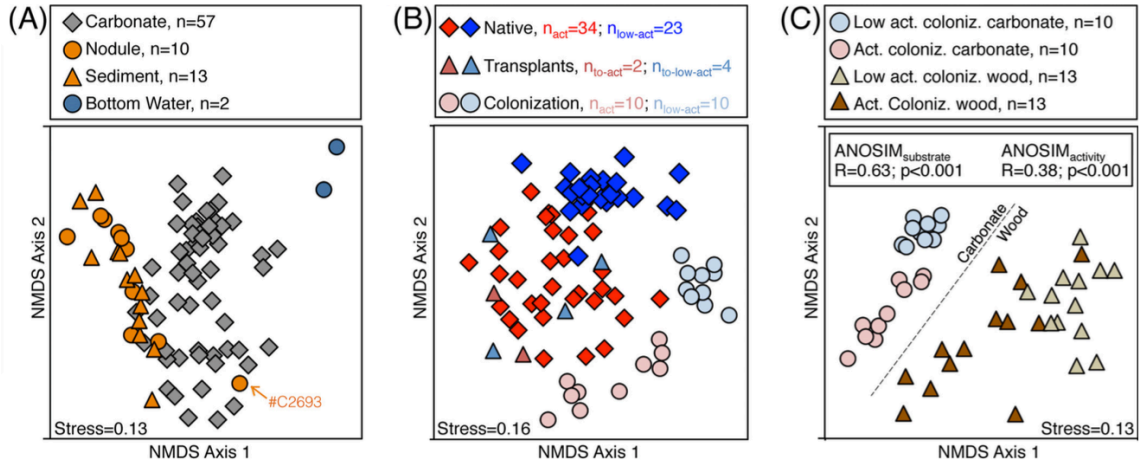


Fig. 1. Non-metric multidimensional scaling ordination of microbial assemblages in this study. Each point represents the entire recovered microbiological assemblage from one sample; samples plotting closer to each other are more similar in microbial composition. Lower stress values indicate better representation of the inter-sample (dis)similarities in two dimensions. (A) Native, unperturbed samples of sediment, nodule, bottom water, and carbonate habitat substrates. Sample #C2693 (identified by orange arrow) represents a nodule-hosted microbial assemblage recently determined to be a biological outlier among sediment and nodules (Mason et al., 2015). (B) Ordination of only carbonate samples, representing the native, transplantation, and colonization treatments. (C) Ordination of only colonization samples, representing carbonate and wood substrates at active and low-activity stations. We cannot rule out that in subplot (A), bottom water microbial assemblages could be different than sediments, nodules, and carbonates because they were extracted by a different method; the same could be true for the observed difference between carbonate- and wood-hosted assemblages in subplot (C).

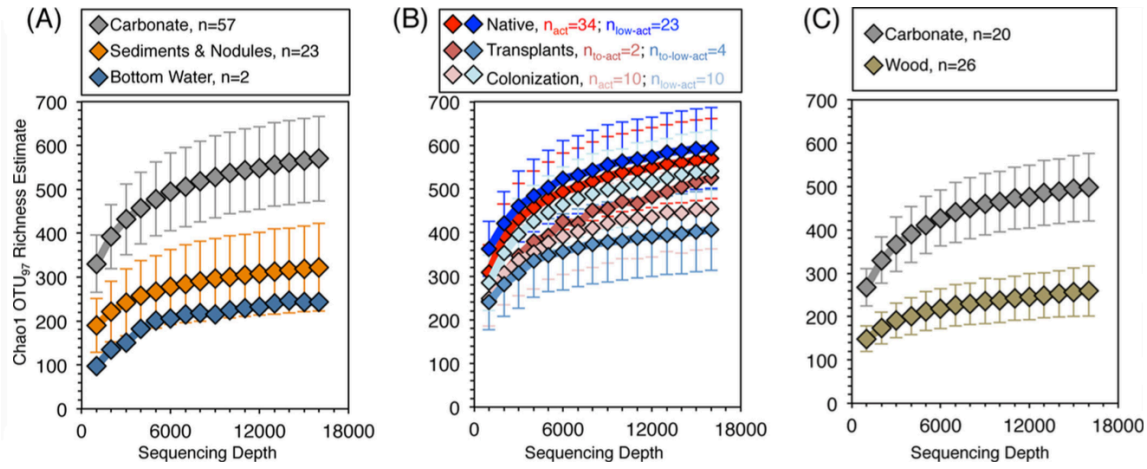


Fig. 2. Collector's curves of estimated Chao1 OTU_{97} richness. 1σ standard deviations are given by the y-axis error bars. (A) Native microbial assemblages associated with carbonates, sediments and nodules, and the two bottom water samples. Sediments and nodules were binned as one group because their associated microbial assemblages were indistinguishable according to ANOSIM tests (Table S2). (B) Carbonate samples in this study, separated by treatment category. (C) Carbonate and wood colonization samples. Standard deviations are not given for bottom water and transplant-to-active sample groups, due to the low number of analyzed samples. Raw OTU rarefaction curves are given in Fig. S2A-C.

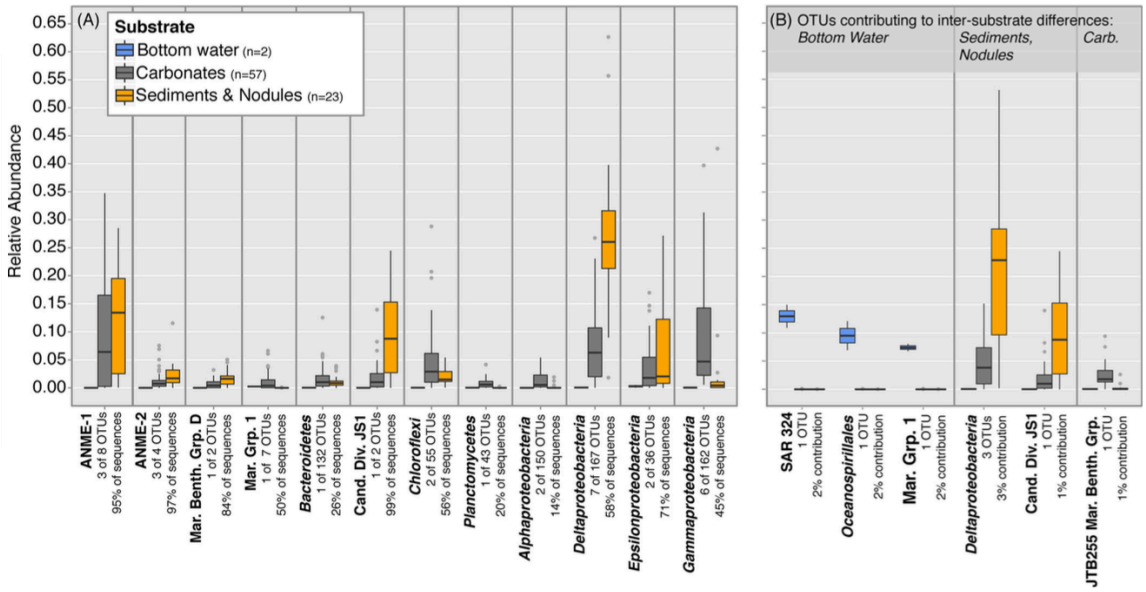


Fig. 3. Boxplot of OTU relative abundances from the 82 native samples in this study. Sediments and nodules are binned as one group because ANOSIM tests revealed their associated microbial assemblages to be statistically indistinguishable. Boxplot centerline represents the median (50th percentile, Q_{50}). The top and bottom hinges represent Q_{75} and Q_{25} quartiles, respectively. The upper and lower whiskers correspond to the highest and lowest data points within 1.5 times the Inter-Quartile Range (Q_{75} minus Q_{25}) from the median. Any data points outside of that range are identified by gray dots. The same plotting format is applied to Fig. 5. (A) Relative abundances of the top 30 most abundant OTUs in the dataset, grouped by taxonomy. The full dataset contains 1,057 OTUs, but the top 30 OTUs account for 1%, 67%, and 43% of the sequences recovered from bottom water, sediment/nodule, and carbonate substrates, respectively. (B) Relative abundances of OTUs revealed to be strongly associated with particular habitat substrates. Inter-substrate differences in microbial assemblages are a cumulative result of contributions from many OTUs; even OTUs strongly associated with a particular habitat substrate only contribute several percent to the total inter-substrate variability. Note that the JS1 OTU is the same in panels (A) and (B) – it is both highly abundant and strongly associated with sediments and nodules. The Marine Group 1 OTU in panels (A) and (B) is different – there is one Marine Group 1 OTU highly abundant in the dataset, subpanel (A), and another Marine Group 1 OTU strongly associated with bottom water samples, subpanel (B). This highlights the variable distribution of phylogenetically similar OTUs. The y-axis of panel (A) also applies to panel (B). Raw OTU data used to generate this plot is available in Table S3.

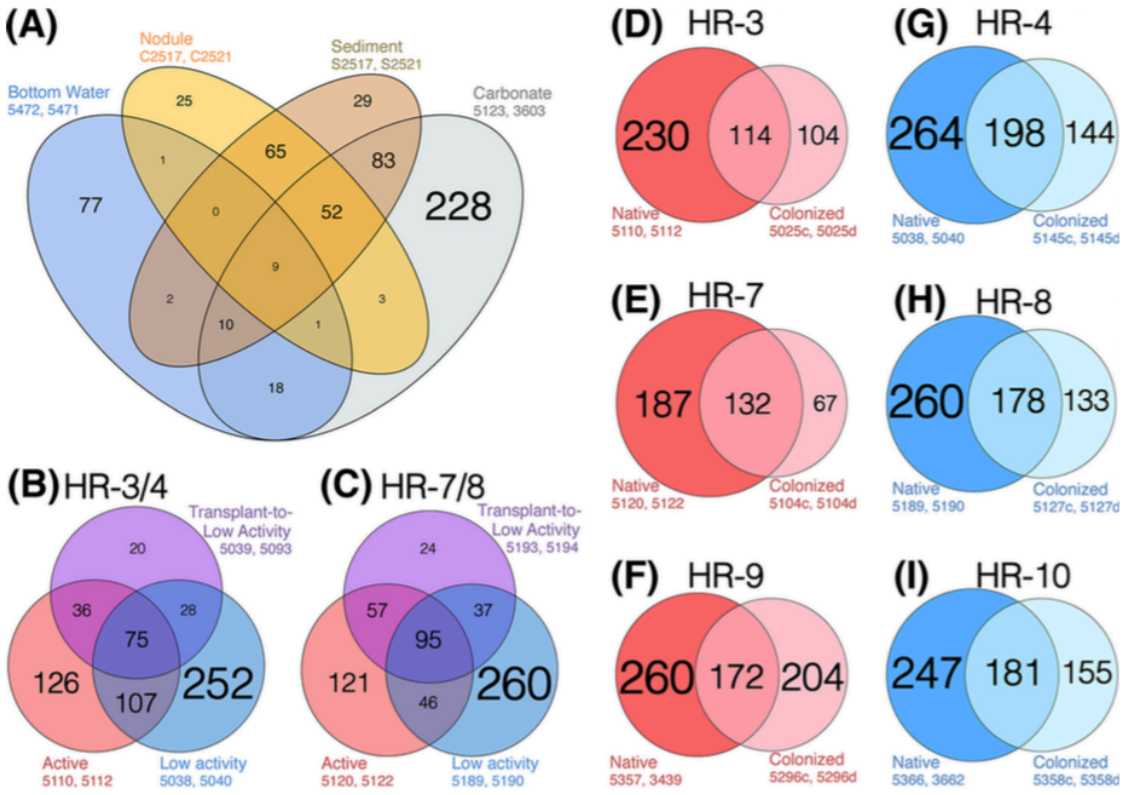


Fig. 4. Comparison of OTU_{97} overlap among various samples and treatments. In order to ensure equal depth of sampling across each substrate type, two representative samples of each substrate were chosen randomly (see sample numbers on the Figure). **(A)** OTU overlap between the four native seep habitat substrates examined in this study: sediments, nodules, carbonates, and bottom water. In order to minimize geographic bias in the analysis, samples were chosen from active stations at Hydrate Ridge South (the only exception was Bottom Water sample #5472, which was from an HR-South low-activity station). Note that carbonates host the richest OTU diversity (c.f. the collector's curve in Fig. 2A), including a large number of OTUs which are distinct to carbonates. Carbonates share more OTUs with sediments and nodules than with bottom waters, possibly indicative of an origin within the sediment column and subsequent exhumation and exposure at the seafloor. Bottom waters contribute more OTUs to carbonates than to either sediments or nodules – consistent with the recovery of our carbonates from directly on the seafloor. (B-C) OTU overlap of active and low-activity control carbonates, and transplant-to-low-activity carbonates, for the HR-3/-4 and HR-7/-8 transplant experiments. Transplant-to-active carbonates were not included due to their low sample number (n=1 for each of HR-3/-4 and HR-7/-8). (D-I) OTUs observed in native carbonate samples vs colonized carbonate samples as a function of Hydrate Ridge station. Stations were included only if they received colonization carbonate deployments *and* we had recovered native carbonates from the same station (these criteria excluded HR-1, HR-2, HR-6, HR-11, and the Southeast Knoll). Left column (red, D-F) are active stations, right column (blue, G-I) are low-activity stations. In each case, the bold color represents the native carbonates and the pale color represents the colonized carbonates. The number in each region denotes the number of OTUs, and font and circle sizes are proportional to OTU count. In most cases, the majority of recovered OTUs from colonization carbonates were also present in native carbonates.

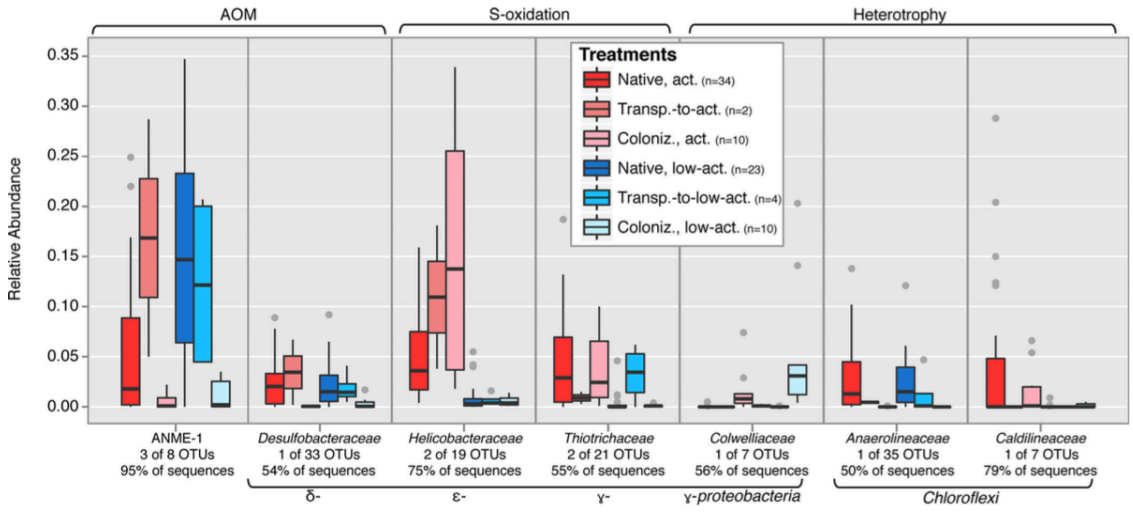


Fig. 5. Boxplot of carbonate-associated relative abundance data of selected key OTUs identified by SIMPER, representing notable taxonomic groups. Note that data for some groups are combined from several OTUs (OTU data are reported individually in Table S3). Although in all cases a minority of the OTUs were identified for presentation (e.g., 3 of 8 for ANME-1), these generally represented the majority of the total sequences recovered among each taxonomy (e.g., 95% of all sequences for ANME-1). When generated with all OTUs associated with each taxa, this plot does not change substantially (data not shown).

2.8 SUPPLEMENTAL MATERIAL: TEXT

2.8.1 SEEP STATION CHARACTERIZATION: ACTIVE vs LOW-ACTIVITY STATION DESIGNATIONS

Higher seep activity is associated with a higher delivery of reduced fluids, including methane and sulfide, from the subsurface. Although common geochemical measurements such as sulfide concentration are difficult to recover from directly below seep carbonates, pore water sulfide concentrations from nearby 0-3 centimeter-below-seafloor (cmbsf) horizons of sediment cores reveal a wide range of concentrations at our active stations (1-14 mM range, 6 mM average, n=9, data from HR-3, -7, -9, and -V1), whereas 0-3 cmbsf horizons of low activity stations contained <1 mM sulfide (0.2 mM average, n=5, data from HR-4 and -V2) (raw data unpublished). Station HR-V2 did not contain any samples used for analysis in this study, so will not be found in Table S1, but represented a seafloor low-activity station not unlike other low-activity stations in this study. Station HR-V2 was located at 44.570374°N, 125.14683°W (HR-South).

Those measured sulfide concentrations are consistent with previous studies of methane seeps (Gieskes et al. 2005; Green-Saxena et al. 2014; Pop Ristova et al. 2015), corroborate our active vs low activity designations, and also support published measurements of AOM in low-activity seep carbonates and sediments (Marlow et al. 2014a). Carbonates have been shown to adsorb methane, resulting in concentrations of methane 100-fold higher than the surrounding sediment (Ijiri et al. 2009). Therefore in areas of active seepage, carbonates might be a particularly methane-replete habitat substrate, magnifying the contrast with methane-depleted conditions in sediments from low-activity seepage areas.

2.8.2 X-RAY DIFFRACTION (XRD)

Among the 57 native seep carbonate samples in this study, 56 were mineralogically characterized (1 lacked enough material for XRD analysis). All six and all twenty of the transplant and colonization carbonates, respectively, were also characterized by XRD. Carbonate powders were analyzed on a Phillips X'Pert Multi Purpose X-Ray Diffractometer in the Division of Applied Physics and Materials Science at Caltech. Measurements were taken from 10˚ to 70˚ 2θ with step size of 0.05˚. SiO_2 standards were run to confirm peak location accuracy. For correlation with microbial assemblage data, samples were coarsely binned into groups based on the presence/absence of diagnostic (104), (221), and (104) peaks for calcite, aragonite, and dolomite respectively (Kontoyannis and Vagenas 2000; Zhang et al. 2010; Marlow et al. 2014b). The results of this mineralogical binning are reported in Table S1, and visualized in Fig. S5 for the native carbonates.

We note a qualitative, apparent correlation between mineralogy and seep station activity (Fig. S4). Low-activity stations are dominated by aragonite-bearing carbonate mineralogies, while active seep stations demonstrate higher carbonate mineralogical variability, including a large proportion of samples found to contain dolomite. Related to this qualitative correlation, we find a strong statistical differentiation of the native carbonate-associated microbial assemblages according to both seep activity and mineralogy (Table S2, see R values). While being sure to note that both seep activity and mineralogy differentiate the native carbonate dataset in the main text of the manuscript, we chose to focus the discussion around the indicator variable of seep activity because it was according to this environmental variable that we explicitly deployed the transplant and colonization experiments.

The links between microbial assemblage, seep activity, and mineralogy remain unresolved and an active area of research. Addressing and illuminating such links is not in the scope of this particular study, although it continues to be a focus of inquiry in our research

program. Broadly, seep activity and mineralogy are hypothesized to be linked to sulfate inhibition of Mg-bearing carbonate phases during precipitation (Peckmann et al. 2001; Naehr et al. 2007; Krause et al. 2012; Bian et al. 2013). However, a variety of outstanding questions remain: Do new carbonates precipitate at low-activity sites even after "high" levels of seepage disappear? What role does secondary alteration play in seep carbonates, and on what timescale? Is carbonate mineralogy further linked to depth and/or microbial regime in the sediment column, and how can this be interpreted or inferred by collecting samples directly from the seafloor? Do particular microorganisms prefer particular carbonate mineralogies, or is the biological differentiation we observe in our dataset mechanistically linked to other environmental parameters (i.e., seep activity)? If microorganisms demonstrate carbonate mineralogical preferences, is this due to chemical or physical aspects of mineralogy?

2.8.3 TRANSPLANTATION AND COLONIZATION DESCRIPTIONS

Exposed carbonates were selected on the seafloor during DSV Alvin dives AD4630, '31, '32 and '34 at HR-North in August 2010 for transplantation. Selected carbonates were picked up with the submersible's arm, temporarily deposited in the payload chassis, and deployed at selected destinations with a marked, weighted stake. The time-intensive nature of this effort, coupled with the difficulty of irrefutably re-identifying transplanted carbonates for recovery 13 months later, limited the total sample number of transplantation experiments. Ultimately, four carbonates were transplanted from active to low activity sites and subsequently recovered, and two from low activity to active sites (two other carbonates which were transplanted from low activity to active stations were unable to be successfully recovered). At each site, we established experimental control carbonates by using the DSV Alvin's manipulator arm to pick up native carbonates from the seafloor, and simply re-depositing them on the seafloor in their original location (identified in Table S1). Upon sequencing, these controls were indistinguishable from other native seep

carbonates (Table S2). Transplanted carbonates and controls were recovered in individual chassis compartments after 13 months during R/V Atlantis cruise 18-10 (September 2011), and frozen at -80˚C upon collection.

Since transplanted carbonates were moved from active- to low activity stations and vice versa, the effect of transplantation can be evaluated by comparing microbial assemblage similarity between the transplanted carbonates and the native, *in situ* carbonates binned into active and low-activity groups. This requires the assumption that the transplanted carbonates hosted an initial assemblage of microorganisms consistent with the characteristic assemblages associated with the native, active and native, low-activity carbonates in our study. This assumption is considered reasonable given the high sampling depth of native carbonates in this study and the strong difference between native microbial assemblages according to seep activity.

During a prior cruise in 2006, massive calcite and dolomite seep carbonates were recovered from Eel River Basin. After dry room temperature storage at Caltech for multiple years, the calcite and dolomite were each subsampled into 10 pieces of approximately 10^3 cm^3. These subsamples were autoclaved, stored aseptically, and brought to sea during R/V Atlantis cruise 15-68 in August 2010. XRD spectra of pre- and post-autoclaving mineral chips indicated no change in structure after sterilization. Pairs of one calcite and one dolomite subsample were deployed in a mesh bag on the seafloor at selected active and low-activity stations at Hydrate Ridge, adjacent to the transplant experimental rocks (Table S1). After 13 months of colonization, these carbonates were recovered in September 2011 during R/V Atlantis cruise 18-10. DNA extraction and PCR were unsuccessful on pre-colonization carbonate negative controls. If historic, remnant DNA were preserved in the colonization carbonates, and subsequently recovered during post-experimental DNA extraction and sequencing, then the colonization samples would not be expected to cluster according to seep activity and, most likely, would cluster within the suite of native carbonates in Fig. 1B and Fig. S3B. Neither of those null hypotheses are

true, further suggesting genuine deposition of new DNA material during the 13-month colonization experiments.

Twenty-six wood samples were also deployed for colonization during the same 13-month time interval. Natural pine (*Pinus* sp.) and natural douglas fir (*Pseudotsuga menziesi*) samples with bark were collected from forests in southwest Washington state, and non-treated wood blocks of douglas fir were purchased in San Diego, CA. Wood samples were deployed at sea in the same manner and at many of the same stations as the carbonate colonization experiments, but were not autoclaved prior to emplacement *in situ*. We are confident that our recovered 16S sequences from the wood experiments reflect genuine colonization (not microbes present prior to deployment) because the wood samples show biological differentiation according to seep activity (Fig. 1C), which would not be predicted from the microbial communities prior to being placed at active and low-activity stations. Furthermore, if any indigenous microbes were living in or on the wood, they were placed in a deep seafloor environment that would not be conducive to their terrestrial origins and growth. Finally, the recovered sequences from the wood samples are similar in composition to those observed in previous deep-sea wood colonization experiments (see Discussion in the main text)

2.8.4 PROCESSING OF MiSeq DATA

Raw sequence data for all 134 samples in this study can be accessed in the Sequence Read Archive under accession numbers SRP055767 and SRP049675. A sequential list of the in-house QIIME.1.8.0 commands used to process the data is given below.

1. Joining paired ends to generate contigs

```
join_paired_ends.py -f <sample_name_R1.fastq> -r
<sample_name_R2.fastq> -m fastq-join -j 50 -p 8 -o
<sample_name>_joined/
```

2. Quality trimming, converting from fastq to fasta format

```
split_libraries_fastq.py -i
<sample_name>_joined/fastqjoin.join.fastq -o
<sample_name>_trimmed/ -m dummy_mapping.txt --sample_id
<sample_name> -q 29 -n 0 --barcode_type 'not-barcoded' --
store_qual_scores
```

3. Chimera checking

```
usearch6.0.203_i86linux32 -uchime_ref
<sample_name>_trimmed/seqs.fna -db
SSURef_NR99_Silva_115_pintailF_ORPHAN.fasta -uchimeout
<sample_name>_chimerachecked/results.uchime --strand plus -
chimeras <sample_name>_chimerachecked/uchime_chimeras.fna -
nonchimeras <sample_name>_chimerachecked/uchime_nonchimeras.fna
```

4. Concatenating all sequences from all samples

```
cat *checked/uchime_nonchimeras.fna > all_seqs_all_samples.fasta
```

5. Picking OTUs at 97% similarity

```
pick_otus.py -i all_seqs_all_samples.fasta -s 0.97 -o
uclust_picked_otus_97/
```

6. Picking representative sequences for each OTU

```
pick_rep_set.py -i
uclust_picked_otus_97/all_seqs_all_samples_otus.txt -f
all_seqs_all_samples.fasta -m most_abundant
```

7. Assigning taxonomy for each OTU

```
assign_taxonomy.py -i all_seqs_all_samples.fasta_rep_set.fasta -t
SSURef_NR99_Silva_115_pintailF_ORPHAN.tax -r
SSURef_NR99_Silva_115_pintailF_ORPHAN.fasta --uclust_similarity
0.9 --uclust_max_accepts 10 --uclust_min_consensus_fraction 0.90
-o uclust_picked_otus_97/uclust_taxa_0.9_10_0.90/
```

8. Making OTU table in biom format

```
make_otu_table.py -i
uclust_picked_otus_97/all_seqs_all_samples_otus.txt -t
uclust_picked_otus_97/uclust_taxa_0.9_10_0.90/all_seqs_all_sample
s.fasta_rep_set_tax_assignments.txt -o
uclust_picked_otus_97/uclust_taxa_0.9_10_0.90/OTU_table_Silva_115
_all_seqs.biom
```

9. Removing singleton OTUs from the dataset

```
filter_otus_from_otu_table.py -i
uclust_picked_otus_97/uclust_taxa_0.9_10_0.90/OTU_table_Silva_115
_all_seqs.biom -n 2 -o
uclust_picked_otus_97/uclust_taxa_0.9_10_0.90/OTU_table_singleton
filtered.biom
```

10. Filtering known PCR contaminant taxa from the dataset by taxonomic name

```
filter_taxa_from_otu_table.py -i
uclust_picked_otus_97/uclust_taxa_0.9_10_0.90/OTU_table_singleton
filtered.biom -o
uclust_picked_otus_97/uclust_taxa_0.9_10_0.90/OTU_table_singleton
_taxafiltered.biom -n
Unassigned,Eukaryota,__Enterobacteriaceae,__Streptococcaceae,__Ps
eudomonadaceae,__Moraxellaceae,__Oxalobacteraceae
```

11. Generating Excel-readable OTU table with taxonomic information (*employs an in-house perl command written by Dr. Connor Skennerton)

```
join_otu_repset.pl
uclust_picked_otus_97/uclust_taxa_0.9_10_0.90/OTU_table_singleton
_taxafiltered.biom all_seqs_all_samples.fasta_rep_set.fasta >
OTU_table_wTaxa_wSeqs_97.txt
```

12. The Excel-readable table from Step 11 was used to identify the OTUs which represent poorly-defined Gammaproteobacteria (defined by QIIME as simply Bacteria;__Proteobacteria;__Gammaproteobacteria), and which have been shown to be PCR contaminants in internal lab control tests. This list of OTUs was then used in this step of filtering in QIIME. We could not simply remove "__Gammaproteobacteria" in Step 10 because it would have resulted in the loss of OTUs which represent genuine Gammaproteobacteria in our dataset. These poorly-defined Gammaproteobacteria represented an average of 4%±5% of the sequences per sample in our dataset.

```
filter_otus_from_otu_table.py -i
uclust_picked_otus_97/uclust_taxa_0.9_10_0.90/OTU_table_singleton
_taxafiltered.biom -e OTUs_to_exclude_GammaF_97.txt -o
uclust_picked_otus_97/uclust_taxa_0.9_10_0.90/OTU_table_singleton
_taxa_Gammafiltered.biom
```

13. Filtering OTUs which appear at less than 0.01% relative abundance in the entire dataset

```
filter_otus_from_otu_table.py -i
uclust_picked_otus_97/uclust_taxa_0.9_10_0.90/OTU_table_singleton
_taxa_Gammafiltered.biom --min_count_fraction 0.0001 -o
uclust_picked_otus_97/uclust_taxa_0.9_10_0.90/OTU_table_singleton
_taxa_Gamma_10000filtered.biom
```

14. Counting number of remaining sequences per sample to identify the level for rarefaction (found to be 16,051 sequences/sample, equal to the smallest sample)

```
biom summarize-table -i
uclust_picked_otus_97/uclust_taxa_0.9_10_0.90/OTU_table_singleton
_taxa_Gamma_10000filtered.biom -o
OTU_table_singleton_taxa_Gamma_10000filtered_SeqSummary.txt
```

15. Rarefying all samples down to the minimum number of sequences per sample

```
single_rarefaction.py -i
uclust_picked_otus_97/uclust_taxa_0.9_10_0.90/OTU_table_singleton
_taxa_Gamma_10000filtered.biom -d 16051 -o
uclust_picked_otus_97/uclust_taxa_0.9_10_0.90/OTU_table_singleton
_taxa_Gamma_10000_rarefied.biom
```

16. Generating Excel-readable OTU table for downstream alpha and beta diversity calculations

```
biom convert -i
uclust_picked_otus_97/uclust_taxa_0.9_10_0.90/OTU_table_singleton
_taxa_Gamma_10000_rarefied.biom -o
uclust_picked_otus_97/uclust_taxa_0.9_10_0.90/OTU_table_singleton
_taxa_Gamma_10000_rarefied.txt -b
```

2.8.5 NMDS, ANOSIM, & SIMPER TESTS IN R

Generic examples of the three commands are given below, including the options employed for this study. All commands require installation of the 'vegan' package in R, as well as two dependencies: 'lattice' and 'permute'. For this study, the following package versions were employed:

'vegan': v2.0-10
'lattice': v0.20-29
'permute': v0.8-3

1. NMDS

```
NMDS_analysis=metaMDS(sqrt_transformed_data,distance="bray",k=2,t
rymax=100,engine=c("monoMDS"),autotransform=FALSE)
```

2. ANOSIM

```
ANOSIM_analysis=with(sample_metadata,anosim(sqrt_transformed_data
,Activity,permutations=999,distance="bray"))
```

3. SIMPER

```
SIMPER_analysis=with(sample_metadata,simper(sqrt_transformed_data
,Activity))
```

2.9 SUPPLEMENTARY MATERIAL: TABLES

Supplementary Table 1-1. All samples in this study are listed with their accompanying metadata. Bold boxes denote the three different experimental treatments in the study: native samples, transplantation samples, and colonization samples.

Sample Number	Experimental Treatment	Habitat Substrate	Activity[c]	Regional Geography[a]	Specific Hydrate Ridge Geography	Local Geography	Latitude (Decimal °N)	Longitude (Decimal °W)	Depth (mbsl)	Habitat Substrate Sub-type[d]
5471		Bottom Water	Active	HR	South	HR-9	44.568466	125.152698	770	n.a.
5472			Low Activity	HR	South	off-HR-9	44.568511	125.151905	794	n.a.
3439		Carbonate	Active	HR	South	HR-9	44.568457	125.152761	775	A
3781				HR	South	HR-11	44.567899	125.153101	795	A
5357				HR	South	HR-9	44.568367	125.152774	774	A
5359				HR	South	HR-11	44.567782	125.153227	798	A
5436				HR	South	HR-11	44.567845	125.153038	794	A
2841				CR	n.a.	Mound 12	8.930590	84.312599	997	AC
2933				CR	n.a.	Mound 11	8.923242	84.303747	1010	AC
3138				CR	n.a.	Mound 12	8.930569	84.312839	996	AC
3530				HR	North	HR-3	44.669544	125.098057	587	AC
3626				HR	North	HR-7	44.667151	125.100020	602	AC
5330				HR	South	HR-V1	44.570239	125.147023	775	AC
5434				HR	South	HR-11	44.567890	125.153063	794	AC
5122[b]				HR	North	HR-7	44.667079	125.100033	601	AC
2781				CR	n.a.	Mound 12	8.929776	84.310799	989	ACD
3007				CR	n.a.	Quepos Mound	9.031787	84.621334	1402	C
3531				HR	North	HR-3	44.669544	125.098057	587	C
3532				HR	North	HR-3	44.669544	125.098057	587	C
3623				HR	North	HR-3	44.669499	125.098220	588	C
3544				HR	North	HR-7	44.667115	125.099995	602	CD
3602				HR	Southeast Knoll	n.a.	44.447635	125.028390	626	CD
3622				HR	North	HR-3	44.669499	125.098220	588	CD
3624				HR	North	HR-3	44.669499	125.098220	588	CD
3625				HR	North	HR-7	44.667151	125.100020	602	CD
3628				HR	North	HR-7	44.667151	125.100020	602	CD
5102				HR	North	HR-7	44.667106	125.099970	600	CD
5103				HR	North	HR-7	44.667106	125.099970	600	CD
5109[b]				HR	North	HR-3	44.669463	125.098120	587	CD
5110[b]				HR	North	HR-3	44.669481	125.098095	587	CD
5112[b]				HR	North	HR-3	44.669481	125.098107	587	CD
5123[b]				HR	North	HR-7	44.667079	125.100033	601	CD
3502				HR	North	HR-3	44.669535	125.098158	587	D
3603				HR	Southeast Knoll	n.a.	44.451163	125.027987	618	D
5120[b]				HR	North	HR-7	44.667079	125.100020	601	D
3665				HR	South	HR-9	44.568421	125.152786	775	n.m.
3079			Low Act.	CR	n.a.	Jaco Scarp	9.172603	84.798496	739	A
3599				HR	Southeast Knoll	n.a.	44.450632	125.029258	612	A
3662				HR	South	HR-10	44.568223	125.152987	778	A
5366				HR	South	HR-10	44.568178	125.152887	787	A
5038[b]				HR	North	HR-4	44.670075	125.098674	595	A
2874	Native			CR	n.a.	Mound 12	8.929939	84.310636	987	AC
2875				CR	n.a.	Mound 12	8.929939	84.312371	995	AC
2876				CR	n.a.	Mound 12	8.929866	84.313026	996	AC
3078				CR	n.a.	Jaco Scarp	9.172603	84.798496	739	AC
3464				HR	North	HR-4	44.670120	125.098686	592	AC
3511				HR	North	HR-5	44.669382	125.103619	620	AC
3541				HR	North	HR-8	44.667726	125.100838	603	AC
3543				HR	North	HR-8	44.667744	125.100926	602	AC
3627				HR	North	HR-8	44.667546	125.100763	603	AC
5153				HR	North	HR-8	44.667582	125.100712	602	AC
5040[b]				HR	North	HR-4	44.670075	125.098674	595	AC
5189[b]				HR	North	HR-8	44.667645	125.100712	604	AC
5190[b]				HR	North	HR-8	44.667645	125.100712	604	AC
3542				HR	North	HR-8	44.667726	125.100838	603	ACD
3545				HR	North	HR-8	44.667744	125.100926	602	ACD
3457				HR	North	HR-4	44.670057	125.098749	595	CD
3458				HR	North	HR-4	44.670057	125.098749	595	D
3604				HR	Southeast Knoll	n.a.	44.448670	125.030466	632	D
C2517[e]		Nodule	Active	HR	South	HR-V1	44.570059	125.147348	800	AC
C2518[e]				HR	South	HR-V1	44.570059	125.147348	800	AC
C2519[e]				HR	South	HR-V1	44.570059	125.147348	800	AC
C2521[e]				HR	South	HR-V1	44.570059	125.147348	800	C
C2689[e]				ERB	n.a.	n.a.	40.811495	124.610823	520	C
C2688[e]				ERB	n.a.	n.a.	40.811495	124.610823	520	CD
5118N				HR	North	HR-7	44.667061	125.099970	600	n.m.
C2520[e]				HR	South	HR-V1	44.570059	125.147348	800	n.m.
C2703[e]			Low Activity	ERB	n.a.	n.a.	40.811495	124.610823	520	CD
C2693[e]			Off-seep	ERB	n.a.	n.a.	40.811495	124.610823	520	C
2686		Sediment	Active	ERB	n.a.	n.a.	40.811495	124.610823	520	n.a.
2687				ERB	n.a.	n.a.	40.811495	124.610823	520	n.a.
5118				HR	North	HR-7	44.667061	125.099970	600	n.a.
S2517[e]				HR	South	HR-V1	44.570059	125.147348	800	n.a.
S2518[e]				HR	South	HR-V1	44.570059	125.147348	800	n.a.
S2519[e]				HR	South	HR-V1	44.570059	125.147348	800	n.a.
S2520[e]				HR	South	HR-V1	44.570059	125.147348	800	n.a.
S2521[e]				HR	South	HR-V1	44.570059	125.147348	800	n.a.
S2688[e]				ERB	n.a.	n.a.	40.811495	124.610823	520	n.a.
S2689[e]				ERB	n.a.	n.a.	40.811495	124.610823	520	n.a.
5163			Low Act.	HR	North	HR-8	44.667447	125.100486	601	n.a.
S2703[e]				ERB	n.a.	n.a.	40.811495	124.610823	520	n.a.
S2693[e]			Off-seep	ERB	n.a.	n.a.	40.811495	124.610823	520	n.a.

Supplementary Table 1-2. All samples in this study are listed with their accompanying metadata. Bold boxes denote the three different experimental treatments in the study: native samples, transplantation samples, and colonization samples.

Sample Number	Experimental Treatment	Habitat Substrate	Activity[c]	Regional Geography[a]	Specific Hydrate Ridge Geography	Local Geography	Latitude (Decimal °N)	Longitude (Decimal °W)	Depth (mbsl)	Habitat Substrate Sub-type[d]
5111	Transplantation	Carbonate	Low Act. -> Act.	HR	North	HR-3	44.669463	125.098120	587	AC
5121				HR	North	HR-7	44.667079	125.100020	601	AC
5193			Act. -> Low Act.	HR	North	HR-8	44.667627	125.100712	603	AC
5039				HR	North	HR-4	44.670093	125.098699	595	CD
5093				HR	North	HR-4	44.670084	125.098686	595	CD
5194				HR	North	HR-8	44.667627	125.100712	603	CD
5025c	Colonization	Carbonate	Active	HR	North	HR-3	44.669454	125.098145	588	C
5104c				HR	North	HR-7	44.667088	125.099995	600	C
5186c				HR	North	HR-6	44.668842	125.108653	613	C
5296c				HR	South	HR-9	44.568412	125.152786	774	C
5302c				HR	South	HR-1	44.568484	125.152635	774	C
5025d				HR	North	HR-3	44.669454	125.098145	588	D
5104d				HR	North	HR-7	44.667088	125.099995	600	D
5186d				HR	North	HR-6	44.668842	125.108653	613	D
5296d				HR	South	HR-9	44.568412	125.152786	774	D
5302d				HR	South	HR-1	44.568484	125.152635	774	D
5127c			Low Act.	HR	North	HR-8	44.667636	125.100737	602	C
5145c				HR	North	HR-4	44.670075	125.098661	600	C
5301c				HR	South	HR-2	44.570302	125.152786	810	C
5308c				HR	North	HR-5	44.669535	125.103946	618	C
5358c				HR	South	HR-10	44.568187	125.152899	788	C
5127d				HR	North	HR-8	44.667636	125.100737	602	D
5145d				HR	North	HR-4	44.670075	125.098661	600	D
5301d				HR	South	HR-2	44.570302	125.152786	810	D
5308d				HR	North	HR-5	44.669535	125.103946	618	D
5358d				HR	South	HR-10	44.568187	125.152899	788	D
5036		Wood	Active	HR	North	HR-3	44.669454	125.098145	588	DougFir
5106				HR	North	HR-7	44.667088	125.099995	600	DougFir
5187				HR	North	HR-6	44.668833	125.108653	613	DougFir
5295				HR	South	HR-9	44.568466	125.152824	775	DougFir
5298				HR	South	HR-1	44.568511	125.152648	774	DougFir
5368				HR	South	HR-11	44.567773	125.153277	800	DougFir
5030				HR	North	HR-3	44.669445	125.098132	588	NatFir
5108				HR	North	HR-7	44.667088	125.099995	600	NatFir
5188				HR	North	HR-6	44.668833	125.108678	613	NatFir
5293				HR	South	HR-9	44.568466	125.152824	775	NatFir
5037				HR	North	HR-3	44.669445	125.098145	588	NatPine
5107				HR	North	HR-7	44.667088	125.099995	600	NatPine
5312				HR	North	HR-6	44.668842	125.108615	615	NatPine
5097			Low Act.	HR	North	HR-4	44.670075	125.098674	595	DougFir
5126				HR	North	HR-8	44.667636	125.100737	601	DougFir
5211				HR	North	HR-5	44.669562	125.103934	618	DougFir
5304				HR	South	HR-2	44.570284	125.152849	810	DougFir
5361				HR	South	HR-12	44.568043	125.153013	792	DougFir
5364				HR	South	HR-10	44.568151	125.153038	788	DougFir
5095				HR	North	HR-4	44.670075	125.098674	595	NatFir
5192				HR	North	HR-8	44.667636	125.100700	604	NatFir
5309				HR	North	HR-5	44.669562	125.103934	618	NatFir
5363				HR	South	HR-10	44.568151	125.153038	788	NatFir
5096				HR	North	HR-4	44.670084	125.098674	595	NatPine
5191				HR	North	HR-8	44.667636	125.100700	604	NatPine
5310				HR	North	HR-5	44.669562	125.103934	618	NatPine

[a]HR=Hydrate Ridge; CR=Costa Rica; ERB=Eel River Basin

[b]These native samples also served as controls for the transplantation expereiments (see main text).

[c]For transplant experiment samples, the seepage activity at their final deployment location is listed.

[d]A=aragonite; AC=aragonite/calcite mixture; C=calcite, CD=calcite/dolomite mixture; D=dolomite; ACD=aragonte/calcite/dolomite mixture.

[e]Sediment and nodule samples previously examined in Mason et al., 2015.

Throughout table, n.a. indicates "not applicable" and n.m. indicates "not measured"

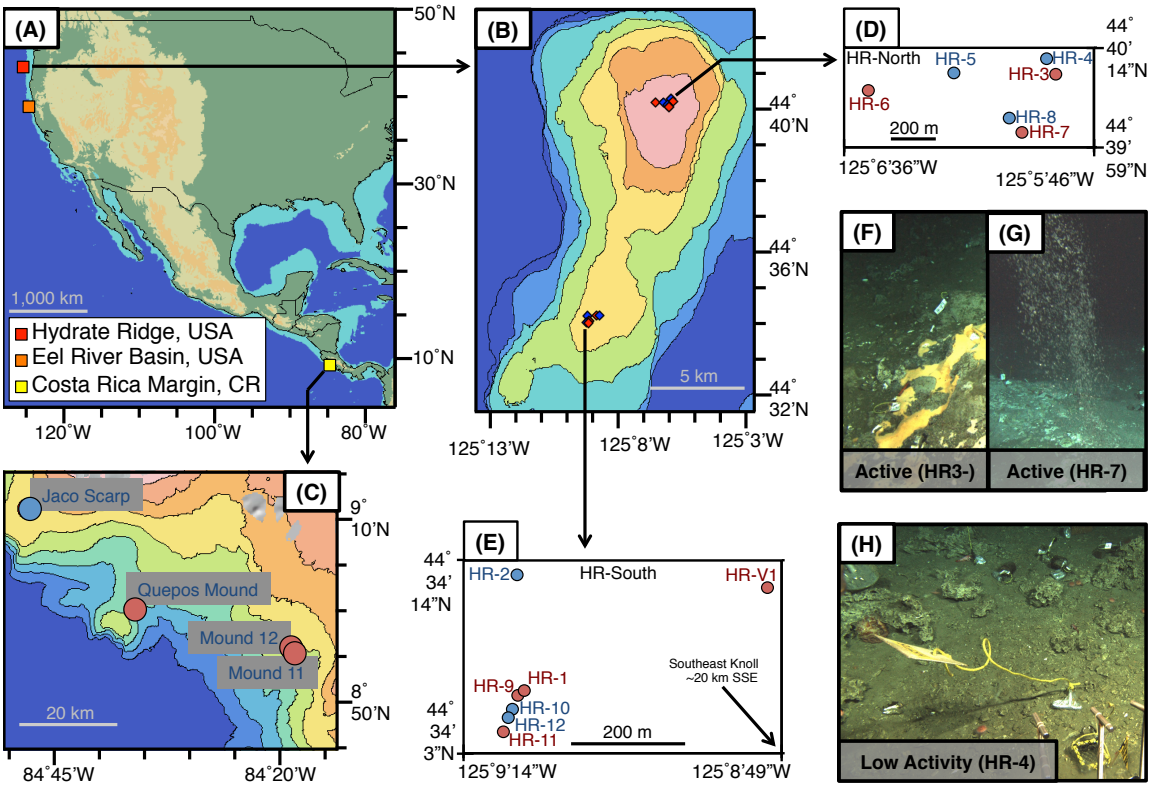
Supplementary Table 2. List of ANOSIM tests. Significance values (p-values) are colored green if <0.05 and red if >0.05.

Environmental Variable Tested	Sample Set	R	p
Habitat Substrate	Native sediments, nodules, carbonates, and bottom water	0.49	0.001
Activity		0.29	0.001
Habitat Substrate	Sediments and nodules	0.09	0.098
Habitat Substrate	Sediments and nodules (binned as one group) and native carbonates	0.49	0.001
Regional Geography	Sediments and nodules	0.51	0.001
Activity		0.43	0.009
Treatment	Transplant controls and native carbonates	-0.01	0.521
Activity	Native carbonates	0.45	0.001
1000s of km biogeography (Costa Rica *vs* Hydrate Ridge)		-0.05	0.713
Mineralogy		0.44	0.001
Mineralogy	Native, active carbonates	0.45	0.001
Mineralogy	Native, low activity carbonates	0.30	0.041
10s of km biogeography (HR-North *vs* HR-South *vs* SE-Knoll)	Native carbonates at Hydrate Ridge (excluding Costa Rica)	-0.06	0.788
Activity		0.50	0.001
Cruise Collected		-0.02	0.652
Specific Collection Site (e.g. HR-3)	Native, active carbonates at Hydrate Ridge (excluding Costa Rica)	0.31	0.002
Specific Collection Site (e.g. HR-4)	Native, low activity carbonates at Hydrate Ridge (excluding Costa Rica)	0.27	0.037
Experimental Treatment	Native and transplanted carbonates	0.49	0.001
Experimental Treatment	Native, active carbonates and transplant-to-low activity carbonates	0.32	0.008
Experimental Treatment	Native, low activity carbonates and transplant-to-low activity carbonates	0.88	0.001
Experimental Treatment	Native sediments, nodules, carbonates, and bottom water (binned as one group) and colonization carbonates and woods (binned as one group)	0.68	0.001
Experimental Treatment	Native carbonates and colonization carbonates	0.65	0.001
Experimental Treatment	Native carbonates and colonizatoin carbonates (Presence/Absence normalized)	0.53	0.001
Habitat Substrate	Colonization carbonates and woods	0.63	0.001
Activity		0.38	0.001
Activity	Colonization carbonates	0.81	0.001
Mineralogy		0.11	0.109
Activity	Colonization woods	0.39	0.001
Wood Type		0.22	0.008
Wood Type	Colonization woods, active	0.12	0.203
Wood Type	Colonization woods, low activity	0.56	0.002

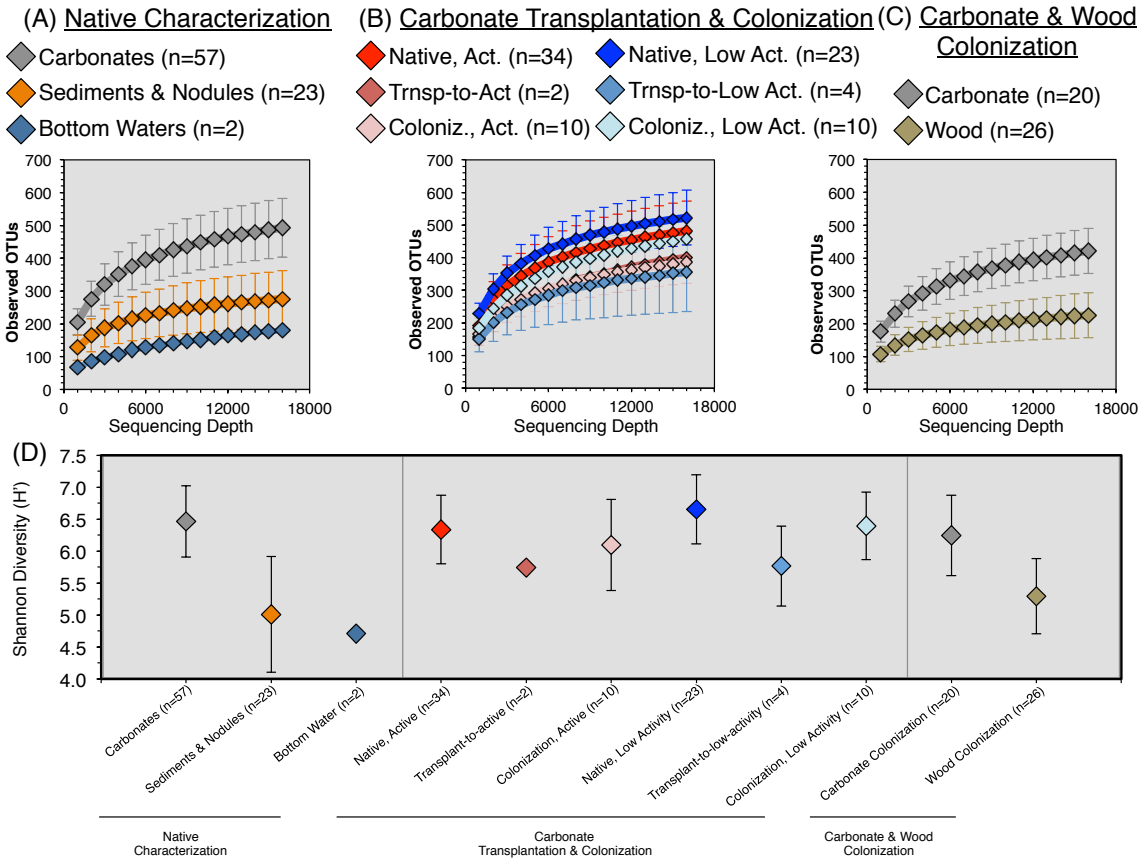
Supplementary Table 3. Relative abundance information is listed for OTUs presented in Fig. 3, Fig. 4, and Fig. 5 of the main text, for all 134 samples in this study. Data is provided in four tabs: (Tab 1) Raw data associated with Fig. 3; (Tab 2) Raw data associated with Fig. 4G; (Tab 3) Raw data associated with Fig. 4H; (Tab 4) Raw data ssociated with Fig. 5. For the data used to generate Fig. 5, in all cases a minority of the OTUs were identified for presentation (e.g., 3 of 8 for ANME-1). However, these represented the majority of the total sequences recovered among each taxonomy (e.g., 95% of all sequences for ANME-1). Bold boxes outline the sets of OTUs which were binned for presentation in Fig. 5. Colors scales are meant to aid the reader in assessing OTU distribution. OTUs of the same phylogeny were combined for presentation in Fig. 5 of the main text.

Supplementary Table 3 can be found in .xlsx format in the Caltech online repository along with this thesis.

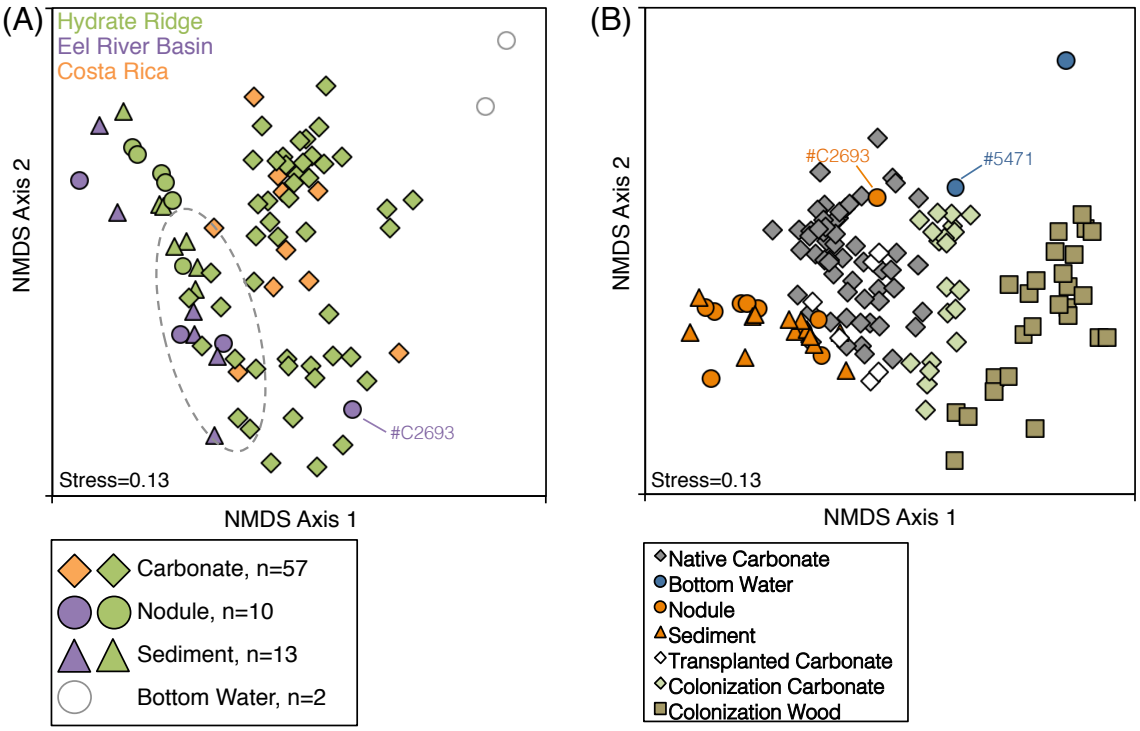
2.10 SUPPLEMENTAL MATERIAL: FIGURES



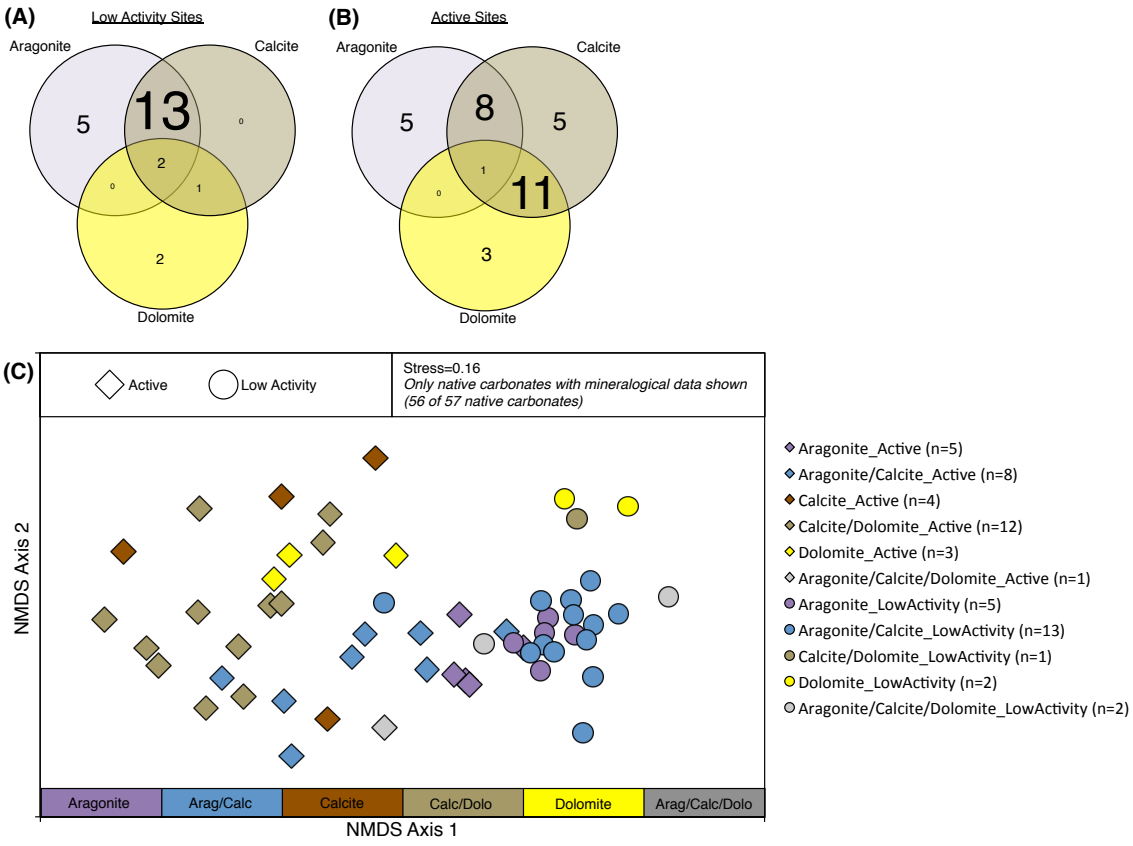
Supplementary Figure 1. Overview map of sampling locations in this study. For seep descriptions, see Section 2.1. (A) Overview the three seep locations sampled in this study, (B) map of Hydrate Ridge including sampling locations at HR-North and HR-South, (C) map of Costa Rica including the four mounds sampled for native carbonates, (D) map of HR-North showing spatial relationship between stations HR-3, -4, -5, -6, -7, and -8, (E) map of HR-South showing spatial relationship between stations HR-1, -2, -9, -10, -11, -12, and -V1, (F-G) seafloor images of active stations HR-3 and HR-7 exhibiting orange bacterial mats (HR-3), clam beds (HR-3 & HR-7), and methane ebullition (HR-7), (H) seafloor image of low-activity station HR-4 lacking the diagnostic indicators seen in (F-G). Subplots (A-C) generated in GeoMapApp (http://www.geomapapp.org, [Ryan et al., 2009]). For subplot (B), contour lines represent 100 m bathymetric relief with HR-North the shallowest at 700 mbsl. For subplot (C), contour lines represent 250 m bathymetric relief with the shallow continental shelf at 250 mbsl in the NE corner. Red and blue points plotted in subplots (B-E) represent active and low-activity stations, respectively.



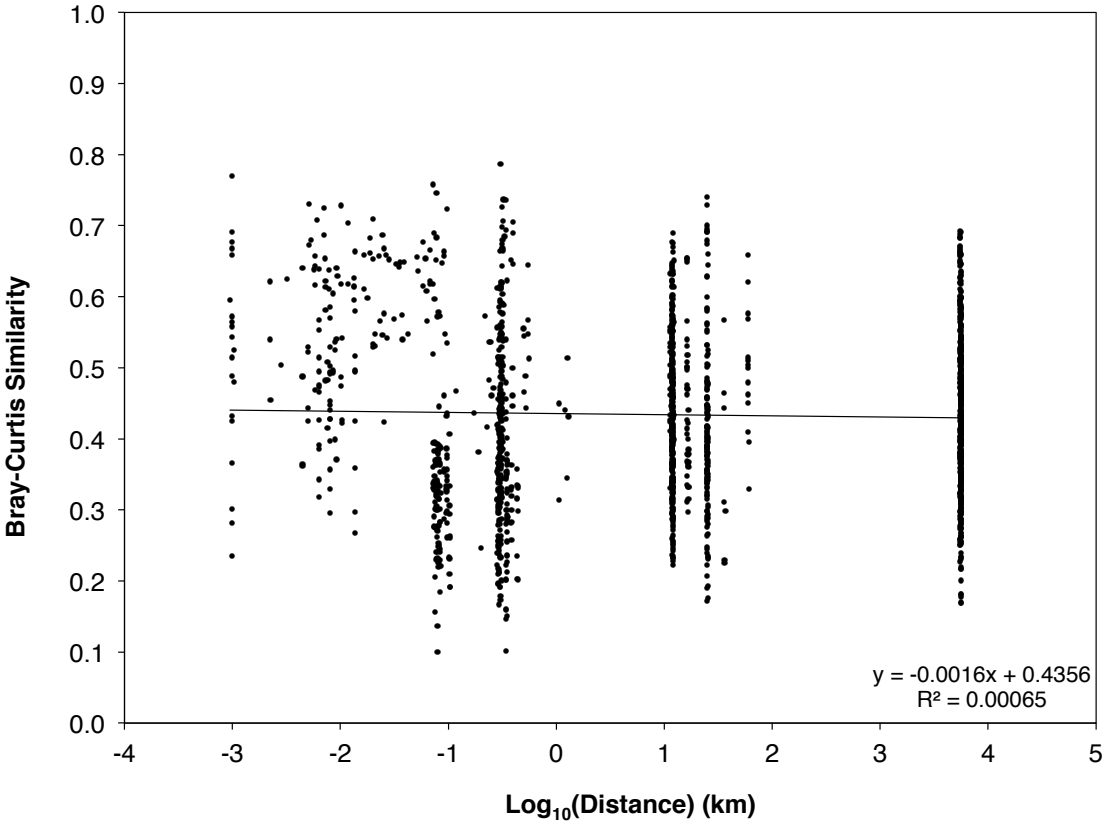
Supplementary Figure 2. Additional alpha diversity metrics for the samples in this study. (A-C) Collector's curves of raw OTU_{97}, for the native samples (A), carbonate samples (B), and colonization samples (C). (D) Shannon Diversity (H') for the major sample groups analyzed in this study. Error bars represent standard deviation among the sample groups, which is not included for the bottom water and transplant-to-active samples due to the low number of replicates.



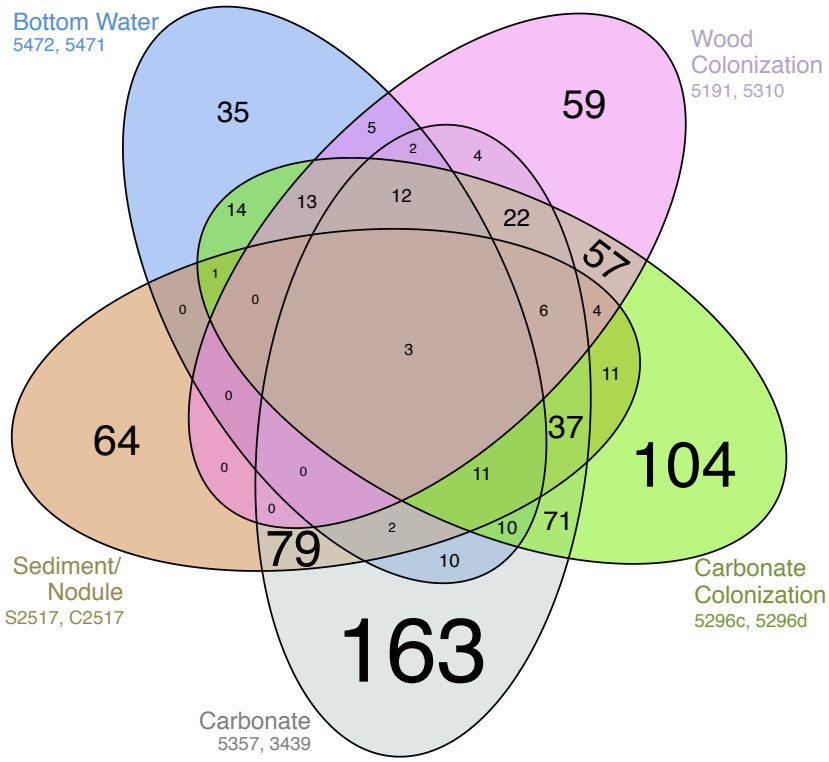
Supplementary Figure 3. Additional non-metric multidimensional scaling analyses (A) Ordination of the sample data from native carbonates, sediments, nodules, and bottom water. Ordination is identical to Fig. 1A, but the carbonates, sediments, and nodules are colored by geographic origin. The overlap between Costa Rica and Hydrate Ridge carbonates, in the same vicinity on the plot as Hydrate Ridge and Eel River Basin sediments and nodules, demonstrates that the region where habitat substrates overlap (roughly, the dashed circle) is not a geographic effect. (B) All 134 samples in this study. The same intra-sample relationships are visible as in Fig. 1. The wood colonization samples plot in a distinct region away from all other samples, though most similar to the colonization carbonates. Sample #C2693 is highlighted as a biological outlier among the nodule-hosted samples. Sample #5471 is highlighted because in this ordination plot of all samples, it appears biologically similar to some of the carbonate colonization samples. Sample #5471 was bottom water from an active seep site (Table S1), which may be related to its similarity to some colonization samples; more bottom water data points would be necessary in order to draw a stronger conclusion.



Supplementary Figure 4. Mineralogical Analysis. (A-B) Distribution of recovered carbonate mineralogies at low-activity (A) and active seep stations (B). Note that aragonite-bearing mineralogies are dominant at low-activity stations, while active seep stations host more varied mineralogies including a greater proportion of dolomite-bearing carbonates. In order to emphasize the distributions, text size is proportional to value. From this data, it appears that carbonate mineralogy may have a loose correlation with seep activity. From our dataset we are unable to determine whether mineralogy and seep activity are truly dependent or independent environmental variables (see Supplemental Text). (C) Non-metric multidimensional scaling analysis of microbial assemblages from native carbonate samples according to seep activity and mineralogy. Samples are separated according to seep activity (c.f. Fig. 1B), but such a difference is also qualitatively correlated with mineralogical differences.



Supplementary Figure 5. Distance-decay plot of the 57 native carbonates collected from Hydrate Ridge and Costa Rica. Geographical distance is calculated in km along a Great Circle between each sample pair (R package 'geosphere' v1.3-13, function 'distHaversine', assuming a spherical Earth of radius of 6,371 km). Bray-Curtis similarity is calculated as described in the main text, presented with a value of '1' being identical similarity, and '0' being complete dissimilarity. Note the y-axis is linear and the x-axis is logarithmic. No correlation is observed between Bray-Curtis similarity and geographic distance. Although this does not preclude such a correlation existing, it means that such a correlation may be aliased by influences of other environmental factors (e.g., seep activity).



Supplementary Figure 6. OTU overlap between colonized substrates at active site HR-9, including representative native substrates. Site HR-9 was chosen because it was the associated site for both bottom water samples, it hosted both carbonate and wood colonization experiments, it hosted multiple native carbonates which were recovered and analyzed, and it was an active seep site. Since sediment and nodule samples were not collected from HR-9, representatives of those substrates were chosen from another HR-South location (HR-V1); shallow (0-3 cmbsf) sediments and nodules were chosen because they were most likely to be relevant for colonization experiments which were performed on the seafloor. In order to keep the Venn diagram relatively simple, and because ANOSIM tests revealed sediments and nodules to be indistinguishable from one another, sediments and nodules were binned together as one “habitat type” for this analysis. In order to ensure equal depth of sampling across each substrate type, two representative samples of each substrate were chosen randomly (see sample numbers on the Figure). In order to emphasize the distributions, text size is proportional to OTU occurrence. For all OTU overlap diagrams, the specific samples analyzed are given in small text next to the diagram bubbles. In order for an OTU to be counted as “present” for a category (e.g., for an OTU to be associated with “Bottom Water”), it had to be present in *both* replicates. This approach ensured that the OTUs under examination were reproducibly recovered from every category, rather than spurious observations.

2.11 REFERENCES

Alain, K., M. Zbinden, N. Le Bris, F. Lesongeur, J. Quérellou, F. Gaill, and M. A. Cambon Bonavita. 2004. Early steps in microbial colonization processes at deep-sea hydrothermal vents. Environmental microbiology **6**: 227–241.

Bekins, B. A., and S. J. Dreiss. 1992. A simplified analysis of parameters controlling dewatering in accretionary prisms. Earth and Planetary Science Letters **109**: 275–287.

Bernard, C., and T. Fenchel. 1995. Mats of colourless sulphur bacteria. II. Structure, composition of biota and successional patterns. Marine Ecology Progress Series **128**: 171–179.

Bian, Y., D. Feng, H. H. Roberts, and D. Chen. 2013. Tracing the evolution of seep fluids from authigenic carbonates: Green Canyon, northern Gulf of Mexico. Marine and Petroleum Geology **44**: 71–81.

Bienhold, C., P. Pop Ristova, F. Wenzhöfer, T. Dittmar, and A. Boetius. 2013. How Deep-Sea Wood Falls Sustain Chemosynthetic Life. PLoS ONE **8**: e53590.

Birgel, D., T. Himmler, A. Freiwald, and J. Peckmann. 2008. A new constraint on the antiquity of anaerobic oxidation of methane: Late Pennsylvanian seep limestones from southern Namibia. Geology **36**: 543–546.

Blumenberg, M., E.-O. Walliser, M. Taviani, R. Seifert, and J. Reitner. 2015. Authigenic carbonate formation and its impact on the biomarker inventory at hydrocarbon seeps – a case study from the Holocene Black Sea and the Plio-Pleistocene Northern Apennines (Italy). Marine and Petroleum Geology **66**: 532–541.

Boetius, A., and E. Suess. 2004. Hydrate Ridge: a natural laboratory for the study of microbial life fueled by methane from near-surface gas hydrates. Chemical Geology **205**: 291–310.

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

Bowman, J. P., and R. D. McCuaig. 2003. Biodiversity, Community Structural Shifts, and Biogeography of Prokaryotes within Antarctic Continental Shelf Sediment. Applied and Environmental Microbiology **69**: 2463–2483.

Caporaso, J. G., C. L. Lauber, W. A. Walters, D. Berg-Lyons, C. A. Lozupone, P. J. Turnbaugh, N. Fierer, and R. Knight. 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proceedings of the National Academy of Sciences **108**: 4516–4522.

Caporaso, J. G., C. L. Lauber, W. A. Walters, D. Berg-Lyons, J. Huntley, N. Fierer, S. M. Owens, J. Betley, L. Fraser, M. Bauer, N. Gormley, J. A. Gilbert, G. Smith, and R. Knight. 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. The ISME Journal **6**: 1621–1624.

Chen, Z., W. Yan, M. Chen, S. Wang, J. Lu, F. Zhang, R. Xiang, S. Xiao, P. Yan, and S. Gu. 2006. Discovery of seep carbonate nodules as new evidence for gas venting on the northern continental slope of South China Sea. Chinese Science Bulletin **51**: 1228–1237.

Clarke, K. R., and R. M. Warwick. 2001. Change in Marine Communities, 2nd ed. PRIMER-E Ltd.

Dekas, A. E., G. L. Chadwick, M. W. Bowles, S. B. Joye, and V. J. Orphan. 2014. Spatial distribution of nitrogen fixation in methane seep sediment and the role of the ANME archaea. Environmental microbiology **16**: 3012–3029.

Distel, D. L., A. R. Baco, E. Chuang, W. Morrill, C. Cavanaugh, and C. R. Smith. 2000. Marine ecology: Do mussels take wooden steps to deep-sea vents? Nature **403**: 725–726.

Fagervold, S. K., C. Romano, D. Kalenitchenko, C. Borowski, A. Nunes-Jorge, D. Martin, and P. E. Galand. 2014. Microbial Communities in Sunken Wood Are Structured by Wood-Boring Bivalves and Location in a Submarine Canyon D.A. Carter [ed.]. PLoS ONE **9**: e96248.

Fagervold, S. K., P. E. Galand, M. Zbinden, F. Gaill, P. Lebaron, and C. Palacios. 2012. Sunken

woods on the ocean floor provide diverse specialized habitats for microorganisms. FEMS Microbiology Ecology **82**: 616–628.

Garrity, G. M., D. J. Brenner, N. R. Krieg, and J. T. Staley, eds. 2005. Bergey's Manual of Systematic Bacteriology, Volume Two: The Proteobacteria, Springer.

Gieskes, J., C. Mahn, S. Day, J. Martin, J. Greinert, T. Rathburn, and B. McAdoo. 2005. A study of the chemistry of pore fluids and authigenic carbonates in methane seep environments: Kodiak Trench, Hydrate Ridge, Monterey Bay, and Eel River Basin. Chemical Geology **220**: 329–345.

Gilbert, J. A., F. Meyer, J. Jansson, J. Gordon, N. R. Pace, J. M. Tiedje, R. E. Ley, N. Fierer, D. Field, N. C. Kyrpides, F. O. Gloeckner, H. P. Klenk, K. E. Wommack, E. Glass, K. Docherty, R. Gallery, R. Stevens, and R. Knight. 2011. The Earth Microbiome Project: Meeting report of the "1st EMP meeting on sample selection and acquisition" at Argonne National Laboratory October 6th 2010. 1–5.

Girguis, P., V. Orphan, S. Hallam, and E. F. DeLong. 2003. Growth and methane oxidation rates of anaerobic methanotrophic archaea in a continuous-flow bioreactor. Applied and Environmental Microbiology **69**: 5472–5482.

Green-Saxena, A., A. E. Dekas, N. F. Dalleska, and V. J. Orphan. 2014. Nitrate-based niche differentiation by distinct sulfate-reducing bacteria involved in the anaerobic oxidation of methane. The ISME Journal **8**: 150–163.

Greinert, J., G. Bohrmann, and E. Suess. 2001. Gas Hydrate-Associated Carbonates and Methane-Venting at Hydrate Ridge: Classification, Distribution, and Origin of Authigenic Lithologies, p. 99–113. *In* C.K. Paull and W.P. Dillon [eds.], Natural Gas Hydrates. American Geophysical Union.

Guilini, K., L. A. Levin, and A. Vanreusel. 2012. Cold seep and oxygen minimum zone associated sources of margin heterogeneity affect benthic assemblages, diversity and nutrition at the Cascadian margin (NE Pacific Ocean). Progress in Oceanography **96**: 77–92.

Heijs, S. K., G. Aloisi, I. Bouloubassi, R. D. Pancost, C. Pierre, J. S. Sinninghe Damsté, J. C. Gottschal, J. D. Elsas, and L. J. Forney. 2006. Microbial Community Structure in Three Deep-Sea Carbonate Crusts. Microbial Ecology **52**: 451–462.

Hinrichs, J. M. Hayes, S. P. Sylva, P. G. Brewer, and E. F. DeLong. 1999. Methane-consuming archaebacteria in marine sediments. Nature **398**: 802–805.

Hovland, M., M. R. Talbot, H. Qvale, S. Olaussen, and L. Aasberg. 1987. Methane-related Carbonate Cements in Pockmarks of the North Sea. Journal of Sedimentary Petrology **57**: 881–892.

Ijiri, A., U. Tsunogai, T. Gamo, F. Nakagawa, T. Sakamoto, and S. Saito. 2009a. Enrichment of adsorbed methane in authigenic carbonate concretions of the Japan Trench. Geo-Marine Letters **29**: 301–308.

Knittel, K., T. Lösekann, A. Boetius, R. Kort, and R. Amann. 2005. Diversity and Distribution of Methanotrophic Archaea at Cold Seeps. Applied and Environmental Microbiology **71**: 467–479.

Kontoyannis, C. G., and N. V. Vagenas. 2000. Calcium carbonate phase analysis using XRD and FT-Raman spectroscopy. The Analyst **125**: 251-255.

Krause, S., V. Liebetrau, S. Gorb, M. Sanchez-Roman, J. A. McKenzie, and T. Treude. 2012. Microbial nucleation of Mg-rich dolomite in exopolymeric substances under anoxic modern seawater salinity: New insight into an old enigma. Geology **40**: 587–590.

Levin, L. A. 2005. Ecology of cold seep sediments: Interactions of fauna with flow, chemistry, and microbes. Oceanography and Marine Biology: An Annual Review **43**: 1–46.

Levin, L. A., G. F. Mendoza, B. M. Grupe, J. P. Gonzalez, B. Jellison, G. W. Rouse, A. R. Thurber, and A. Waren. 2015. Biodiversity on the Rocks: Macrofauna Inhabiting Authigenic Carbonate at Costa Rica Methane Seeps. PLoS ONE 1–31.

Levin, L. A., G. F. Mendoza, J. P. Gonzalez, A. R. Thurber, and E. E. Cordes. 2010. Diversity of

bathyal macrofauna on the northeastern Pacific margin: the influence of methane seeps and oxygen minimum zones. Marine Ecology Progress Series **31**: 94–110.

Levin, L. A., V. J. Orphan, G. W. Rouse, A. E. Rathburn, W. Ussler, G. S. Cook, S. K. Goffredi, E. M. Perez, A. Waren, B. M. Grupe, G. Chadwick, and B. Strickrott. 2012. A hydrothermal seep on the Costa Rica margin: middle ground in a continuum of reducing ecosystems. Proceedings of the Royal Society of London B 1–9.

Levin, L. A., W. Ziebis, G. F. Mendoza, V. Growney, M. D. Tryon, K. M. Brown, C. Mahn, J. M. Gieskes, and A. E. Rathburn. 2003. Spatial heterogeneity of macrofauna at northern California methane seeps: influence of sulfide concentration and fluid flow. Marine Ecology Progress Series **265**: 123–139.

Li, L., C. Kato, and K. Horikoshi. 1999. Microbial Diversity in Sediments Collected from the Deepest Cold-Seep Area, the Japan Trench. Marine Biotechnology **1**: 391–400.

Marlow, J. J., J. A. Steele, W. Ziebis, A. R. Thurber, L. A. Levin, and V. J. Orphan. 2014a. Carbonate-hosted methanotrophy represents an unrecognized methane sink in the deep sea. Nature Communications **5**: 1–12.

Marlow, J., J. A. Steele, D. Case, S. A. Connon, L. A. Levin, and V. J. Orphan. 2014b. Microbial abundance and diversity patterns associated with sediments and carbonates from the methane seep environments of Hydrate Ridge, OR. Frontiers in Marine Science **1**: 1–16.

Marlow, J., J. Peckmann, and V. Orphan. 2015. Autoendoliths: a distinct type of rock-hosted microbial life. Geobiology **13**: 303–307.

Mason, O. U., D. H. Case, T. H. Naehr, R. W. Lee, R. B. Thomas, J. V. Bailey, and V. J. Orphan. 2015. Comparison of Archaeal and Bacterial Diversity in Methane Seep Carbonate Nodules and Host Sediments, Eel River Basin and Hydrate Ridge, USA. Microbial Ecology **70**: 776–784.

Morono, Y., T. Terada, M. Nishizawa, M. Ito, F. Hillion, N. Takahata, Y. Sano, and F. Inagaki. 2011. Carbon and nitrogen assimilation in deep subseafloor microbial cells. Sci. Dril. **108**: 18295–18300.

Naehr, T. H., P. Eichhubl, V. J. Orphan, M. Hovland, C. K. Paull, W. Ussler, T. D. Lorenson, and H. G. Greene. 2007. Authigenic carbonate formation at hydrocarbon seeps in continental margin sediments: a comparative study. Deep Sea Research Part II **54**: 1268–1291.

Niemann, H., P. Linke, K. Knittel, E. MacPherson, A. Boetius, W. Brückmann, G. Larvik, K. Wallmann, U. Schacht, E. Omoregie, D. Hilton, K. Brown, and G. Rehder. 2013. Methane-Carbon Flow into the Benthic Food Web at Cold Seeps – A Case Study from the Costa Rica Subduction Zone. PLoS ONE **8**: e74894.

Nunoura, T., Y. Takaki, H. Kazama, M. Hirai, J. Ashi, H. Imachi, and K. Takai. 2012. Microbial Diversity in Deep-sea Methane Seep Sediments Presented by SSU rRNA Gene Tag Sequencing. Microbes and Environments **27**: 382–390.

Orphan, V. J., Hinrichs, W. Ussler, C. K. Paull, L. T. Taylor, S. P. Sylva, J. M. Hayes, and E. F. Delong. 2001. Comparative Analysis of Methane-Oxidizing Archaea and Sulfate-Reducing Bacteria in Anoxic Marine Sediments. Applied and Environmental Microbiology **67**: 1922–1934.

Orphan, V. J., K. A. Turk, A. M. Green, and C. H. House. 2009. Patterns of 15N assimilation and growth of methanotrophic ANME-2 archaea and sulfate-reducing bacteria within structured syntrophic consortia revealed by FISH-SIMS. Environmental microbiology **11**: 1777–1791.

Orphan, V., W. Ussler, T. H. Naehr, C. H. House, Hinrichs, and C. K. Paull. 2004. Geological, geochemical, and microbiological heterogeneity of the seafloor around methane vents in the Eel River Basin, offshore California. Chemical Geology **205**: 265–289.

Peckmann, J., A. Reimer, C. Luth, C. Luth, B. T. Hansen, C. Heinicke, J. Hoefs, and J. Reitner. 2001. Methane-derived carbonates and authigenic pyrite from the northwestern Black Sea.

Marine Geology **177**: 129–150.

Pop Ristova, P., F. Wenzhöfer, A. Ramette, J. Felden, and A. Boetius. 2015. Spatial scales of bacterial community diversity at cold seeps (Eastern Mediterranean Sea). The ISME Journal **9**: 1306–1318.

R Core Team. 2014. R: A language and environment for statistical computing.

Rajala, P., M. Bomberg, R. Kietäväinen, I. Kukkonen, L. Ahonen, M. Nyyssönen, and M. Itävaara. 2015. Rapid Reactivation of Deep Subsurface Microbes in the Presence of C-1 Compounds. Microorganisms **3**: 17–33.

Reitner, J., J. Peckmann, A. Reimer, G. Schumann, and V. Thiel. 2005. Methane-derived carbonate build-ups and associated microbial communities at cold seeps on the lower Crimean shelf (Black Sea). Facies **51**: 66–79.

Rossel, P. E., M. Elvert, A. Rammette, A. Boetius, and Hinrichs. 2011. Factors controlling the distribution of anaerobic methanotrophic communities in marine environments: Evidence from intact polar membrane lipids. Geochimica et Cosmochimica Acta **75**: 164–184.

Ruff, S. E., J. Arnds, K. Knittel, R. Amann, G. Wegener, A. Ramette, and A. Boetius. 2013. Microbial Communities of Deep-Sea Methane Seeps at Hikurangi Continental Margin (New Zealand). PLoS ONE **8**: e72627.

Ruff, S. E., J. F. Biddle, A. P. Teske, K. Knittel, A. Boetius, and A. Ramette. 2015. Global dispersion and local diversification of the methane seep microbiome. Proceedings of the National Academy of Sciences 1–6.

Ryan, W. B. F., S. M. Carbotte, J. O. Coplan, S. O'Hara, A. Melkonian, R. Arko, R. A. Weissel, V. Ferrini, A. Goodwillie, F. Nitsche, J. Bonczkowski, and R. Zemsky. 2009. Global Multi-Resolution Topography synthesis. Geochemistry Geophysics Geosystems **10**: 1-9.

Sahling, H., D. G. Masson, C. R. Ranero, V. Hühnerbach, W. Weinrebe, I. Klaucke, D. Bürk, W. Brückmann, and E. Suess. 2008. Fluid seepage at the continental margin offshore Costa Rica and southern Nicaragua. Geochemistry Geophysics Geosystems **9**: 1–22.

Sahling, H., D. Rickert, R. W. Lee, P. Linke, and E. Suess. 2002. Macrofaunal community structure and sulfide flux at gas hydrate deposits from the Cascadia convergent margin, NE Pacific. Marine Ecology Progress Series **231**: 121–138.

Schauer, R., C. Bienhold, A. Ramette, and J. Harder. 2010. Bacterial diversity and biogeography in deep-sea surface sediments of the South Atlantic Ocean. The ISME Journal **4**: 159–170.

Schmale, O., I. Leifer, J. Schneider von Deimling, C. Stolle, S. Krause, K. Kießlich, A. Fram, and T. Treude. 2015. Bubble transport mechanism: Indications for a gas bubble-mediated inoculation of benthic methanothrophs into the water column. Continental Shelf Research **103**: 70–78.

Sievert, S. M., E. B. A. Wieringa, C. O. Wirsen, and C. D. Taylor. 2007. Growth and mechanism of filamentous-sulfur formation by Candidatus Arcobacter sulfidicus in opposing oxygen-sulfide gradients. Environmental microbiology **9**: 271–276.

Sorokin, D. Y., T. P. Tourova, E. Y. Bezsoudnova, A. Pol, and G. Muyzer. 2007. Denitrification in a binary culture and thiocyanate metabolism in Thiohalophilus thiocyanoxidans gen. nov. sp. nov. – a moderately halophilic chemolithoautotrophic sulfur-oxidizing Gammaproteobacterium from hypersaline lakes. Archives of microbiology **187**: 441–450.

Stadnitskaia, A., G. Muyzer, B. Abbas, M. J. L. Coolen, E. C. Hopmans, M. Baas, T. C. E. van Weering, M. K. Ivanov, E. Poludetkina, and J. S. Sinninghe Damsté. 2005. Biomarker and 16S rDNA evidence for anaerobic oxidation of methane and related carbonate precipitation in deep-sea mud volcanoes of the Sorokin Trough, Black Sea. Marine Geology **217**: 67–96.

Suess, E., B. Carson, S. D. Ritger, J. C. Moore, M. L. Jones, L. D. Kulm, and G. R. Cochrane. 1985. Biological communities at vent sites along the subduction zone off Oregon. Bulletin of the Biological Society of Washington **6**: 475–484.

Sylvan, J. B., B. C. Pyenson, O. Rouxel, C. R. German, and K. J. Edwards. 2012a. Time-series analysis of two hydrothermal plumes at 9°50′N East Pacific Rise reveals distinct,

heterogeneous bacterial populations. Geobiology **10**: 178–192.

Sylvan, J. B., B. M. Toner, and K. J. Edwards. 2012b. Life and Death of Deep-Sea Vents: Bacterial Diversity and Ecosystem Succession on Inactive Hydrothermal Sulfides. mBio **3**: e00279–11.

Tavormina, P., W. Ussler, S. Joye, B. Harrison, and V. Orphan. 2010. Distributions of putative aerobic methanotrophs in diverse pelagic marine environments. The ISME Journal **4**: 700–710.

Taylor, C. D., C. O. Wirsen, and F. Gaill. 1999. Rapid Microbial Production of Filamentous Sulfur Mats at Hydrothermal Vents. Applied and Environmental Microbiology **65**: 2253–2255.

Teichert, B., G. Bohrmann, and E. Suess. 2005. Chemoherms on Hydrate Ridge--Unique microbially-mediated carbonate build-ups growing into the water column. Palaeogeography, Palaeoclimatology, Palaeoecology **227**: 67–85.

Thiel, V., J. Peckmann, H. H. Richnow, U. Luth, J. Reitner, and W. Michaelis. 2001. Molecular signals for anaerobic methane oxidation in Black Sea seep carbonates and a microbial mat. Marine Chemistry **73**: 97–112.

Thurber, A. R., L. A. Levin, V. J. Orphan, and J. J. Marlow. 2012. Archaea in metazoan diets: implications for food webs and biogeochemical cycling. The ISME Journal **6**: 1602–1612.

Trembath-Reichert, E., D. H. Case, and V. J. Orphan. 2016. Characterization of microbial associations with methanotrophic archaea and sulfate-reducing bacteria through statistical comparison of nested Magneto-FISH enrichments. PeerJ **4**: 31913-31.

Treude, T., A. Boetius, K. Knittel, K. Wallmann, and B. B. Jørgensen. 2003. Anaerobic oxidation of methane above gas hydrates at Hydrate Ridge, NE Pacific Ocean. Marine Ecology Progress Series **264**: 1–14.

Tryon, M. D., K. M. Brown, and M. E. Torres. 2002. Fluid and chemical flux in and out of sediments hosting methane hydrate deposits on Hydrate Ridge, OR, II: Hydrological processes. Earth and Planetary Science Letters **201**: 541–557.

Tryon, M. D., K. M. Brown, M. E. Torres, A. M. Tréhu, J. McManus, and R. W. Collier. 1999. Measurements of transience and downward fluid flow near episodic methane gas vents, Hydrate Ridge, Cascadia. Geology **27**: 1075–1078.

Watanabe, Y., S. Nakai, A. Hiruta, R. Matsumoto, and K. Yoshida. 2008. U–Th dating of carbonate nodules from methane seeps off Joetsu, Eastern Margin of Japan Sea. Earth and Planetary Science Letters **272**: 89–96.

Zhang, F., H. Xu, H. Konishi, and E. E. Roden. 2010. A relationship between d104 value and composition in the calcite-disordered dolomite solid-solution series. American Mineralogist **95**: 1650-1656.

C h a p t e r T h r e e

OBSERVED AND MODELED SHIFT OF CARBONATE-ASSOCIATED MICROBIAL BIOMARKERS DURING *IN SITU* SIMULATED METHANE SEEP QUIESCENCE

David H. Case[1]

in collaboration with,

Marcos Y. Yoshinaga[2], Alexis L. Pasulka[1], Katherine S. Dawson[1], Andreas J. Greve[2], Kai-Uwe Hinrichs[2], and Victoria J. Orphan[1]

[1]Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA, USA

[2]Organic Geochemistry Group, Department of Geosciences, University of Bremen, 28334 Bremen, Germany

This chapter is in preparation for publication.

3.0 ABSTRACT

Carbonate pavements at marine methane seeps preserve microbial signatures in the geologic record long after seep quiescence. However, the taphonomy of biomarkers associated with seep microorganisms is poorly understood. We characterized microbial intact polar lipid (IPL) profiles associated with carbonates from a well-studied methane seep ecosystem, Hydrate Ridge (OR, USA), and determined aspects of these profiles were well-differentiated by seepage activity and by mineralogy. We further contextualized IPL distributions with parallel interpretation of bacterial and archaeal 16S rRNA genes and core archaeal lipids. *In situ*, time-resolved transplant experiments simulated methane seep quiescence, enabling direct observation of shifts in biomarker profiles on short (13 month) timescales, as well as forward modeling of continued biomarker profile changes beyond 13 months.

Bulk IPL profiles exhibit less change than 16S rRNA gene biomarkers upon transplantation simulating seep quiescence. However, differences between IPL and core lipid profiles also indicate IPLs do not represent an integrative record of microorganisms with the longevity of core lipids. The majority of IPLs are cosmopolitan to all seep conditions, but the distribution of some IPLs are especially sensitive to seep activity (e.g., phosphatidylcholine-diacylglycerol-C34:2 [PC-DAG-C34:2]) and respond to quiescence of methane seepage on the order of years. Estimates of cell concentrations from total lipid amounts indicate populations of ANME-1 ($\sim 10^7$ to 10^8 cells/cm^3) and ANME-2 ($\sim 10^7$ to 10^{10} cells/cm^3) in carbonates from active seeps are similar to sediments from similar environments, emphasizing the likely importance of carbonate-associated anaerobic methanotrophs in seep methane cycling. Relative proportions of ANME-1 IPL biomarkers appear to be robust for determination of whether active vs low-activity seep conditions are recorded. Core lipids appear biased toward ANME-1 over ANME-2 probably a result of differential lipid degradation rates. Some bacterial IPLs likely associated with sulfate-reducing deltabacteria exhibit a closer association with dolomite than calcite mineralogy. An

integrated understanding of microbial biomarker distribution and time-dependent behavior is critical for accurate interpretations of paleo-seep carbonates.

3.1 INTRODUCTION

The central microbial metabolism in marine methane seeps is the sulfate-coupled anaerobic oxidation of methane (AOM) which is estimated to consume >80% of contemporary subsurface methane, thereby preventing its release into the ocean/atmosphere system (Reeburgh 2007). Marine methane seeps have been proposed to play many roles throughout Earth history. These have included implications for the ancient evolution of life (Rasmussen 2000; Peckmann and Goedert 2005), extreme warming events such as the Paleocene-Eocene Thermal Maximum (Katz et al. 1999) or Neoproterozoic post-glacial warm periods (Jiang et al. 2003), and/or perturbations in Earth's carbon cycle (Hinrichs 2002; Schrag et al. 2013). Although these hypotheses continue to be debated (Barstow et al. 2011; Slotznick and Fischer 2016), it is true that methane seeps have been pervasive for hundreds of millions of years (or longer) and during that time have likely played important roles in cycling Earth's carbon inventory.

Carbonate precipitates are ubiquitous features of the seafloor landscape at methane seeps (Suess et al. 1985; Moore et al. 1990; Boetius and Suess 2004). These precipitates, ranging in size scale from 10^{-3} to 10^{2} meters, are hypothesized to form as a result of alkalinity generated during sulfate-coupled AOM (Berner 1980; Ritger et al. 1987). Carbon isotopes provide supporting evidence for this, with seep carbonates often exhibiting depleted $\delta^{13}C$ values diagnostic of methanotrophy (Kulm and Suess 1990; Gieskes et al. 2005). A variety of carbonate mineralogies have been observed in seep settings including aragonite, calcite, dolomite, and mixtures thereof (e.g., Naehr et al. 2007). It is hypothesized that dolomites form deep in the sediment column, below AOM zones, while aragonites form in sulfate-replete near-surface sediments and calcites precipitate at or near the zones of highest AOM (Burton 1993; Greinert et al. 2001; Naehr et al.

2007; Bayon et al. 2009; Blättler et al. 2015). In addition to their occurrence and distribution at or near the seabed, seep carbonates also extend many meters beneath the seafloor and effectively represent the dominant volumetric fraction of habitat substrate at methane seeps (Marlow et al. 2014a and references therein). The links between microbial activity and mineralogy are not well understood, but pore fluid geochemistry and flux, as well as microbial assemblage composition, have been identified as possible drivers of mineralogy (Aloisi et al. 2002; Teichert et al. 2005; Reitner et al. 2005; Stadnitskaia et al. 2008; Leefmann et al. 2008; Birgel et al. 2011; Hagemann et al. 2012).

Censuses of seep sediment-hosted microbial diversity based on the 16S rRNA gene have recently revealed a dominance of *Methanomicrobia*, *Deltaproteobacteria*, and candidate divisions Hyd24-12 and JS1 as compared to other marine environments (Pop Ristova et al. 2015; Ruff et al. 2015). However, 16S rRNA gene profiles in exhumed seep carbonates recovered from the seabed are distinct from those in seep sediments (Marlow et al. 2014b; Case et al. 2015). These carbonate-associated microbial assemblages are less dominated by Deltaproteobacterial and bacterial Candidate Division JS1 16S rRNA gene sequences than seep sediments and are characteristically rich in the uncultured gammaproteobacterial JTB255 Marine Benthic Group (Case et al. 2015). Carbonate-associated microbial assemblages, like those in seep sediments, appear to be strongly shaped by the magnitude of local seepage (Lloyd et al. 2010; Rossel et al. 2011; Case et al. 2015). Recent measurements of active AOM by carbonate-hosted microbial assemblages established that viable methanotrophic microorganisms inhabit seep carbonates (Marlow et al. 2014a). Furthermore, seep carbonate 16S rRNA gene diversity does not appear to record historically active microbial populations: microbial 16S rRNA gene biomarker richness decreases markedly upon imposed seep quiescence during *in situ* ecological transplant experiments and 16S rRNA gene profiles are distinct between active and low-activity seepage sites (Case et al. 2015).

The presence of microbial lipid biomarkers recovered from Phanerozoic carbonates, often highly depleted in $\delta^{13}C$, have been used to link geologic outcrops to historic methane seepage sites (Peckmann et al. 1999; Thiel et al. 1999; Peckmann et al. 2002; Goedert et al. 2003; Birgel et al. 2006b; a; Birgel et al. 2008b; a; Kiel et al. 2013; Natalicchio et al. 2015; Little et al. 2015). Some studies have further attempted to reconstruct the strength of past seepage conditions (Leefmann et al. 2008; Peckmann et al. 2009; Birgel et al. 2011; Hagemann et al. 2012) or microbial ecology from detailed analysis of recovered microbial lipids (e.g., ANME-1 vs ANME-2 ratios; Blumenberg et al. 2004; Birgel et al. 2006a; Niemann and Elvert 2008; Peckmann et al. 2009; Birgel et al. 2011; Natalicchio et al. 2015). However, the interpretation of lipid biomarkers in ancient carbonates as being representative of past biochemical and ecological environments is hampered by several factors. First, co-incorporation of lipids from carbonate-specific cells with lipids from exogenous sedimentary and planktonic cells (Peckmann et al. 2002; Hoffmann-Sell et al. 2011; Blumenberg et al. 2015) can expand and complicate the environmental signal recorded in lipid biomarkers. Similar results have been reported for marine glycerol-diphytanyl-glycerol-tetraether (GDGT) lipid profiles, where sedimentary and planktonic signals appear to be mixed (Pearson et al. 2016). Second, post-depositional thermal alteration and diagenesis can alter lipid profiles, thereby impacting interpretation of environmental conditions at the time of deposition (Goedert et al. 2003; Hagemann et al. 2012). A third process has remained relatively unexplored: genuine shifts in the seep-associated microbial assemblages after the period of seep activity but prior to incorporation into the geologic record. For example, ANME-1 archaea are often observed to predominate over ANME-2 in low fluid flux regimes (and vice versa in high flux seeps), and thus the recovery of ANME-1 vs ANME-2 lipids has been used to infer past seepage magnitude (Blumenberg et al. 2004; Stadnitskaia et al. 2008; Peckmann et al. 2009). However, if, during progressive seep quiescence, the carbonate-hosted microbial assemblage and associated biomarkers were to shift in reflection of the diminishing seepage flux as is observed in 16S rRNA

genes, the resulting recorded biomarker inventory would not fully represent the environmental history of methane seepage.

IPLs have generally been considered to degrade quickly after cell death due to the instability of the bond between the glycerol backbone and polar head group, leading to the assumption that IPLs reflect living biomass (White et al. 1979; 1997; Zink et al. 2003; Sturt et al. 2004). However, recent studies have called this into question. In one study, laboratory incubations of anoxic sandy North Sea sediments were amended with dead cells of the archaeon *Haloferax volcanii* and the eukaryote *Saccharomyces cerevisiae* (eukaryotic non-isoprenoidal ester-bound IPLs are structurally similar to those of bacteria). After 100 days, >75% of ester-bound IPLs had degraded, while no degradation was measured of the archaeal ether-bound IPLs (Logemann et al. 2011). A second study spiked a radioactive synthetic analog of a monoglycosidic archaeol (^{14}C-1G-AR) into North Sea and deep subsurface sediments, then measured IPL degradation over 300 days. The IPL degradation half-life was extrapolated to be ~3-300 ky, depending on environmental conditions (Xie et al. 2013). Finally, diglycosidic (2G) GDGTs were recently measured to have slow production rates and hypothesized to be synthesized by ANME-1 in stationary phase rather than active growth (Kellermann et al. 2016). The taphonomy of IPL biomarkers remains an active area of debate and is critical to understand for application in modern and/or ancient systems, particularly because IPLs (and core lipids) can be preserved in authigenic carbonate minerals.

Methane seepage itself is known to fluctuate on an extremely wide range of timescales, from diurnal (tidal forcing) to interglacial (changes in mean sea level), suggesting the activation and quiescence of seepage is a highly relevant process impacting microbial assemblages at methane seeps (Tryon et al. 1999; 2002; Teichert et al. 2003). A laboratory experiment with seep sediments from the North Sea, in which methane was supplied for 120 days, removed for 36 days, resupplied for 40 days, and then removed permanently, showed AOM to respond on the scale of days to varying methane flux (Wegener and Boetius 2009). However, the study only examined rates of AOM and not biomarkers of the constituent methanotrophic microbial consortia. Adding

further complexity, modeling and empirical measurements indicate individual carbonate pavements precipitate over timescales of 10^2-10^3 years, implying that a single carbonate can record periods spanning multiple cycles of seepage activation and quiescence in cases of rapidly shifting seep activity (Luff et al. 2004; Bayon et al. 2009). Understanding whether and to what extent carbonates record the extant microbial community structure vs an integrative record of all previously inhabiting microorganisms is therefore of critical importance. Furthermore, the history of microbial assemblages recorded in seep carbonates may vary depending on the type of biomarker analyzed (DNA vs IPL vs core lipid).

In this study we (i) describe the IPL profiles associated with modern seep carbonates at Hydrate Ridge, OR, USA, and link differences in IPL profiles to environmental factors, (ii) employ 13-month, ecological *in situ* transplantation experiments to test the impact of seep quiescence on lipid biomarker profiles, and (iii) forward model the response of both IPL and 16S rRNA gene biomarkers to extended conditions of quiescence, including comparison to archaeal core lipids from our sample set. Ultimately this enables hypothesis development regarding the changes in microbial community structure that occur when a seep becomes dormant – a phase in the seep life cycle common to carbonates which eventually enter the geologic record.

3.2 MATERIAL AND METHODS

3.2.1 FIELD EXPERIMENTS AND SAMPLE COLLECTION

All carbonate experiments and sample collections were conducted at Hydrate Ridge (HR), a well-studied natural laboratory of methane seepage offshore OR, USA (Suess et al. 1985; Tryon et al. 1999; Boetius et al. 2000; Sahling et al. 2002; Tryon et al. 2002; Treude et al. 2003; Boetius and Suess 2004; Gieskes et al. 2005; Levin et al. 2010; Pasulka et al. 2015; Case et al. 2015; Table 1). Twenty-two of our 23 samples were collected from a site referred to as Hydrate

Ridge North (HR-N), a promontory at ~600 meters below sea level (mbsl) just above the oxygen minimum zone (~0.7 mL L^{-1} O_2; Levin et al. 2010). One sample was recovered from a deeper promontory (775 mbsl) demonstrating active seepage located approximately 12 km farther south (HR-S) and bathed in oxygen-depleted waters (0.2-0.3 mL L^{-1} O_2; Levin et al. 2010). Specific sampling stations (e.g., "HR-3"; see Case et al. 2015 for coordinates) were identified as active (HR-3 and HR-7) or low-activity (HR-4 and HR-8) based on the presence or absence, respectively, of diagnostic seafloor characteristics such as methane ebullition and/or chemosynthetic communities such as mats and clam beds (as described in Orphan et al. 2004). Later pore water measurements confirmed that active stations exhibit higher sulfide concentrations than low-activity stations, corroborating our activity designations (Pasulka et al. 2015). High throughput Illumina sequences of the 16S rRNA gene V4 region for all 23 samples have recently been published in a study of seep microbial 16S rRNA gene diversity (Case et al. 2015).

The 23 seep carbonates were assigned to one of three categories: native, transplantation, or colonization (following designations in Case et al. 2015). Native carbonates were exhumed carbonate slabs and blocks recovered directly from the seafloor and provide essential context for interpreting the biomarker signatures in the transplantation and colonization experiments in active and low-activity sites. Transplantation carbonates were transferred from the same stations as native carbonates via the DSV *Alvin* from active to low-activity stations, and vice versa (paired active and low-activity stations were separated by 10^1-10^2 meters on the seafloor). These replicated experiments simulated rapid seep quiescence (active to low-activity) and seep activation (low-activity to active). The transplanted samples were incubated for 13 months on the seafloor before recovery. Colonization carbonates represent samples of a large dolomitic and calcitic carbonate slab that had been collected from a seep site during a previous cruise and subsequently sterilized by autoclaving in the lab prior to deployment as part of this study. The colonization carbonates were placed on the seafloor for 13 months at the same stations as the transplantation

experiments. All collection and experimentation was conducted during R/V *Atlantis* cruises AT15-68 (2010) and AT18-10 (2011). Upon recovery onboard ship, all carbonates were processed into subsamples and frozen at -80˚C for subsequent DNA and lipid biomarker analysis.

3.2.2 BIOMARKER EXTRACTION AND QUANTIFICATION

Two independent lipid datasets were generated: a quantitative IPL dataset that included both bacterial and archaeal lipids (hereafter "IPL dataset"), and a semi-quantitative dataset exclusive to Archaea which consisted of both intact polar and core lipids (hereafter "archaeal semiquantitative lipids dataset" or "ASL dataset"; relative abundance values within this dataset are robust, but not absolute core lipid concentrations). To prepare samples for total lipid extraction, subsamples of frozen carbonates were lyophilized overnight (~10-30 g/sample), followed by pulverization for 4 minutes in a tungsten-carbide shatter box. The box was cleaned thoroughly with isopropyl alcohol between samples to remove debris. Powders were then extracted as total lipid extracts (TLEs) using a modified Bligh and Dyer protocol (Sturt et al. 2004), after adding an internal standard (phosphatidylcholine $C_{21:0}/C_{21:0}$). The obtained TLEs were characterized by high-performance liquid chromatography mass spectrometry (HPLC-MS). TLEs were measured in positive ionization mode, while scanning a mass-to-charge (*m/z*) range of 150–2,000, with automated data-dependent MS/MS fragmentation of base peak ions. Compound detection was conducted on a Bruker maXis Ultra-High Resolution qToF-MS, equipped with an electrospray ionization (ESI) interface. For quantitative IPL analysis, separation of polar lipids was achieved on a Dionex Ultimate 3000 UHPLC equipped with a Waters Acquity UPLC BEH Amide column (150 x 2.1 mm, 1.8 µm particle size), operating in normal phase. Rings and unsaturation patterns of archaeal tetraether lipids were evaluated by reverse phase chromatography with a Waters Acquity BEH C_{18} column. Details of the chromatographic conditions and analyses are described in Wörmer et al. 2013). Compound identification was

achieved by monitoring exact masses of possible parent ions (present mainly as H^+ and NH_4^+ adducts) in combination with characteristic fragmentation patterns (Sturt et al. 2004; Yoshinaga et al. 2011). The reported concentrations of microbial lipids are based on the peak areas of molecular ions, accounting for potential differences in ionization during HPLC-MS routines. A list of commercially available and purified standards used to determine the response factors of polar lipids are provided in the Supplementary Material (Table S1). It is worth mentioning that carbonate samples were additionally extracted using HCl (e.g. Birgel et al. 2006b) and analyzed for comparison with the Bligh and Dyer method. Similar or even higher yields for both polar lipids and archaeal core lipids were obtained using the Bligh and Dyer method relative to the HCl treatment (data not shown).

3.2.3 MINERALOGICAL AND ISOTOPIC CHARACTERIZATION

All samples were subjected to bulk mineralogical and carbon isotopic analysis. Carbonates were first ground in sterile ceramic mortar and pestle to generate a homogenous powder. Mineralogy was then determined by X-ray diffraction (XRD) on a Phillips X'Pert Multi Purpose instrument. Measurements were taken from 10˚ to 70˚ 2θ with step size of 0.05˚. SiO_2 standards were run to confirm peak location accuracy. Similar to previous studies, areas of diagnostic peaks for aragonite (2θ = 26.3˚), calcite (2θ = 29.5˚), and dolomite (2θ = 31.0˚) were used to determine quantitative mixing ratios of bulk mineralogy for each rock (Tennant and Berger 1957; Bergmann 2013; Marlow et al. 2014b; Table 1). Ultimately, the carbonates were coarsely divided into two categories – aragonitic and calcitic/dolomitic (Supplemental Text).

Bulk organic carbon isotopic composition, as well as weight percent organic carbon, was determined for each carbonate. Organic carbon was isolated by digesting ~4 mg bulk powder in 2N H_3PO_4 three times. All samples were analyzed via continuous flow (He; 100 mL/min) on a Costech Instruments Elemental Combustion System model 4010 (EA) by oxidation at 980°C over

chromium (III) oxide and silvered cobalt (II, III) oxide followed by reduction over elemental copper at 650 °C. CO_2 was subsequently passed through a water trap and then a 5 Å molecular sieve GC at 50 °C to separate N_2 from CO_2. CO_2 was diluted with helium in a Conflo IV interface/open split prior to analysis. $\delta^{13}C$ values were measured on a Thermo Scientific Delta V Plus IR-MS. $\delta^{13}C$ values were corrected for sample size dependency and then normalized to the VPDB scale with a two-point calibration and internal standards.

3.2.4 STATISTICAL ANALYSIS

Biomarker "richness", a measure of ecological alpha diversity (i.e., within-sample diversity) was defined as a simple presence/absence count of the total lipids (or 16S rRNA gene OTUs) present in a given sample, regardless of the biomarkers' relative abundance. Bray-Curtis similarity, a measure of ecological beta diversity (i.e., between-sample diversity), was calculated as the similarity of total biomarker profiles (incorporating not only presence but also abundance of biomarkers) between samples. This was done by two methods: first, Bray-Curtis similarities were calculated based on the raw quantitative IPL data (in ng/g). Second, Bray-Curtis similarities were calculated on IPL data which had previously been transformed into relative abundance. Then, for both calculation methods, nonmetric multidimensional scaling (NMDS) ordinations and Analysis of Similarity (ANOSIM) tests were computed in R using the 'vegan' package v2.0-10 (Oksanen et al. 2013; R Core Team 2014). NMDS ordinations were computed with the 'metaMDS' function on a Bray-Curtis similarity matrix (distance="bray") set to two dimensions (k=2), with a maximum of 100 iterations (trymax=100) and the 'monoMDS' engine. ANOSIM calculations were also executed on a Bray-Curtis similarity matrix with a maximum of 999 permutations. Each lipid from both the relative abundance IPL and archaeal ASL datasets was tested for correlation against each OTU from the iTag 16S rRNA gene dataset using the 'cor' function in R (method="pearson"), with p-values calculated using the 'cor.test' function (method="pearson").

3.2.5 MODELING TIME-DEPENDENT BIOMARKER PROFILE SHIFTS

We took two modeling approaches in interpreting our time-resolved transplantation experiments. First, carbonate-associated biomarker richness (the presence/absence of biomarkers and the degree to which they are detected across multiple sample types) was used to track and forward model whole-community microbial responses to seep quiescence. Second, we examined variations in concentration of individual IPL and 16S rRNA gene biomarkers which, based on our dataset, may be diagnostic of active or low-activity methane seep environments.

Because richness is highly sensitive to sampling depth (discovery opportunity), the richness analyses were restricted to an equal number of samples per category examined. For example, Fig. 5 was limited to two samples per category because of the low sample number of our transplant-to-low-activity experiments. Therefore, the results of our richness analyses (Fig. 5, 6) are technically only representative of the specific samples examined. However, we tested multiple permutations of samples and found similar results for the majority of sample combinations, such that the results presented in Fig. 5 and Fig. 6 are generally supported across the whole dataset (Supplemental Text).

We modeled total biomarker richness as a function of time (t) in carbonates during seep quiescence by the following general equation:

$$\text{Richness}(t) = a + b + c \qquad \text{(Eq. 1)}$$

where a is the constant richness never lost or gained (the cosmopolitan richness found in all samples; Table 2), b is the remaining active-type richness (a vector of decreasing richness with time; Table 2), and c is the richness gained from low-activity stations (a vector of increasing richness with time; Table 2; Supplemental Text). Our time-resolved transplantation data gives us three time points necessary to empirically derive an equation to fit the microbial community dynamics observed at Hydrate Ridge: t=0 years, the community structure of native-active carbonates prior to seep quiescence, t=1.08 years, the community structure after 13 months of

seep quiescence, and $t \gg 1.08$ years, the community structure expected when the microbial assemblage has fully turned over to low-activity-type conditions (Table 2; Supplemental Text).

We assume a log-linear response rate of microbial biomarkers over time, congruent with previous observations of microbial processes (Shade et al. 2013). Ecologically, a log-linear model implies the microbial communities are responding most rapidly during immediate onset of seep quiescence, with slower response over time. A rapid metabolic response of AOM-related microorganisms is supported by previous laboratory incubations (Wegener and Boetius 2009). Log-linear responses to environmental change are well-supported in the other biological systems as well (Benincà et al. 2008; Korhonen et al. 2010), although the availability of high-throughput next generation sequencing has only recently allowed microbial ecologists to begin fitting models to high resolution time series data (Faust et al. 2015). More experimental time points would be necessary in order to fully test and validate this assumption for microbial community responses at Hydrate Ridge. These models were built to match the experimental observation of an initial drop in biomarker richness upon quiescence and predict a gradual transition to the characteristic biomarker richness of the native-low-activity carbonates, although the models were not forced to recreate the input data points (Fig. 6c). We applied the modeling to our parallel 16S rRNA gene (Fig. 6b), IPL (Fig. 6c), and archaeal core lipid (Fig. 6d) datasets.

In our second modeling approach we examined our data to identify specific IPLs (and 16S rRNA gene OTUs) characteristic of active and low-activity seepage conditions and which could be useful biomarkers for future geobiological studies of seepage activity and AOM. In order to identify key IPL biomarkers of interest (using the quantitative IPL dataset), we applied three criteria to the entire suite of native-active, native-low-activity, and transplant-to-low-activty carbonates (Q1=first quartile; Q3=third quartile):

(i) *This criterion ensures there is a significant difference in concentration of the IPL between active and low-activity seep conditions.*
In order to qualify as an active-type IPL, Q1 of the IPL among the native-active carbonates must be greater than Q3 of the IPL among the native-low-activity carbonates. In order to qualify as a low-activity-type IPL, Q3 of the IPL among the native-active carbonates must be less than Q1 of the IPL among the native-low-activity carbonates.

(ii) *This criterion ensures that upon seep quiescence the concentration of the IPL is changing in a manner consistent with the observed difference between active and low-activity conditions.*
The median of the IPL among the transplanted (active to low-activity) carbonates must be between the medians of the IPL among the native-active and native-low-activity carbonates.

(iii) *This criterion ensures that spurious IPL biomarkers are not identified with extremely low concentrations among the dataset.*
In order to qualify as an active-type IPL, the median of the IPL among native-active carbonates must be greater than 4.43 ng/g in concentration. In order to qualify as a low-activity-type IPL, the median of the IPL among native-low-activity carbonates must be greater than 4.43 ng/g in concentration. For this criterion, the value of 4.43 ng/g was chosen because it is 1% of the median total IPL concentration for all native and transplantation carbonates in this study.

In order to identify key 16S rRNA genes of interest, we employed the same criteria with one modification on the third criterion in order to account for the relative abundance nature of the 16S rRNA gene dataset: rather than a cutoff value of 4.43 ng/g (irrelevant to the 16S rRNA gene data), we applied a cutoff of 1% relative abundance. After identifying specific biomarkers (Fig. 7), we applied a log-linear approach in order to generate biomarker-specific models (Fig. 8; Table 3).

3.3 RESULTS

The cumulative per-sample concentrations of 201 archaeal and bacterial IPLs ranged three orders of magnitude between 25 ng/g and 12,189 ng/g (Fig. 1a; Table 1; Table S2). Among active seep stations both aragonitic and calcitic/dolomitic-type carbonates were recovered (median=954 ng/g; max=12,189 ng/g; min=57 ng/g), but among the low-activity seep stations sampled here only aragonitic carbonates were recovered (median=436 ng/g, max=522 ng/g, min=377 ng/g). The three native-active-station carbonates which were aragonitic in composition exhibited the highest total IPL abundance (>4,180 ng/g; Fig. 1a). This high total IPL abundance was mirrored in the single rock transplanted from an active to low-activity station (11,757 ng/g), which happened to also be aragonitic. The remaining 19 of 23 samples all yielded IPL concentrations of less than 1,462 ng/g (mean=351 ng/g, σ=356 ng/g). Besides the three native carbonates with very high IPL abundances, the remaining 10 native carbonates were not differentiated in total IPL concentration by activity or mineralogy ($mean_{active,calcitic/dolomitic}$=445 ng/g; $mean_{low\text{-}activity,aragonitic}$ =432 ng/g), although native-calcitic/dolomitic carbonates from active seep stations exhibited 10-fold wider variability (σ=588 ng/g) than native-aragonitic carbonates from low-activity seep stations (σ=59 ng/g). Transplanted carbonates yielded similar total IPL abundances to the native carbonates (Fig. 1a), while all four autoclaved colonization carbonates yielded very low amounts of IPLs (mean=76 ng/g, σ=37 ng/g). Pre-deployment negative controls of the colonization carbonates after autoclaving yielded no bacterial IPLs but detectable concentrations of uncharacterized archaeal IPLs (data not shown).

The four carbonates which hosted the highest total IPL concentrations (3 native and 1 transplant) also exhibited the four most depleted bulk organic carbon isotopic compositions (Table 1; Fig. 1c). The total dataset of 23 carbonates demonstrated a linear correlation between bulk $\delta^{13}C_{org}$ and total IPL concentration (R^2=0.60; $p \ll 0.01$) but no linear correlation between bulk organic $\delta^{13}C_{org}$ and total IPL concentration when the four carbonates with very high IPL

concentrations were excluded (R^2=0.05; p=0.36). Aragonitic carbonates, regardless of seep activity or experimental treatment, exhibited significantly more depleted bulk organic $\delta^{13}C$ (mean=-49.0±11.3‰) than calcitic/dolomitic carbonates (mean=-26.3±3.3‰).

Native carbonates with high total IPL concentrations exhibited high proportions of archaeal IPLs and low $\delta^{13}C_{org}$ values (Fig. 1d; Fig. S3; Table S2). A list of archaeal IPLs and their likely sources in cold seep systems is provided in the Supplementary Material (Table S3). These included phosphatidyl-inositol, phosphatidyl-glycerol, and phosphatidyl-serine hydroxyarchaeols (PI-, PG-, and PS-OH-AR), common among ANME-2 archaea, as well as monoglycosidic, diglycosidic, and phosphatidyl-glycerol glycerol-diphytanyl-glycerol-tetraethers (1G-, 2G-, and PG-GDGT) typically assigned to ANME-1 archaea (Fig. 1d; Table S3). Bacterial fractions of total IPLs were composed of a variety of non-isoprenoidal diacyl-glycerols (DAG), acyl-ether-glycerols (AEG), and diether-glycerols (DEG) with phospholipid head groups (Fig. 1d). The chemotaxonomy of these bacterial IPLs is not fully understood, although previous studies of seep core lipids have associated non-isoprenoidal glycerol ethers with sulfate-reducing *Deltaproteobacteria* (e.g. Hinrichs et al. 2000; Orphan et al. 2001; Elvert et al. 2003).

Archaeal IPLs were a more abundant fraction of the total IPLs on native-aragonitic carbonates as compared with native-calcitic/dolomitic carbonates (Fig. 2a), which was also reflected in the higher archaeal IPL proportion associated with carbonates from low-activity vs active stations (Fig. 2b). The median proportion of ANME-2-derived IPLs, identified as the sum of AR (including OH-AR) IPLs, among all IPLs, was also higher on aragonitic carobnates but with less of a clear differentiation by seep activity (Fig. 2). Among archaeal IPLs only, the median proportion of ANME-2-derived IPLs was only slightly higher in the native-aragonitic carbonates as compared to native-calcitic/dolomitic carbonates (Fig. 2a). Relative proportions of IPLs affiliated with ANME-2 among total archaeal IPLs were slightly higher in carbonates from active stations than low-activity stations (Fig. 2b). The median proportion of IPLs attributed to ANME-1 (all GDGTs divided by all archaeal IPLs) among all bacterial and archaeal IPLs was higher

among low-activity seep carbonates than active seep carbonates, but equivocal by mineralogy (Fig. 2). ANME-1 fractions among solely archaeal IPLs demonstrated similar, though less strong, distribution with regard to seep activity. Native-aragonitic rocks exhibited a higher and more variable fraction of PG-GDGTs among intact polar tetraethers than calcitic/dolomitic rocks, consistent with more contemporary/active ANME-1 biomass (Yoshinaga et al. 2015; Kellermann et al. 2016; Fig. 2a). Absolute abundances of the 2G-GDGT IPL on its own were higher in native-low-activity than native-active carbonates (Fig. 1; Fig. 7).

Over the whole ASL dataset, archaeal core lipids were more abundant than archaeal IPLs ($Avg_{core/IPL}$=5; $Max_{core/IPL}$=16, $Min_{core/IPL}$=1; Fig. 1e; Table S2). The core lipids were dominantly composed of GDGTs, while the IPLs were rich in AR and OH-AR (Fig. 1e; Table S2). ANME-2/$Archaea_{Total}$ ratios were consistently lower by a factor of 3.2±1.6 in the core lipids as compared to the IPLs recovered in the ASL dataset (R^2=0.80; Table S6). The ANME-2/$Archaea_{Total}$ ratio was not consistent between IPLs in the ASL and IPL datasets, which may be derived from methodological differences in the data acquisition or represent differences in fossil vs extant archaea.

Whether analyzing the IPL or ASL datasets, aragonites were observed to host the highest richness of lipid biomarkers (Fig. S1). However, in all three of the IPL, ASL, and 16S rRNA gene datasets the plurality of lipids were cosmopolitan to all mineralogy types (Fig. S1; raw data used to generate each Venn diagram in Fig. S1 is given in Table S5). Decomposing the IPL and ASL datasets from the native-active carbonates by mineralogy again revealed the archaeal IPLs to be more associated with aragonitic than calcitic/dolomitic carbonates (Fig. 3), with bacterial IPLs (IPL dataset) and core archaeal lipids (ASL dataset) more associated with calcitic/dolomitic than aragonitic carbonates. Some abundant non-isoprenoidal ester- and ether- bound bacterial IPLs (e.g., PC-DAG-C33:1, PC-DAG-C31:0, PG-DEG-C36:2), tentatively assigned to sulfate reducing bacteria, appeared to be more strongly associated with dolomitic than calcitic mineralogy (Fig. 3a).

Non-metric multidimensional scaling (NMDS) ordinations of IPL profiles from the native and transplantation carbonates revealed some differences as compared to corresponding 16S rRNA gene diversity surveys (Fig. 4). Unlike 16S rRNA gene trends, the native samples are not statistically differentiated by seep activity among the IPL dataset. This difference is driven by the native-active samples clustering into two distinct groups (separated by mineralogy) while the native-low-activity carbonates cluster into one distinct group (all of which exhibit aragonitic mineralogy). Mineralogy was also observed to differentiate the 16S rRNA gene profiles, although in that dataset seep activity was more strongly associated with inter-sample DNA-based differences. IPL profiles from replicate transplanted carbonates at duplicate seafloor stations did not generally demonstrate a shift in microbial community over the 13 month experiment. IPL diversity profiles from neither native nor transplant carbonates were differentiated by $\delta^{13}C_{org}$ or seafloor sampling station (Fig. S2).

When examining IPL richness, we find a large proportion of lipids to be cosmopolitan to all samples – present in the transplants and both activity classes of native carbonates (n=47 in Fig. 5a; n=91 in Fig. 5b). However, clear loss of active-type biomarkers and gain of low-activity-type biomarkers is apparent from transplant experiments upon 13 months of seep quiescence. Twenty-eight percent (station HR-3/-4 in Fig. 5a) and 30% (station HR-7/-8 in Fig. 5b) of the IPLs present in native-active carbonates were lost after 13 months of seep dormancy in transplanted carbonates. Meanwhile, the appearance of IPLs characteristic of background carbonates (e.g., from low-activity sites) is slower – after 13 months, only 11% (station HR-3/-4 in Fig. 5a) and 4% (station HR-7/-8 in Fig. 5b) of the IPLs present in native-low-activity carbonates have appeared in the transplant carbonates.

Modeling reveals highly similar time-dependent richness evolution for both the 16S rRNA gene and IPL biomarkers, at both experimental stations (Fig. 6). However, 16S rRNA gene biomarkers incur a more dramatic decrease in richness, down to ~40% of maximum richness, whereas IPL biomarkers do not decrease beyond ~70% of maximum richness (Fig. 6b-c). This

may be due to partly to genuinely higher magnitude of loss and gain within the 16S rRNA gene biomarker pool and partly to the level of taxonomic depth afforded by gene sequencing vs IPL analysis. Notably, the archaeal core lipids do not exhibit this behavior, instead only accumulating biomarker richness over time, consistent with higher recalcitrance to degradation. After ~100 years of modeled time, very little remaining biomarker richness change is predicted.

Identification of specific IPL and 16S rRNA gene biomarkers characteristic of active or low activity conditions revealed 6 diagnostic IPLs and 11 diagnostic 16S rRNA gene OTUs in the dataset (out of 201 IPLs and 1,057 16S rRNA gene OTUs total) that may represent robust biomarkers characteristic of seep activity (Fig. 7; Fig. 8; Table 3). Of the IPL biomarkers, which included 5 bacterial (all diacylglycerol [DAG] IPLs) and 1 archaeal IPL (2G-GDGT), the 5 bacterial IPLs were diagnostic of active seepage conditions and the 1 archaeal IPL was diagnostic of low-activity conditions. Of the 16S rRNA gene biomarkers, which also included bacterial and archaeal OTUs (e.g., Desulfobulbaceae, and ANME-1b), 5 were characteristic of active seepage conditions and 6 were characteristic of low-activity conditions. All ten of the combined IPL and 16S rRNA gene biomarkers diagnostic of active seep conditions responded relatively rapidly to seep quiescence, with modeling revealing them to be >90% transitioned (i.e., decreased in abundance to low-activity seep levels) within 10 years (Fig. 8a,e). Biomarkers diagnostic of low-activity seepage responded more slowly, with only 2 of the 7 biomarkers (the ANME-1b and *Brocadiaceae* 16S rRNA gene OTUs) being >90% transitioned (i.e., increased in abundance to low-activity levels) within 10 years of modeled time (Fig. 8f).

3.4 DISCUSSION

3.4.1 COMPARISON OF 16S rRNA GENE, IPL, AND CORE LIPID BIOMARKERS

The taphonomic process by which microbial biomarkers are recoded in seep carbonates is not well understood, and the extent of preservation may vary for 16S rRNA genes vs intact

polar lipids vs core lipids. Accurate interpretation of ancient seep carbonates for the purpose of reconstructing paleo-microbiota and carbon cycles is dependent on whether and to what extent seep carbonates record microbial bio-signatures reflective of the assemblages during active seepage. This study presents several lines of evidence suggesting that IPL biomarker profiles in native seep carbonates integrate a period of time only slightly longer than 16S rRNA genes, and that neither represent an integration time as long as core lipids.

Three primary observations from the 16S rRNA gene data described in Case et al. (2015) indicated that gene biomarkers principally recorded the extant microbial community and that carbonate-associated microbial communities underwent succession upon seep quiescence in our 13-month *in situ* transplantation experiments: (i) a difference in 16S rRNA gene community structure (beta diversity) was observed in microbial assemblages inhabiting native active vs low-activity carbonates, (ii) 16S rRNA gene richness (alpha diversity) decreased upon seep quiescence, and (iii) specific activity-sensitive OTUs (e.g., *Helicobacteraceae*) clearly responded to imposed changes in seep activity (Fig. 1,2,5 in Case et al. 2015). In order to examine whether IPLs, like DNA-based diversity surveys, also reflect the extant microbial community's shifts on short (13-month) timescales, we analyzed IPLs from the same carbonate sample set as described in Case et al. (2015) to look for similar patterns of biomarker behavior. In addition, the quantitative nature of the IPL dataset allowed us to probe shifts in absolute abundance and estimate total cell concentrations – interpretations not possible in PCR-based 16S rRNA gene surveys which rely on relative abundance.

Unlike in the 16S rRNA gene data, ordination of the IPL data did not reveal seepage activity to statistically differentiate the native-active and native-low-activity assemblages ($p=0.07$ for relative abundance IPL data; $p=0.13$ for absolute abundance IPL data; Fig. 4b-c). This either suggests that IPLs record a more homogenized (time-integrated) biomarker inventory than 16S rRNA genes or that IPL biomarkers are recorded in carbonates exclusively during periods of active seepage and subsequently change very little upon seep quiescence. However, we note that

the ordination does not reveal a highly homogenized mixture of native samples, but rather three distinct clusters: one of native-low-activity carbonates (all aragonitic) and two of native-active carbonates, which are separated by mineralogy (Fig. 4b-c). Interestingly, the two mineralogy-separated clusters of native-active carbonates are differentiated by, among others, the absolute abundances of PC-DAG-C34:2 (calcite/dolomite) and PS-OH-AR (aragonite; a putative ANME-2 biomarker). Overall, while mineralogy is strongly associated with differences in microbial assemblages (as was also the case for 16S rRNA genes), the effect of seep activity also seems to play a role. The clustering of native-low-activity carbonates separate from both native-active clusters implies storage of distinct IPL biomarker inventories as a function of seep activity, which would be the case if IPL inventories eventually shift between active and low activity seepage periods. Moreover, the aragonitic carbonates are well-differentiated into active and low-activity groups. This is supported by analysis of specific IPL ratios, which indicate the proportion of archaea to total microorganisms as well as the proportion of ANME-1 among total IPLs is different in active vs low-activity environments (ANME-1, specifically, higher in low-activity settings; Fig. 2b).

Complex cycling of biomarkers (simultaneous loss of active-type and gain of low-activity-type biomarkers), rather than passive biomarker accumulation over time, would be supported by a decrease in richness upon imposed seep quiescence (our transplant-to-low-activity experiments) as the microbial assemblage responds to changing conditions. Indeed this was observed in the 16S rRNA gene data, and it is also observed in the IPL data (Fig. 5; Raw data used to generate Fig. 5 is given in Table S5). While both biomarkers indicate a decrease in richness upon seep quiescence, we note that the majority of IPL richness is cosmopolitan to all samples (Fig. 5), while the majority of 16S rRNA gene richness was specific to either active or low-activity seepage conditions (Fig. 3b-c in Case et al. 2015). This fundamental difference complements the NMDS ordinations, implying that IPL biomarkers overall have distributions similar to 16S rRNA genes but with more tendency to exhibit cosmopolitan, shared, and/or time-integrated aspects.

If microbial assemblages respond to seep quiescence, and the response is reflected in biomarker patterns on short timescales, then specific biomarkers ought to be identifiable which exhibit a rise (or fall) in abundance during the transition. This was clear in 16S rRNA gene data, where OTUs associated with the *Helicobacteraceae* and *Methylococcales* (putative sulfur and methane oxidizers, respectively), as well as ANME-1b (anaerobic methanotrophs) clearly shifted in relative abundance during the transplantation experiments (Fig. 4 in Case et al. 2015). Not only are IPL biomarkers also identifiable which demonstrate short timescale response to changes in seep activity, but some IPL biomarkers have similar putative chemotaxonomy as the 16S rRNA OTUs (Fig. 7). For example, IPLs and OTUs associated with ANME-1 both increased in relative abundance. This similar biomarker recovery suggests IPLs are recording contemporary changes in the microbial community, at least for some microorganisms. Furthermore, the quantitative nature of the IPL dataset enables examination of similar (or different) lipid biomarker shifts in absolute vs relative abundance. Among the ANME-1 IPLs, for example, an increase upon transition to low-activity conditions was observed in both absolute and relative abundance, but the magnitude of shift was lesser in absolute than relative units. Among bacterial IPLs, we observed that those which decrease in absolute abundance upon seep quiescence (e.g., PC-DAG-C34:2) also decrease in relative abundance (Fig. 7; Fig. 8).

It therefore appears that IPL biomarkers are distributed in a broadly similar manner as 16S rRNA genes and exhibit roughly similar shifts in pattern upon imposed changes in seep activity. As such, IPLs may reflect extant microbial communities as previously suggested (White et al. 1979; 1997; Zink et al. 2003; Sturt et al. 2004). However, other lines of evidence imply a degree of recalcitrance within the IPL biomarker pool that would require IPLs to reflect a somewhat longer time-integration than the 16S rRNA genes.

Firstly, IPLs appear recalcitrant to degradation as evidenced by the recovery of high IPL concentration in the aragonitic carbonate transplanted from an active to low-activity seep station (Fig. 1a). Secondly, in NMDS ordination the transplant-to-active carbonates plot near the native

carbonates from the low-activity regimes in which they originated (Fig. 4b). This was not the case in 16S rRNA gene data, where transplants simulating seep activation revealed a rapid response of the microbial assemblages to renewed methane flux (Fig. 1b in Case et al. 2015). Thirdly, as already pointed out above, the majority of IPL richness is cosmopolitan to all carbonates (Fig. 5), implying longer-term storage of biomarker inventories than 16S rRNA gene OTUs where the majority of richness was specific to either active or low-activity conditions (Fig. 3b-c in Case et al. 2015). Finally, the proportion of GDGTs with phosphatidyl-glycerol head groups (PG-GDGT), which are suggested to be a proxy for active ANME-1 cells (Kellermann et al. 2016), are a minority of all GDGT IPLs (Fig. 2). This points to a large proportion of GDGT IPLs being either historic or perhaps produced in stationary phase rather than by actively growing cells (Kellermann et al. 2016).

It appears that IPLs represent a biomarker inventory with somewhat greater longevity than 16S rRNA genes, but how do IPLs compare to core lipids, which are hypothesized to be the best biomarker recording ancient seep microorganisms and processes (e.g., Birgel et al. 2008b)? The independent archaeal lipid dataset (ASL) is well-suited to this inquiry – archaeal intact polar and core lipids can be directly contrasted. The ASL dataset reveals seep carbonates to host a higher amount of archaeal core lipids than IPLs by a factor of up to 16 (Fig. 1e), supporting a fossil record of historic microorganisms (the core lipids) stored in seep carbonates. However, the specific profiles of archaeal lipids differ whether viewed as IPLs or core lipids, specifically the ANME-2/Archaea$_{Total}$ ratio which tends to be lower in core lipids than IPLs (Table S6). This ratio is of interest in geobiological studies due to the interpretation that ANME-1 and ANME-2 are physiologically adapted to low and high flux methane supplies, respectively (Niemann and Elvert 2008; Birgel et al. 2011; Natalicchio et al. 2015). The marked difference in ANME-2/Archaea$_{Total}$ ratios between archaeal intact polar and core lipid data is evidence that the core lipids likely preserve a biomarker inventory with even longer integration time than IPLs. The core

lipid record is likely impacted by the specific longevity of GDGTs as hypothesized by Kellermann et al. (2016).

3.4.2 MODELED TIME-DEPENDENT BIOMARKER PROFILE CHANGES

3.4.2.1 SHIFTS IN COMMUNITY RICHNESS UPON SEEP QUIESCENCE

Our parallel biomarker datasets indicate increasing longevity of biomarkers from 16S rRNA genes to IPLs to core lipids. By extrapolating our transplant experiment results beyond 13 months, we can further inform the timescales of biomarker profile changes upon seep quiescence – both the richness of the entire microbial assemblage as well as the concentrations of specific "active-type" or "low-activity-type" biomarkers.

From the richness models, it is possible to define three phases of microbial community change upon seep quiescence: (i) a phase of rapid richness loss (approximately the first and second years), (ii) a phase of relatively rapid richness gain (approximately through the first decade), and (iii) a relatively slow transition to the new stable state of the microbial assemblage (approximately year 10 onward). 16S rRNA gene biomarkers demonstrate a larger relative decrease in richness – Phase (i) – than IPL biomarkers (Fig. 6b,c). Both biomarkers are reduced in richness to approximately the level of the core "cosmopolitan" richness (dashed horizontal lines in Fig. 6) before rebounding to higher richness levels. This core richness could be interpreted as either evidence for a pool of stored fossil biomarkers, or as evidence of a core microbial community whose constituent members have no sensitivity to seep flux, and therefore whose presence does not change upon seep quiescence.

A fundamental difference between the IPL and 16S rRNA gene datasets is that the majority of richness in IPLs is cosmopolitan to all samples, but for 16S rRNA genes the majority of richness is specific to either active or low-activity conditions. Therefore, the process of biomarker profile change is inherently different: IPLs are undergoing a subtle shift in the

presence/absence of a minority of the overall IPL profile, while 16S rRNA gene biomarkers are turning over the majority of the presence/absence community profile between two significantly different microbial assemblage types (active and low-activity). The longevity of IPLs in subseafloor settings is not well constrained, with some estimates that IPLs represent living biomass and some that IPLs, especially archaeal IPLs, may degrade extremely slowly (Logemann et al. 2011; Xie et al. 2013). Our data over 13-month experiments are consistent with IPLs exhibiting longevity more similar to DNA than core lipids, and therefore that on these approximately year-long timescales IPLs tend to reflect extant microbial assemblages. This incongruity with recent evidence of long-lived IPLs is perhaps rooted in different methodological approaches: our experiments were performed *in situ*, with analysis of biomarkers from genuine seep microorganisms, while recent studies either employed synthetic analogs of a single archaeal IPL type (Xie et al. 2013) or amended laboratory incubations with cells not representative of methane seep microbial diversity (Logemann et al. 2011). Although even DNA has been questioned as reflective of extant microorganisms (Dell'Anno et al. 1998; Levy-Booth et al. 2007; Torti et al. 2015), evidence of microbial assemblage shifts in our 16S rRNA dataset strongly suggests a reflection of contemporary microbial community dynamics.

3.4.2.2 CHANGES OF BIOMARKERS CHARACTERISTIC TO ACTIVE OR LOW-ACTIVITY SEEP REGIMES

It appears from 16S rRNA gene data (Case et al. 2015) and modeling (this study) that microbial assemblages do change upon seep quiescence, and that several decades are necessary for community structure to fully respond to changing conditions. Concurrent with these changes, carbonates at low-activity seep sites nonetheless still host viable anaerobic methanotrophs (Marlow et al. 2016a). IPL profiles appear to preserve a slightly more historic whole-assemblage record than 16S rRNA genes, with a smaller proportion of the IPL profile than the 16S rRNA

gene profile shifting upon seep quiescence (Fig. 6). However, concentrations of a subset of individual IPL (and 16S rRNA gene) biomarkers do appear to be sensitive to seep activity (Fig. 7). These included bacterial and archaeal IPLs (e.g., PC-DAG-C34:2 and 2G-GDGT) as well as bacterial and archaeal 16S rRNA gene OTUs (e.g., Desulfobacteraceae and ANME-1b). Cross-correlation of the lipid and 16S rRNA gene datasets supports that some of these characteristic biomarkers may in fact represent the same microbial clades. The PE-DAG-C30:1 IPL, though not strongly correlated with any of the diagnostic 16S rRNA gene OTUs, is strongly associated with another *Helicobacteraceae* OTU in the 16S rRNA gene dataset (Table S4), a sulfur-oxidizing epsilonproteobacterial family known to be sensitive to seep activity (Case et al. 2015). All detected lipid-DNA correlations are provided in full in Table S4, although we note that apparent correlations may arise not because an IPL is specifically produced by a microbial species, but because distributions of some microbial species are correlated with each another (e.g., Trembath-Reichert et al. 2016) and therefore their constituent IPLs may be correlated as well. In addition, it is difficult to identify specific chemotaxonomic connections between IPLs and specific microbial taxa without pure cultures of diverse environmental microorganisms with which to probe species-specific IPL production.

Modeling for the 11 16S rRNA gene biomarkers predicted that 7 of 11, including all 5 active-type biomarkers, would be >90% transitioned to values observed in the low-activity sites within 10 years (Fig. 8). Thus, the response of key activity-sensitive IPL and 16S rRNA gene biomarkers to seep quiescence is, as expected, shorter than the whole-community response (Fig. 6, Fig. 8). It is possible this approach could, in future studies, be applied to differentiate endogenous vs exogenous inputs to microbial biomarker inventories in other marine environments, for example where pelagic and sedimentary signals are mixed (Blumenberg et al. 2015; Pearson et al. 2016).

Of the five characteristic active-type bacterial IPLs (Fig. 7), PC-DAG-C43:2 was recovered in especially high relative abundance on the colonization carbonates (mean=8%,

min=2%, max=17%, Table S2), notable because bacterial IPLs were not recoverable from pre-incubation sterilized negative controls of the colonization experiments. However, patterns of colonization by microorganisms producing PC-DAG-C34:2 did not seem to be activity-dependent, as both the maximum and minimum relative abundance values were recovered from the two carbonates placed for colonization at low-activity seep stations (the same was true for absolute abundance values; Table S2).

Notably, in both IPL and 16S rRNA gene analyses, putative biomarkers for ANME-1 archaea (2G-GDGT among the IPLs, ANME-1 OTUs among the 16S rRNA genes) are predicted to increase in abundance during the shift to a low-activity seepage environment (Fig. 7). This is in concordance with previous studies in which ANME-1 biomarkers were assumed to represent a low-flux seepage regime (Blumenberg et al. 2004; Peckmann et al. 2009) and in which ANME-1 OTUs were observed to both natively inhabit and actively colonize low-activity seep carbonates at a higher relative abundance than at active seeps (Marlow et al. 2014b; Case et al. 2015). We can determine from our quantitative IPL data that the increase in 2G-GDGT relative abundance is at least partly a function of decreasing total IPL content rather than in-growth of new organisms in low-activity conditions, similar to 16S rRNA gene data from Marlow et al. (2014a). Median absolute abundance of the 2G-GDGT IPL is 14 times higher in the native-low-activity (median=33.0 ng/g; Q1=27.7 ng/g; Q3=55.5 ng/g) than in the native-active (median=2.4 ng/g; Q1=0.5 ng/g; Q3=21.9 ng/g) carbonates. However, median relative abundances of 2G-GDGT are 20 times higher in native-low-activity (median=7.6%; Q1=7.3%, Q3=10.6%) than native-active (median=0.4%; Q1=0.4%; Q3=0.5%) carbonates. Therefore of the increase in relative abundance of the 2G-GDGT IPL upon seep quiescence, approximately 70% of the signal is supported by a genuine increase in absolute abundance and approximately 30% of the signal is due to the degradation of other IPLs in the dataset. This is also reflected in estimates of cell concentrations (see below). ANME-1b OTUs were observed to actively colonize sterilized seep carbonates on short timescales in 16S rRNA gene data (Case et al. 2015), but

similar interpretation is difficult with the IPL dataset due to the recovery of archaeal IPLs in the pre-colonization negative controls. The observations in the IPL dataset are supported by the independent archaea-specific ASL dataset, in which 2G-GDGT IPLs are also observed to increase in relative abundance in the native-low-activity carbonates (Fig. 1), but absolute abundances are not available for the ASL dataset. Furthermore, an independent stable isotope labeling study of sediment-hosted seep microbial communities concluded that 2G-GDGT IPLs were likely produced by ANME-1 cells in stationary phase rather than actively growing (Kellermann et al. 2016). Over time the core lipids appear biased toward ANME-1 GDGT lipids over ANME-2 AR and OH-AR lipids, probably as a result of slow degradation rates of GDGTs. Thus caution must be used when interpreting ancient seep biomarker profiles as fully representative of historic seep conditions and processes, an application which has already begun to be explored (Birgel et al. 2006a; Niemann and Elvert 2008; Peckmann et al. 2009; Birgel et al. 2011; Natalicchio et al. 2015).

3.4.3 ESTIMATES OF MICROBIAL CELL CONCENTRATION ASSOCIATED WITH SEEP CARBONATES

Previous evidence, though limited in scope, suggests that carbonates from active seepage regimes exhibit higher porosity, permeability, rates of methane oxidation, total microbial cell concentration, and ANME-2 relative abundance than carbonates from low activity settings (Marlow et al. 2014a; b). In order to convert from ng/g to cell/cm^3, we applied multiple theoretical and empirical conversion factors for the density of seep carbonates and the cellular mass of IPLs (Supplemental Text; Simon and Azam 1989; Zink et al. 2008; Lipp et al. 2008; Meador et al. 2014; Marlow et al. 2014a) (Fig. 9). These estimates corroborate the finding by Marlow et al. (2014a) that active carbonates can host microbial abundances on the order of 10^9-10^{10} cell/cm^3. Our data also suggest that some seep carbonates may host cell concentrations

several orders of magnitude lower than estimated by DAPI cell counts in Marlow et al. (2014a) (Fig. 1b; Supp. Text S3). Differences could be explained by natural heterogeneity, under-efficiency of lipid extraction (which can not be ruled out), or overestimation of cell concentration in Marlow et al. (2014a), for example by extrapolating the number of cells from raw counts of microbial aggregates. With our quantified IPL dataset we were able to estimate median ANME-1 absolute cell concentrations, calculated as the sum of GDGTs (Table S3), of $1.1x10^7$ cell/cm^3 ($Q1=1.0x10^7$; $Q3=1.1x10^8$) in native-active carbonates and $9.1x10^7$ cell/cm^3 ($Q1=8.3x10^7$; $Q3=1.7x10^8$) in native-low-activity carbonates (Fig. 9a). Likewise ANME-2 IPLs (calculated as the sum of all AR and OH-AR) gave median absolute cell concentrations of $3.9x10^7$ cell/cm^3 ($Q1=1.2x10^7$; $Q3=4.2x10^9$) in native-active carbonates and $1.9x10^8$ cell/cm^3 ($Q1=1.8x10^8$; $Q3=1.9x10^8$) in native-low-activity carbonates (Fig. 9a). Although is it surprising that the ANME-2 IPLs reveal a higher cell concentration of ANME-2 in low-activity than active seep stations, we note that the variability of ANME-2 IPLs is very wide among native-active carbonates. In fact, the description of the data is improved by binning not only according to seep activity but additionally by mineralogy. This reveals three groupings: active-calcitic/dolomitic, active-aragonitic, and low-activity-aragonitic (unfortunately, low-activity-calcitic/dolomitic carbonates were not present in the sample set). When plotting the data within these groupings, it is clear that the carbonates from active stations are separated into two groups: aragonitic, with very high cell concentrations, and calcitic/dolomitic with lower cell concentrations (Fig. 9b). This was apparent, of course, in the total cell concentrations (Fig. 1a) and in the NMDS ordinations of the data (Fig. 4b-c). Within solely aragonitic rocks (thus removing the affect of mineralogy), ANME-2 cell counts are shown to decrease from active to low-activity stations, as expected (Fig. 9b). Within only aragonitic rocks the ANME-1 cell counts are observed to be highly similar between active and low-activity conditions, and when including the calcitic/dolomitic carbonates it further appears that over the whole data set, the ANME-1 increase in abundance from active to low-activity conditions.

In order to make a comparison to active-type Hydrate Ridge seep sediments, we averaged aggregate counts (aggregates/cm^3) from the top 3 centimeters below seafloor of *Beggiatoa* fields at Hydrate Ridge reported in Boetius et al. 2000 and Treude et al. 2003, and converted to total cells/cm^3 using conversion factors reported for active-type sediments in Marlow et al. (2014a). We then calculated the proportion of total cells in near-surface Hydrate Ridge active-type sediments attributable to ANME-1 and ANME-2 using proportions reported in Knittel et al. 2005, assuming reported aggregates were 50:50 ANME-2:bacteria and single cells were 100% ANME-1. Propagating uncertainty, we find ANME-1 and ANME-2 to be present at $5.9x10^7$-to-$1.5x10^8$ and $2.6x10^9$-to-$6.6x10^9$ cell/cm^3, respectively (Fig. 9). These numbers are remarkably comparable to our calculations of ANME-1 and ANME-2 cell concentrations estimated from IPLs recovered in Hydrate Ridge active seep carbonates, suggesting carbonates host similar magnitudes of biomass as seep sediments.

3.4.4 ASSOCIATIONS BETWEEN BIOMARKERS AND MINERALOGY

Beyond the induction of authigenic carbonate formation by alkalinity generation during sulfate-coupled AOM (Berner 1980), the mechanistic links between mineralogy, microbial assemblage, and precipitation processes remain unclear (Teichert et al. 2005; Reitner et al. 2005; Leefmann et al. 2008; Birgel et al. 2011; Hagemann et al. 2012). Carbonates are hypothesized to form within the sediment column, with aragonites likely to form nearest the sediment/water interface where high sulfate concentrations inhibit precipitation of Mg-bearing minerals such as dolomite and high-Mg calcite (Burton 1993; Greinert et al. 2001; Aloisi et al. 2002). That our dataset shows some IPLs to be preferentially associated with dolomite is intriguing (Fig. 3a), as sulfate-reducing bacteria (SRB) are known to be involved in the precipitation of dolomite (Vasconcelos et al. 1995; Van Lith et al. 2003; Warthmann et al. 2005). Sulfate-reducing deltaproteobacteria *Desulfosarcina variabilis* and *Desulforhabdus amnigenus* produce non-isoprenoidal

ether- and ester-bound lipids (AEG, DEG, DAG) with a variety of polar head groups (PC, PE, PG; Rütters et al. 2001). Core and intact polar versions of these lipids have also been recovered in environmental samples heavily enriched in the *Desulfosarcina*, *Desulfococcus*, and *Desulfobulbus* genera which are known synthrophic partners of anaerobic methanotrophs (Hinrichs et al. 2000; Orphan et al. 2001; Pancost et al. 2001; Elvert et al. 2003; Rossel et al. 2008; Niemann and Elvert 2008; Schubotz et al. 2011). Association of similar lipids in this study with dolomitic carbonates renews questions regarding the timing and location of dolomite precipitation in seep sediments. Alkalinity generation by AOM consortia could help promote dolomite precipitation, but dolomites are often observed to have relatively enriched $\delta^{13}C$ values relative to methanotrophy (e.g. Kulm and Suess 1990; Greinert et al. 2001). A separate SRB community could be involved in the precipitation of dolomite, but presumably would need to be near the sediment/water interface where sulfate concentrations remain high (Schubotz et al. 2011). Previous models, in contrast, have suggested dolomites form deep in the sediment column, below the sulfate-methane transition zone (e.g. Greinert et al. 2001). Thus, the interrelationship between SRB, AOM, sulfate, and dolomite remains enigmatic.

3.5 CONCLUSIONS: OPPORTUNITIES AND CHALLENGES FOR RECONSTRUCTION OF PAST METHANE SEEPS

Interpretation of biological and chemical signatures recorded in carbonates from ancient methane seeps has great potential to inform our understanding of historic microbial community structure and carbon cycling on Earth. However, the taphonomy of microbial biomarkers has not been experimentally tested. Lipid biomarker evidence presented here from Hydrate Ridge, OR, suggests that IPL biomarkers demonstrate behavior more similar to 16S rRNA genes than complementary core lipids. Relative abundances of some individual IPLs (e.g., PC-DAG-C34:2, 2G-GDGT) are especially sensitive to changing seep flux and therefore informative as to whether

an "active-type" or "low-activity-type" microbial assemblage is recorded in seep carbonates. In the future, transplant experiments with more time points and better resolved geochemical data can further shed light on shifts in microbial community structure during response to environmental change.

IPL distributions were correlated partly with carbonate mineralogy. Aragonitic carbonates hosted IPLs most representative of active seep conditions, including depleted $\delta^{13}C_{org}$ values and high ANME-2 abundances (i.e., ARs and OH-ARs). In contrast some bacterial IPLs, including those with ester- and ether-linkages, were preferentially recovered on dolomitic carbonates (e.g., PC-DAG-C33:1, PC-DAG-C31:0, PG-DEG-C36:2). Although chemotaxonomy of these lipids is challenging, the preponderance of evidence suggests they are likely associated with sulfate-reducing bacteria.

Our quantitative IPL dataset confirms the carbonate-hosted cell concentrations determined previously by DAPI staining (Marlow et al. 2014b), but suggests greater variability among carbonates than previously observed. The concentrations of ANME-1 and ANME-2 are similar to estimates from active Hydrate Ridge seep sediments, further corroborating that seep carbonates likely play critical roles in carbon cycling both today and, potentially, in Earth's past. Accurately interpreting biomarker records of paleo-seepage, however, remains challenging due to variable rates of methane flux, microbial response, and biomarker degradation.

3.6 ACKNOWLEDGMENTS

We thank Lisa Levin for helping orchestrate the experimental design and field work. In addition, we thank the crews of *R/V* Atlantis cruises 15-68 and 18-10 for their help at sea, without which this work would not have been possible. Members of the VO research group, including Stephanie Connon, Joshua Steele, Fuyan Li, and Jeffrey Marlow helped with at-sea processing and laboratory work. DC was supported by a National Science Foundation (NSF) Graduate

Research Fellowship. Research funding for VO was provided by the NSF (OCE-0825791), the NASA Astrobiology Institute (NNA13AA92A), and the Gordon and Betty Moore Foundation Marine Microbiology Initiative #3780. This is NAI-Life Underground Publication Number 077. Funding for AP was provided through an NSF-Division of Ocean Sciences Postdoctoral Research Fellowship.

Table 1. All samples in this study are listed with their accompanying metadata. Samples are organized by the three different experimental treatments in the study: native samples (exhumed *in situ* carbonates), transplantation samples (seafloor manipulation experiments), and colonization samples (deployed sterile carbonate substrates).

Experimental Treatment	Activity	Sample Number	Specific Hydrate Ridge Geography	Seafloor Station	Latitude (Decimal °N)	Longitude (Decimal °W)	Depth (mbsl)	Carbonate Mineralogy[a]	Percent Aragonite (±15%)	Percent Calcite (±15%)	Percent Dolomite (±15%)	Total IPL Recovery (ng/g)	Cell Concentration (cells/cm^3)[b]	Bulk $\delta^{13}C_{org}$ (‰)
Native	Active	3530	North	HR-3	44.669544	125.098057	587	Aragonitic	70	23	7	11031	1.31E+09	-65.5
		5330	South	HR-V1	44.570239	125.147023	775	Aragonitic	80	18	2	4180	5.44E+08	-54.9
		5122	North	HR-7	44.667079	125.100033	601	Aragonitic	68	28	4	12189	1.59E+09	-61.5
		5109	North	HR-3	44.669463	125.098120	587	Calcitic/Dolomitic	40	38	22	144	1.88E+07	-25.1
		5110	North	HR-3	44.669481	125.098095	587	Calcitic/Dolomitic	14	57	29	117	1.52E+07	-24.0
		5112	North	HR-3	44.669481	125.098107	587	Calcitic/Dolomitic	44	38	18	57	7.42E+06	-23.9
		5123	North	HR-7	44.667079	125.100033	601	Calcitic/Dolomitic	24	26	50	446	5.81E+07	-28.1
		5120	North	HR-7	44.667079	125.100020	601	Calcitic/Dolomitic	44	0	56	1462	1.90E+08	-26.2
	Low Act.	5038	North	HR-4	44.670075	125.098674	595	Aragonitic	86	4	10	436	5.68E+07	-53.4
		3511	North	HR-5	44.669382	125.103619	620	Aragonitic	81	18	0	443	5.77E+07	-30.6
		5040	North	HR-4	44.670075	125.098674	595	Aragonitic	76	23	0	383	4.98E+07	-49.0
		5189	North	HR-8	44.667645	125.100712	604	Aragonitic	70	21	9	522	6.80E+07	-35.7
		5190	North	HR-8	44.667645	125.100712	604	Aragonitic	62	36	2	377	4.91E+07	-42.9
Transplantation	Low Act. -> Act.	5111	North	HR-3	44.669463	125.098120	587	Aragonitic	85	4	12	220	2.86E+07	-43.6
		5121	North	HR-7	44.667079	125.100020	601	Aragonitic	66	29	4	959	1.25E+08	-40.7
	Act. -> Low Act.	5193	North	HR-8	44.667627	125.100712	603	Aragonitic	71	22	6	11757	1.53E+09	-60.9
		5039	North	HR-4	44.670093	125.098699	595	Calcitic/Dolomitic	48	32	20	115	1.50E+07	-22.7
		5093	North	HR-4	44.670084	125.098686	595	Calcitic/Dolomitic	18	57	25	504	6.56E+07	-28.1
		5194	North	HR-8	44.667627	125.100712	603	Calcitic/Dolomitic	0	19	81	171	2.23E+07	-24.1
Colonization	Active	5025c	North	HR-3	44.669454	125.098145	588	Calcitic/Dolomitic	34	51	15	112	1.59E+07	-34.1
		5025d	North	HR-3	44.669454	125.098145	588	Calcitic/Dolomitic	0	4	96	90	1.17E+07	-24.3
	Low Act.	5145c	North	HR-4	44.670075	125.098661	600	Calcitic/Dolomitic	41	49	10	76	9.84E+06	-30.3
		5145d	North	HR-4	44.670075	125.098661	600	Calcitic/Dolomitic	13	0	86	25	3.24E+06	-24.1

[a]See Supplemental Text for details regarding mineralogical assignments; carbonates less >50% aragontie were binned "aragonitic" and those with <50% aragonite were binned "calcitic/dolomitic".
[b]See Supplemental Text for details regarding conversion of IPL concentration to volumetric cell abundance.

Table 2. Vales of terms *a*, *b*, and *c* in the model of biomarker richness as a function of time (Equation 1). Values for IPL biomarkers are from Fig. 5 in this study. Values for 16S rRNA gene biomarkers are from Fig. 4 in Case et al., 2015. Term *a* is the cosmopolitan richness never lost nor gained (a constant). Term *b* is the remaining amount of active-type richness, a vector decreasing with time. Term *c* is the gained low-activity-type richness, an increasing vector with time. Time point t=0 is the richness of the native-active carbonates. Time point t=1.08 is the richness of the transplant-to-low-activity carbonates. The “infinite” time point is the eventual value to be achieved in the native-low-activity carbonates. See Supplemental Text for additional details.

16S rRNA Biomarkers

	HR-3/-4			HR-7/-8		
Time (Years)	Term *a*	Term *b*	Term *c*	Term *a*	Term *b*	Term *c*
t=0	75	269	0	95	224	0
t=1.08	75	36	28	95	57	37
t=inf.	75	0	387	95	0	343
	$a(t) = 47$	$b(t) = \exp(-1.8565t+5.5947)$	$c(t) = 387 - \exp(-0.0693t+5.9584)$	$a(t) = 95$	$b(t) = \exp(-1.2633t+5.4116)$	$c(t) = 343 - \exp(-0.1054t+5.8377)$

IPL Dataset (IPLs of Archaea and Bacteria)

	HR-3/-4			HR-7/-8		
Time (Years)	Term *a*	Term *b*	Term *c*	Term *a*	Term *b*	Term *c*
t=0	47	38	0	91	58	0
t=1.08	47	14	10	91	13	5
t=inf.	47	0	48	91	0	45
	$a(t) = 47$	$b(t) = \exp(-0.9217t+3.6376)$	$c(t) = 48 - \exp(-0.2156t+3.8712)$	$a(t) = 91$	$b(t) = \exp(-1.3805t+4.0604)$	$c(t) = 45 - \exp(-0.1087t+3.8067)$

ASL Dataset (Intact polar and core lipids of Archaea only)

	HR-3/-4			HR-7/-8		
Time (Years)	Term *a*	Term *b*	Term *c*	Term *a*	Term *b*	Term *c*
t=0	16	0	0	25	0	0
t=1.08	16	0	10	25	0	5
t=inf.	16	0	48	25	0	45
	$a(t) = 16$	$b(t) = 0$	$c(t) = 15 - \exp(-0.0636t+2.7080)$	$a(t) = 25$	$b(t) = 0$	$c(t) = 8 - \exp(-0.4338t+2.0794)$

Table 3. Values of terms *a*, *b*, and *c* in the model of diagnostic biomarker shift as a function of time. IPL values are absolute abundance (ng/g). 16S rRNA gene values are relative abundance. For the 16S rRNA gene biomarkers, multiple OTUs of some taxa (e.g., *Helicobacteraceae*) were identified. Biomarkers characteristic of low-activity conditions are given in blue; biomarkers characteristic of active conditions are given in red. Bold and italicized text indicate archaeal and bacterial biomarkers, respectively. Values at t=0 are medians from the native-active carbonates. Values at t=1.08 are median values from the transplant-to-low-activity carbonates. Values at t=infinity are the median values from the native-low-activity carbonates. Low-activity-type biomarkers increase in abundance with time; active-type biomarkers decrease in abundance with time. Three criteria were used to identify characteristic biomarkers for this analysis (see main text):

(i) In order to qualify as an active-type IPL, Q1 of the IPL among the native-active carbonates must be greater than Q3 of the IPL among the native-low-activity carbonates. In order to qualify as a low-activity-type IPL, Q3 of the IPL among the native-active carbonates must be less than Q1 of the IPL among the native-low-activity carbonates.

(ii) The median of the IPL among the transplanted (active to low-activity) carbonates must be between the medians of the IPL among the native-active and native-low-activity carbonates.

(iii) In order to qualify as an active-type IPL, the median of the IPL among native-active carbonates must be greater than 4.43 ng/g in concentration. In order to qualify as a low-activity-type IPL, the median of the IPL among native-low-activity carbonates must be greater than 4.43 ng/g in concentration. For this criterion, the value of 4.43 ng/g was chosen because it is 1% of the median total IPL concentration for all native and transplantation carbonates in this study.

In order to identify key 16S rRNA genes of interest, we employed the same criteria with one modification on the third criterion in order to account for the relative abundance nature of the 16S rRNA gene dataset: rather than a cutoff value of 4.43 ng/g (irrelevant to the 16S rRNA gene data), we applied a cutoff of 1% relative abundance. After identifying specific biomarkers (Fig. 7), we applied a log-linear approach in order to generate biomarker-specific models (Fig. 8; Table 3).

	Lipid Biomarkers					
	Characteristic of Low-Activity Seep Environmen	Characteristic of Active Seep Environments				
Time (Years)	2G-GDGT	PE-DAG-C30:1	PE-DAG-C34:2	PE-DAG-C32:1	PE-DAG-C32:2	PC-DAG-C34:2
t=0	2.3761	5.2461	14.1018	13.6130	20.4486	48.3031
t=1.08	2.9358	3.1324	8.3764	10.4892	10.8297	16.1597
t=inf.	32.9653	0.4526	1.6326	0.9942	0.0000	11.1897

	16S rRNA Biomarkers										
	Characteristic of Low-Activity Seep Environments					Characteristic of Active Seep Environments					
Time (Years)	ANME1b	MG1	MBGD	*JTB255*	*SAR324*	*Brocadiaceae*	*Helicobacteraceae*	*Helicobacteraceae*	*Methylococcales*	*Thiohalophilus*	*Desulfobacteraceae*
t=0	0.0193	0.0001	0.0007	0.0002	0.0002	0.0021	0.0796	0.0150	0.0173	0.0144	0.0140
t=1.08	0.0727	0.0006	0.0018	0.0004	0.0009	0.0048	0.0023	0.0004	0.0117	0.0107	0.0099
t=inf.	0.1581	0.0222	0.0140	0.0118	0.0103	0.0104	0.0017	0.0001	0.0002	0.0002	0.0031

3.8 FIGURES

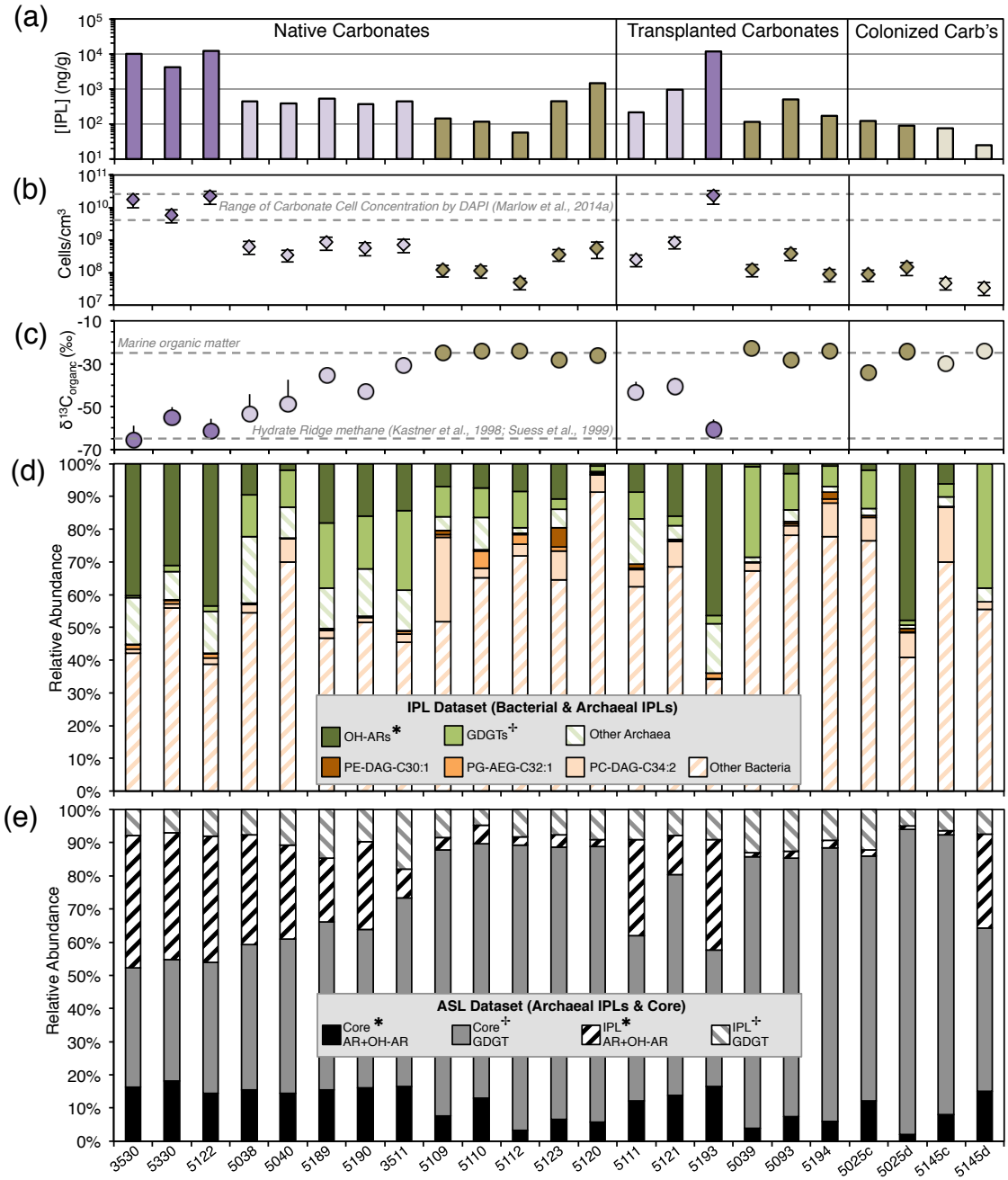


Fig. 1: Lipid and geochemical data from authigenic carbonates at Hydrate Ridge. (a) Summed quantitative bacterial and archaeal IPL concentration is reported in ng IPL per g dry carbonate powder. Purple and brown colors represent aragonitic and calcitic/dolomitic mineralogy groups, respectively. The dark and light color shading indicate active and low-activity seep stations, respectively. For transplant carbonates, these shadings represent the seep station from which the rocks were transplanted (their origin). These colors and shadings also apply in panels (b) and (c). In panel (b), IPL concentrations are converted to cells/cm^3 using multiple conversion metrics (see Supplemental Text for calculation details and assumptions). As a comparison, horizontal dashed gray lines indicate the cell concentration range determined for methane seep carbonates by direct DAPI cell counts in Marlow et al., 2014a. (c) $\delta^{13}C$ for carbonate-hosted microorganisms ($\delta^{13}C_{org}$) for each sample, with instrumental uncertainty represented by vertical black lines. (d) Relative abundance IPL profile for each carbonate sample, with IPLs putatively associated with ANME-1 (summed GDGTs) and ANME-2 (summed OH-ARs) identified by superscript symbols (+ and *, respectively; also applied in (e)). Green and orange bars represent archaeal and bacterial IPLs, respectively. (e) Relative abundance profiles of the Archaeal Semiquantiative Lipids (ASL) dataset, with both archaeal IPL (striped) and core lipids (filled) identified.

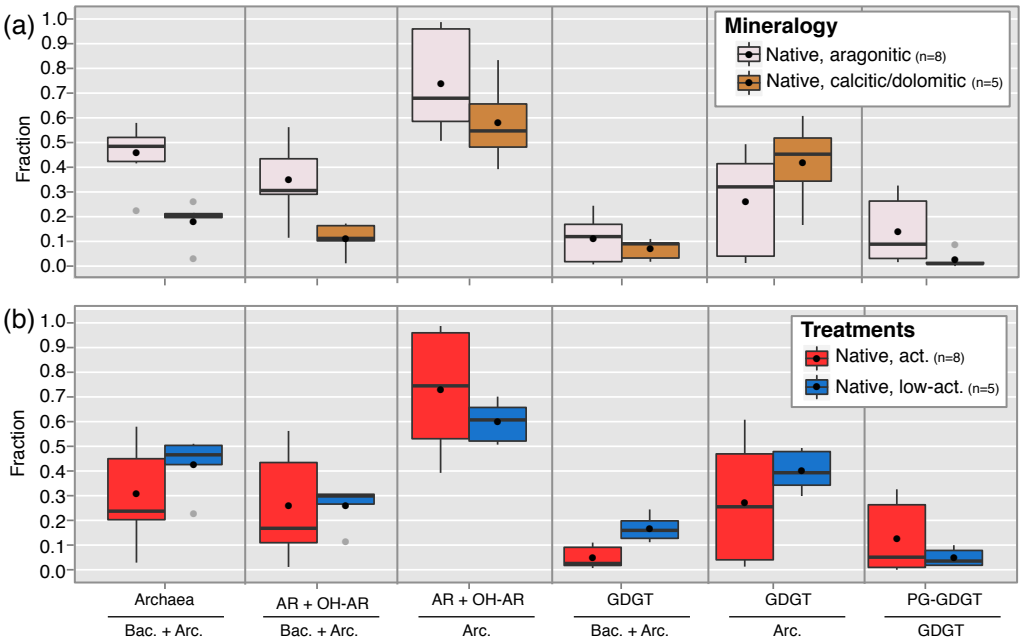


Fig. 2: Specific IPL ratios of native carbonates. (a) Samples are binned according to bulk mineralogy, (b) Samples are binned according to seep activity. Median values are given by the horizontal black line, with the colored boxes encompassing the first through third quartiles of the data. Upper and lower whiskers denote the highest and lowest data points within 1.5 times the interquartile (Q_{75} minus Q_{25}) range. Any data points outside this range are identified by gray circles. Mean values are identified by black circles, but the non-normal distribution of IPL abundances among seep carbonates implies that medians are a better statistical measure than means for this dataset. All ratios are calculated from quantitative bacterial and archaeal IPL data. The sum of AR and OH-AR (all archaeols and all hydroxyarchaeols) was used to infer the contribution of ANME-2 among in carbonate samples (Table S3). The ratio of all GDGT IPLs was calculated to infer the contribution of ANME-1 among carbonate samples. The ratio of $PG\text{-}GDGT_{All}$ (PG-GDGT, G-GDGT-PG, and 2G-GDGT-PG) over total GDGTs was calculated as a proxy for active ANME-1 biomass as has previously been suggested (Kellermann et al., 2016; Yoshinaga et al., 2015).

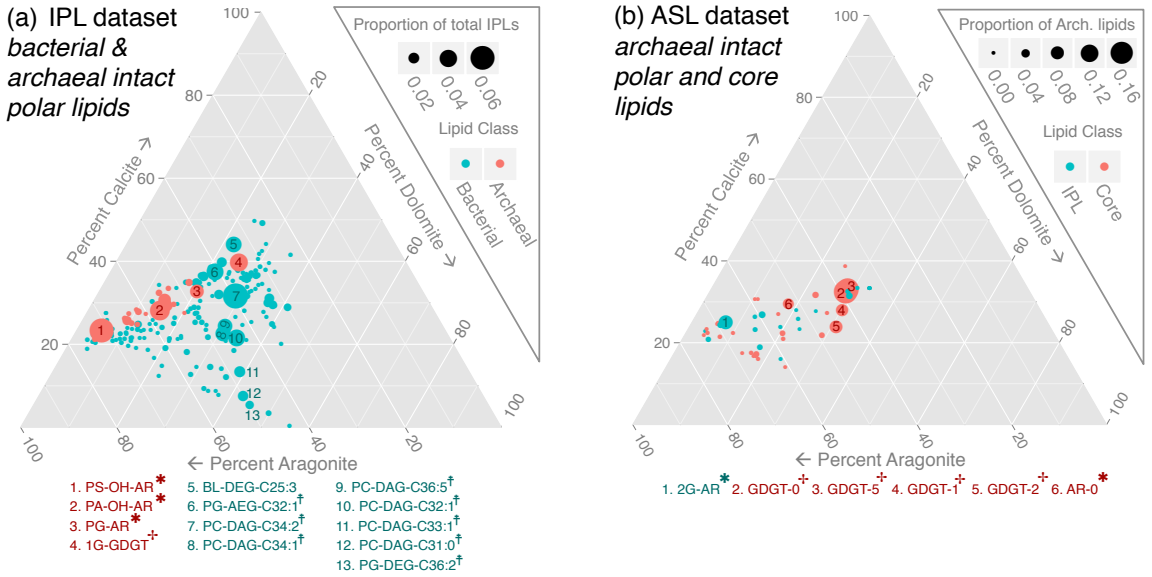


Fig. 3: Ternary diagrams of lipid biomarker data associated with carbonate mineralogy. (a) Bacterial and archaeal IPL data and (b) archaeal IPL and core lipid data according to mineralogical distributions. Only native-active samples were included in order to isolate the affect of mineralogy and avoid any influences of experimental treatment or seep activity. To calculate the percent association of each lipid with each mineralogy, for each carbonate the relative abundance of the lipid was divided into weighted proportions according to the mineralogy of the sample. These proportions were then summed for each mineralogy across all samples. Selected lipids are identified in (a) and (b) as dominant lipids in the respective datasets and identified below the plots. Bacterial non-isoprenoidal ether- and ester-bound lipids with a mass range of between 30 and 36 carbons (identified in (a) with double cross) would be consistent with SRB, according to numerous and extensive methane seep environmental characterizations (e.g., Rossel et al., 2008). As in Fig. 1d-e, archaeal lipids likely associated with ANME-1 and ANME-2 are denoted with a single cross and asterisk, respectively.

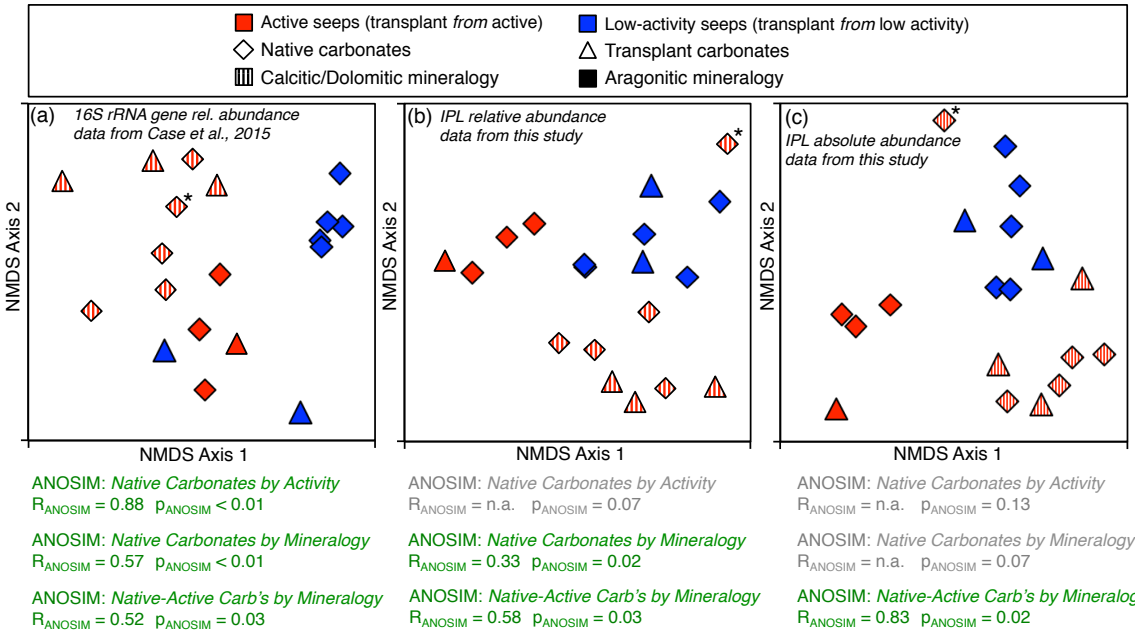


Fig. 4: Non-metric multidimensional scaling ordinations. (a) seep carbonate-associated 16S rRNA gene biomarkers from Hydrate Ridge, for the same set of samples from which lipids were extracted in this study (stress = 0.16; data published in Case et al., 2015); (b) Relative abundance IPL profiles from this study (stress = 0.15); (c) Absolute abundance IPL profiles from this study (stress = 0.07). The native-active carbonate in the upper portion of (b) and (c) is sample #5120 (marked with an asterisk in (a) also), which has previously been determined to host a microbial community anomalously high in ANME-1 compared to other native-active carbonates (Marlow et al., 2014a; Case et al., 2015). ANOSIM test results are given in green where statistically significant and in gray where statistically insignificant.

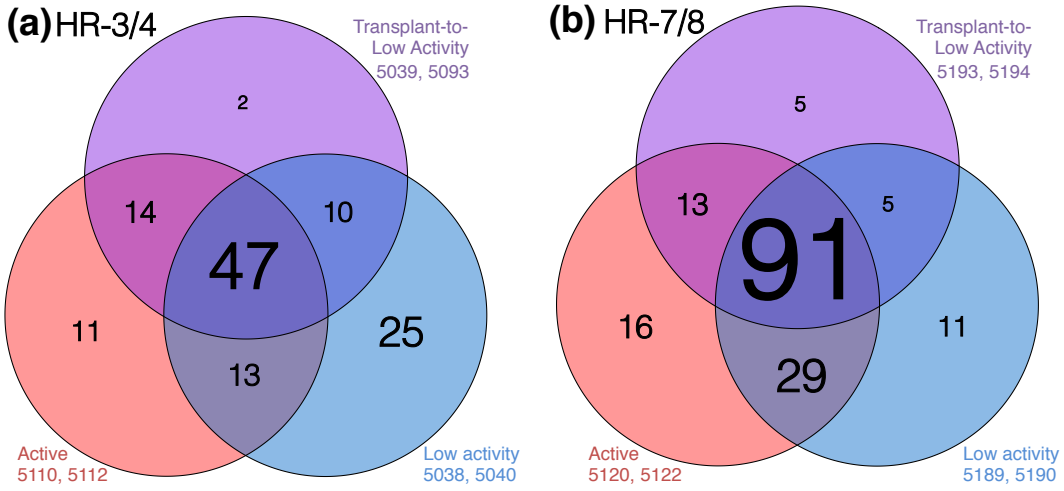


Fig. 5: Venn diagrams of lipid richness for parallel transplant experiments. (a) Transplant experiments at the paired sites HR-3 (active) and HR-4 (low-activity), (b) transplant experiments at the paired sites HR-7 (active) and HR-8 (low-activity). For each field of the Venn diagram, two duplicate samples were included (sample numbers are given in the figure). In order for a lipid to count as "present" for this presence/absence analysis, it must have been recovered from both replicates. This conservative requirement took advantage of having carbonates representing duplicate *in situ* treatment conditions on the seafloor. Although transplant-to-active experiments were conducted, they were not included in this analysis due to low replicate number. In each field of the Venn diagrams, text size is proportional to value. Twenty-eight percent (n=(11+13)/(11+13+14+47) in panel (a)) and 30% (n=(16+29)/(16+29+13+91) in panel (b)) of the IPLs present in native-active carbonates were lost after 13 months of seep dormancy in transplanted carbonates. After 13 months, only 11% (n=10/(10+25+13+47) in panel (a)) and 4% (n=5/(5+11+29+91) in panel (b)) of the IPLs present in native-low-activity carbonates have appeared in the transplant carbonates.

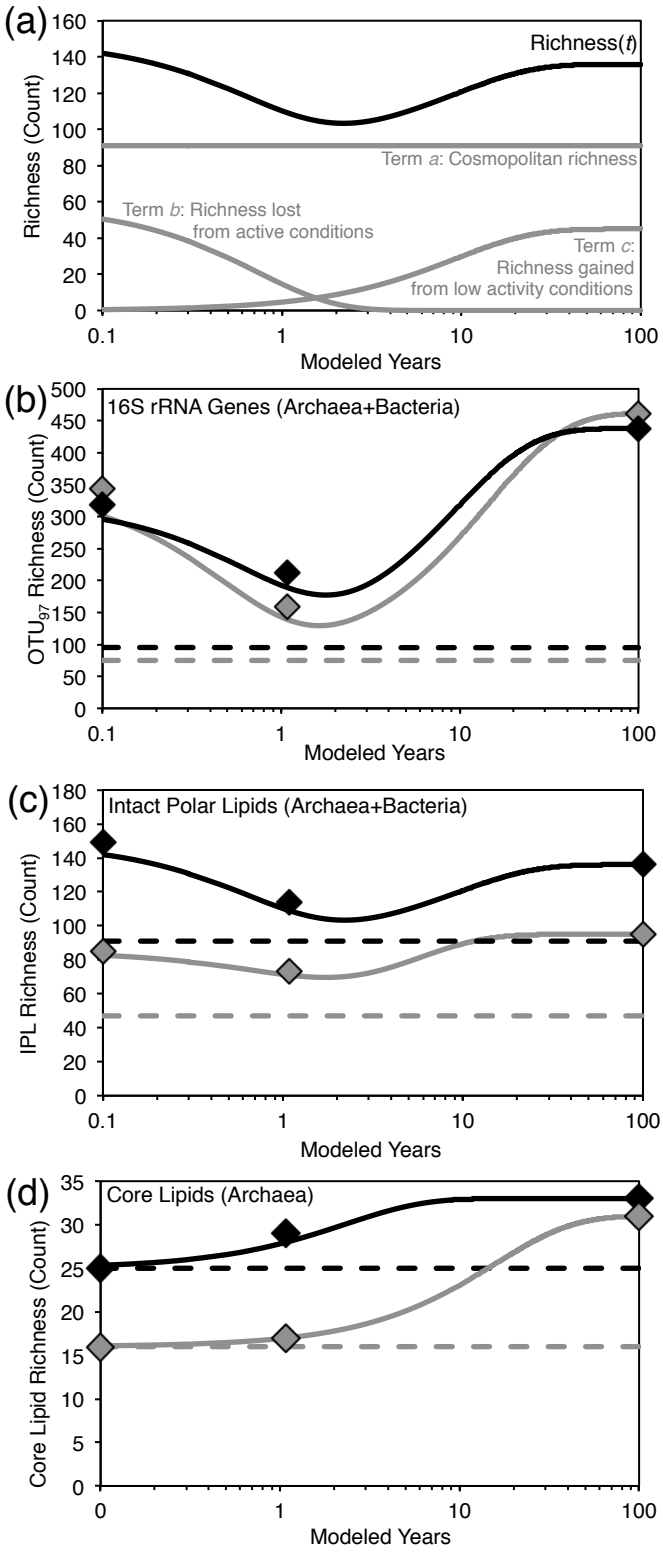


Fig. 6: Model of richness over time for 16S rRNA gene, IPL, and core lipid biomarkers. (a) Demonstration of Terms *a*, *b*, and *c* in Equation 1, (b) 16S rRNA gene richness, (c) IPL (bacteria+archaea) richness, and (d) core archaeal lipid richness. In (b-d), the model is given by the curved solid line and the data points used to generate the model are given by diamonds. The dashed lines represent the core cosmopolitan biomarker richness. Results from parallel *in situ* experiments at HR-3/4 (gray) and HR-7/8 (black) are given. The model is not *a priori* required to intercept each data point. A higher number of data points could increase predictive accuracy of the model and/or reveal highly complex temporal dynamics not currently observed. A maximum model length of 100 years was chosen because by that time, little remaining change in biomarker richness occurred. See Supplemental Text for modeling results from other sample permutations.

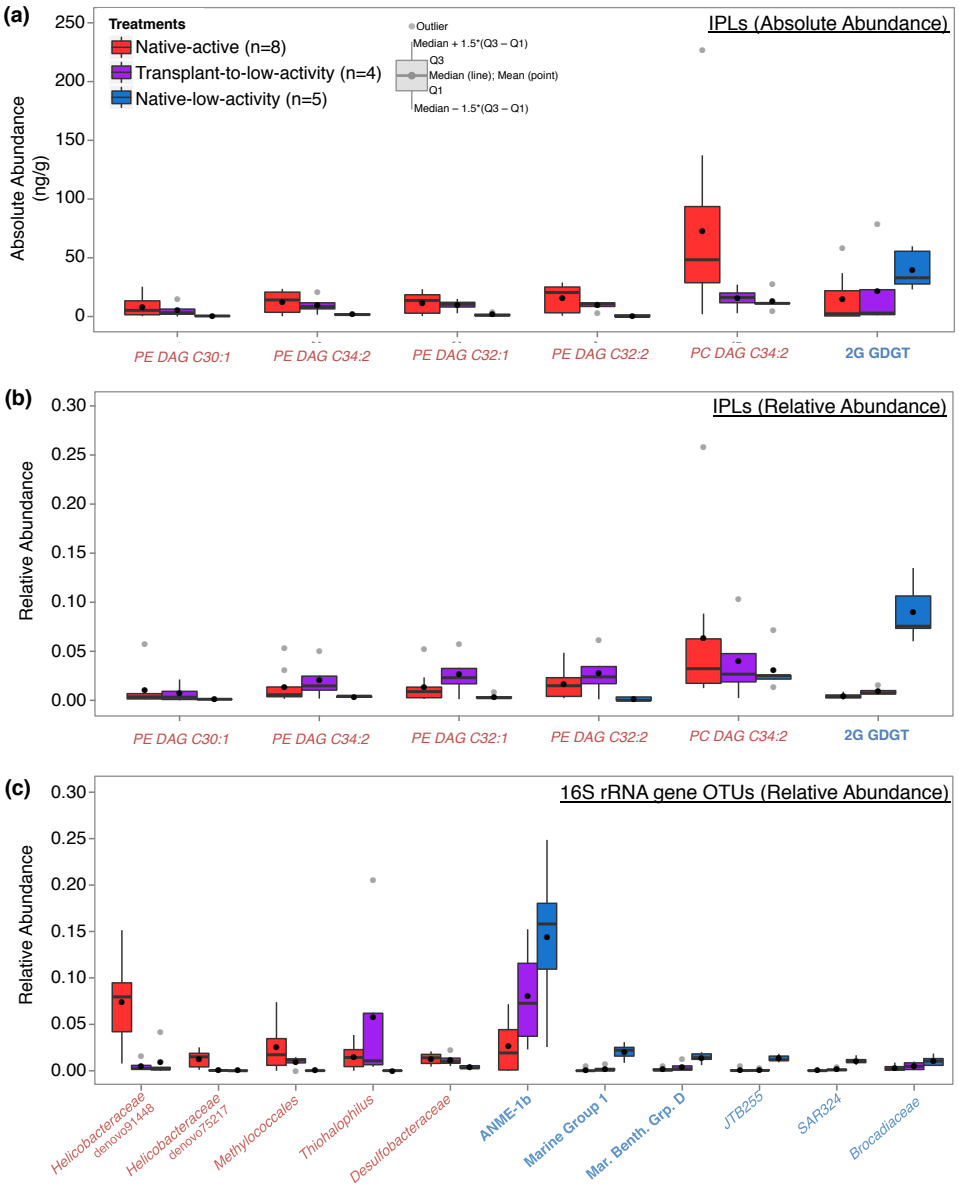


Fig. 7: Box plots of biomarkers identified as characteristic of active or low-activity seep conditions. Absolute abundances of IPL biomarkers are given in (a), with relative abundances of the same biomarkers given in (b). Panel (c) gives relatives abundances of 16S rRNA gene biomarkers. Three criteria were used to identify characteristic biomarkers for this analysis (see main text):

(i) In order to qualify as an active-type IPL, Q1 of the IPL among the native-active carbonates must be greater than Q3 of the IPL among the native-low-activity carbonates. In order to qualify as a low-activity-type IPL, Q3 of the IPL among the native-active carbonates must be less than Q1 of the IPL among the native-low-activity carbonates.

(ii) The median of the IPL among the transplanted (active to low-activity) carbonates must be between the medians of the IPL among the native-active and native-low-activity carbonates.

(iii) In order to qualify as an active-type IPL, the median of the IPL among native-active carbonates must be greater than 4.43 ng/g in concentration. In order to qualify as a low-activity-type IPL, the median of the IPL among native-low-activity carbonates must be greater than 4.43 ng/g in concentration. For this criterion, the value of 4.43 ng/g was chosen because it is 1% of the median total IPL concentration for all native and transplantation carbonates in this study.

In order to identify key 16S rRNA genes of interest, we employed the same criteria with one modification on the third criterion in order to account for the relative abundance nature of the 16S rRNA gene dataset: rather than a cutoff value of 4.43 ng/g (irrelevant to the 16S rRNA gene data), we applied a cutoff of 1% relative abundance. After identifying specific biomarkers (Fig. 7), we applied a log-linear approach in order to generate biomarker-specific models (Fig. 8; Table 3).

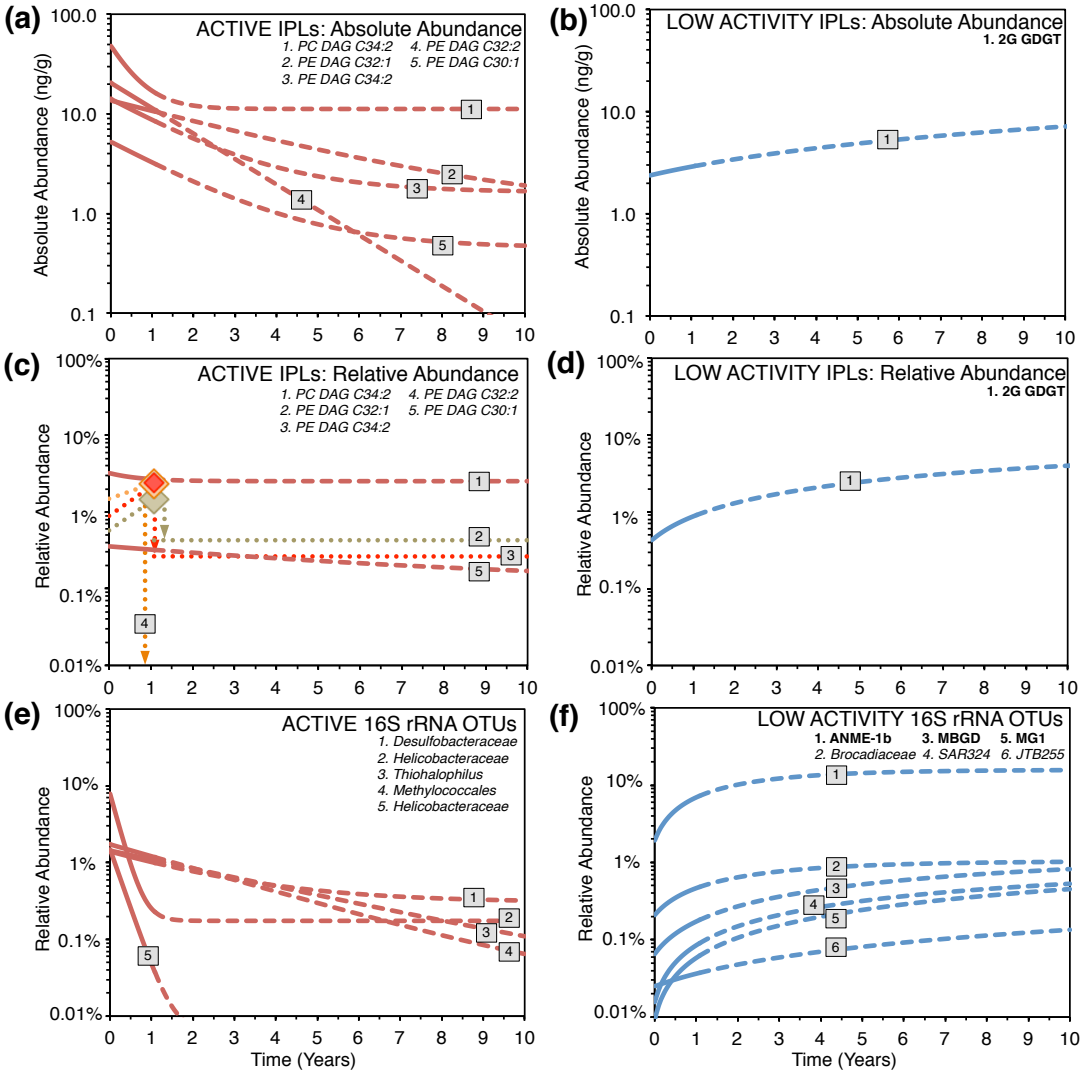


Fig. 8: Model of specific biomarker turnover for IPL and 16S rRNA gene biomarkers. Panels (a) and (b) give absolute abundances of IPLs identified as characteristic of active or low-activity seepage, respectively. Panels (c) and (d) give the same IPLs in relative abundance. Panels (e) and (f) show 16S rRNA gene OTUs identified as diagnostic of low-activity conditions. See main text and caption to Fig. 7 for description of criteria used to identify active-type and low-activity-type biomarkers. Biomarkers identified in italics are bacterial, biomarkers in bold are archaeal. Red lines indicate active-type biomarkers; blue lines indicate low-activity-type biomarkers. In all panels, solid model lines represent the timescale of the transplantation experiments (13 months), and the dashed lines represent extrapolation beyond 13 months. In panel (c), dotted lines indicate behavior of IPLs that demonstrated first a rise and then a drop in relative abundance. This behavior was not compatible with the modeling algorithm in which the median values at t=1.08 derived from the transplant carbonates must be between the median values from the native-active and native-low-activity carbonates, and therefore dotted lines are used to demonstrate behavior rather than the modeled solid and dashed lines elsewhere in the figure. In panels (a) and (c), IPL #4 (PE-DAG-C32:2) goes to zero at infinite time (i.e., the IPL is not observed in the native-low-activity carbonates).

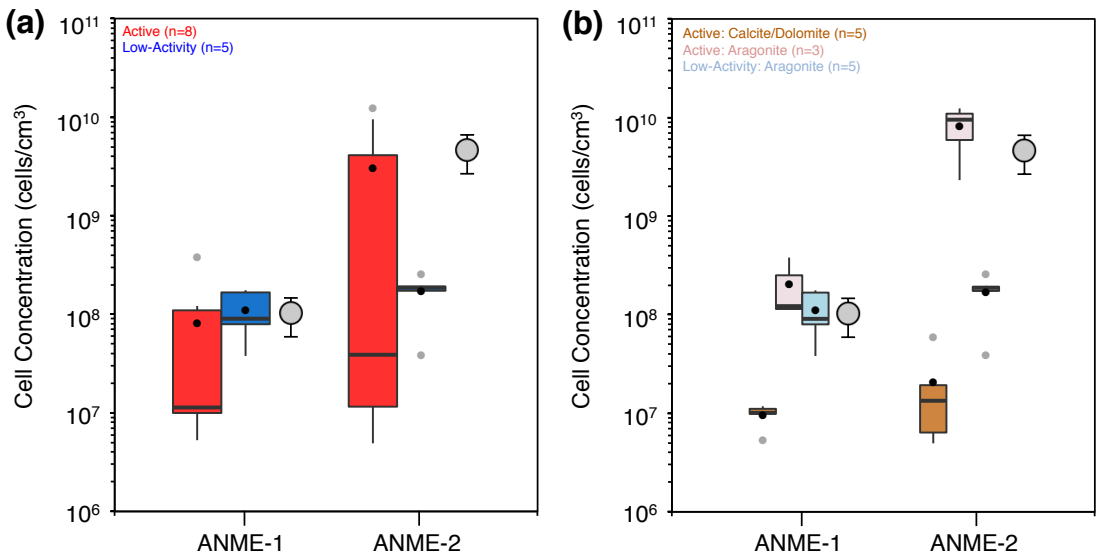


Fig. 9: Calculated cell concentrations of ANME-1 and ANME-2. Cell concentrations from seep carbonates in this study were calculated by applying conversion factors for the density of carbonate and the specific IPL content of bacterial and archaeal cells (see Supplemental Text for details). Panel (a) gives data for seep carbonates parsed by seep activity in red and blue box plots indicating active and low-activity seep conditions, respectively. Panel (b) gives data for seep carbonates parsed by both activity and mineralogy. Gray circles in both panels indicate average and standard deviations of calculated cell concentrations from reported literature values of active seep shallow sediments at Hydrate Ridge, OR (See main text for additional details; Boetius et al. 2000; Treude et al. 2003; Knittel et al. 2005; Marlow et al. 2014a). Box plots parameters are given in the caption to Fig. 2 and sketched schematically in Fig. 7a.

3.9 SUPPLEMENTAL MATERIAL: TEXT

3.9.1 MINERALOGICAL DETERMINATION BY X-RAY DIFFRACTION

In order to quantify the proportions of aragonite, calcite, and dolomite within our samples, we created a standard set of carbonate powder mixes composed of powders from pure carbonate samples. Through XRD analysis of these standards, we could use empirical data to develop a quantitative algorithm for calculating carbonate composition in our environmental samples. This approach follows naturally from previous studies which used the ratio of diagnostic calcite and dolomite peaks to determine calcite/dolomite mixing ratios in carbonate samples (Tennant and Berger 1957; Bergmann 2013). However, we expand this approach to also include aragonite inter-mixing. Recent studies have probed aragonite, calcite, and dolomite presence in seep carbonate samples, but have not reported quantitative mixing fractions of the morphologies (Marlow et al. 2014b; Case et al. 2015).

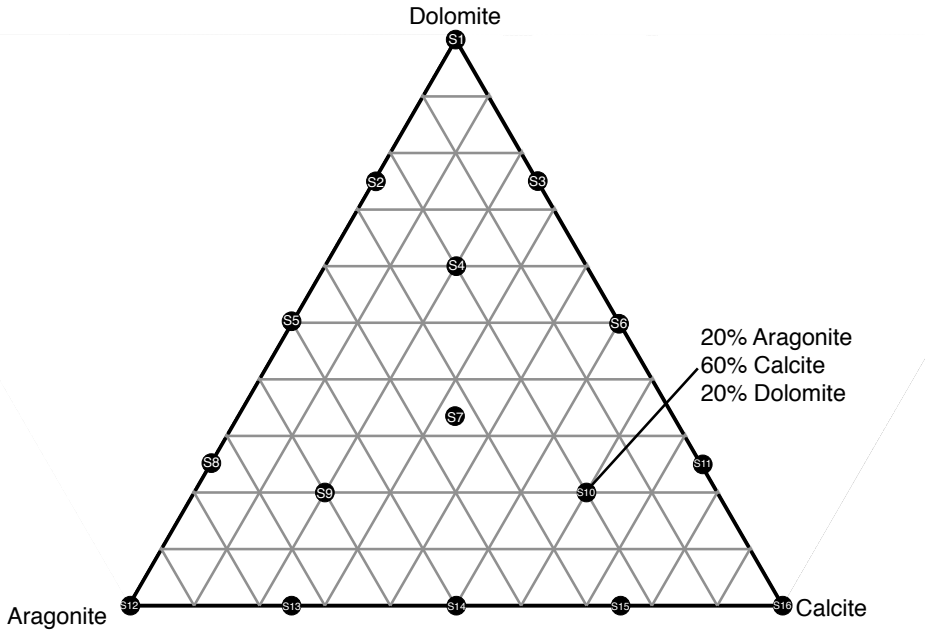


Fig. 3.9-1. Mixing fractions of 16 standard powders for developing carbonate bulk XRD quantitation.

For each morphology, we chose as diagnostic for the most dominant peak in pure-powder XRD spectra (S1, S12, and S16): aragonite (26.3° 2θ), calcite (29.5° 2θ), and dolomite (31.0° 2θ):

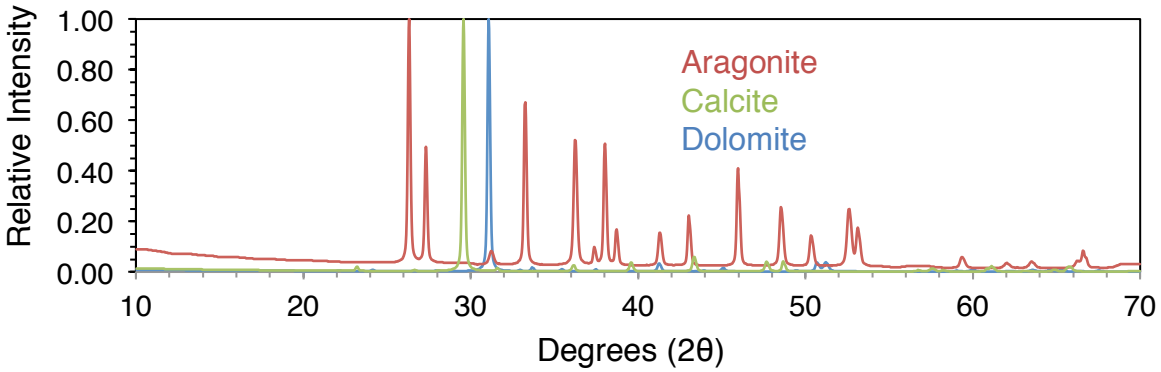


Fig. 3.9-2. Bulk XRD spectra of pure powders of aragonite, calcite, and dolomite.

By quantifying the area of the diagnostic peak for each morphology, we were able to calculate simple mixing fractions to develop a quantitative framework. Firstly, we calculate the mixing fraction of aragonite (f_{arag}) in each standard powder and compare to the known concentration ($\%_{arag}$):

$$f_{arag} = Area_{arag}/(Area_{arag}+Area_{calc}+Area_{dolo})$$

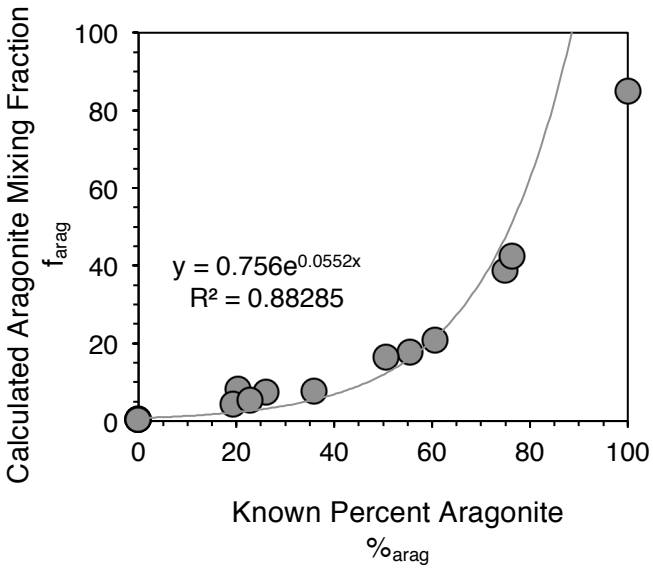


Fig. 3.9-3. Calibration curve for quantifying percent aragonite in a carbonate sample of mixed mineralogy with aragonite, calcite, and dolomite.

To a good degree, we have thus established an empirical relationship to determine the percent aragonite in a sample based on the mixing ratio of the aragonite peak at 26.3° 2θ. We can subtract this percent aragonite from 100%, and the remainder can be assigned to calcite and dolomite according to two-endmember mixing as has been described previously (Tennant and Berger 1957; Bergmann 2013):

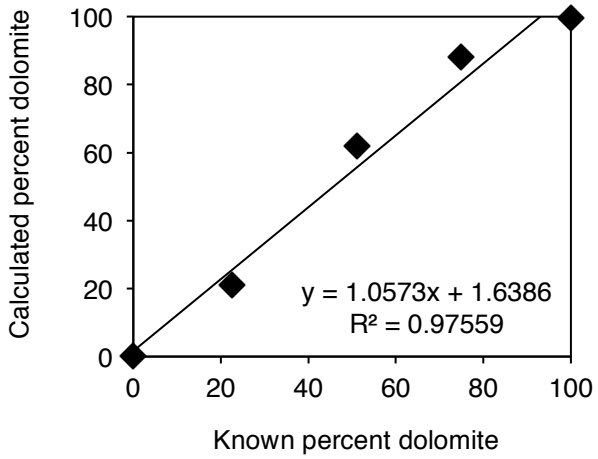


Fig. 3.9-4. Verification of accuracy of calculating mixing fractions of calcite and dolomite in standard samples of only calcite and dolomite composition.

Finally, we can compare our algorithm to the known fractions of aragonite, calcite, and dolomite in each of the 16 standard mixtures:

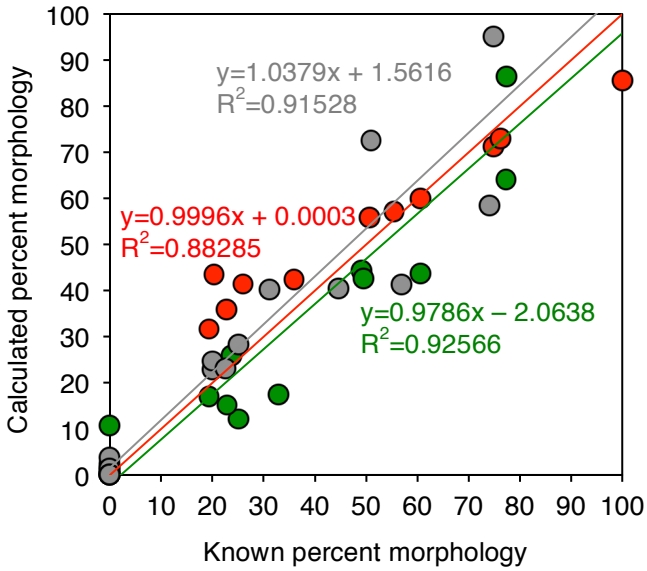


Fig. 3.9-5. Verification of our full quantitation pipeline on the suite of 16 standard powders. Red=Aragonite, Gray=Dolomite, Green=Calcite.

There is substantial scatter away from a 1:1 line among the data points. Ultimately, this empirical algorithm is able to quantitatively estimate carbonate morphology mixing fractions to within +/- 15%. Although the uncertainty is relatively high, we are able to apply this algorithm to our environmental samples to get an estimate of carbonate composition.

Plotting our environmental samples on a ternary diagram, we see that the carbonates fall into two categories: aragonitic (red) and calcitic/dolomitic (teal). The aragonitic samples are >60% aragonite, where as calcitic/dolomitic samples are <60% aragonite. Manual analysis of XRD spectra for every environmental sample confirmed the binning into aragonitic and calcitic/dolomitic bins made logical sense.

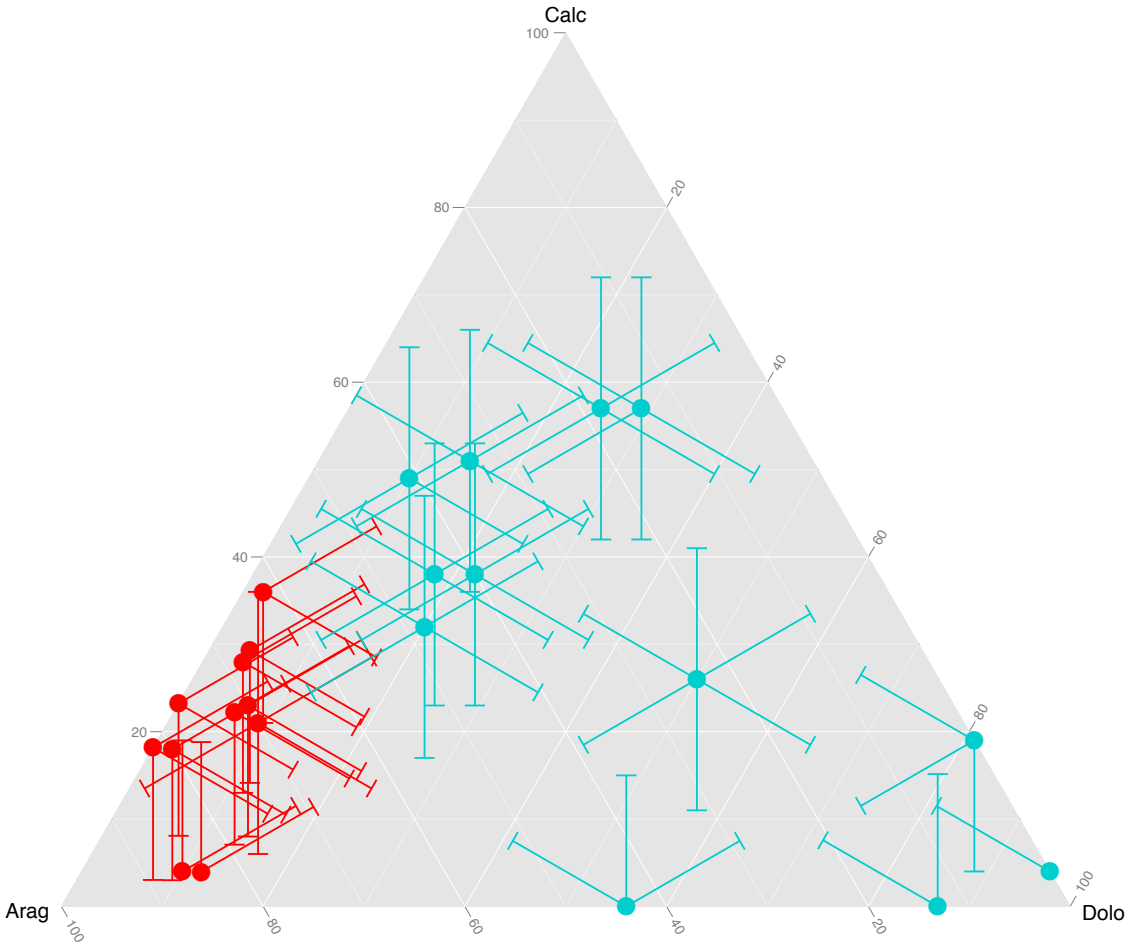
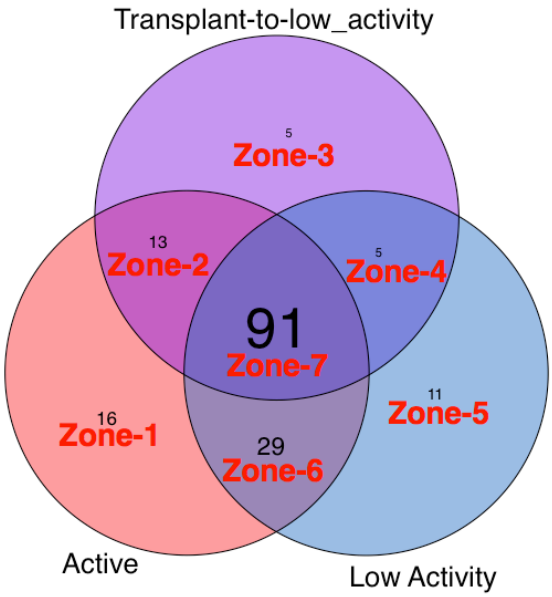


Fig. 3.9-6. Ternary diagram of environmental samples according to their mixing proportions of aragonite, calcite, and dolomite.

3.9.2 DESCRIPTION OF RICHNESS NODELING APPROACH

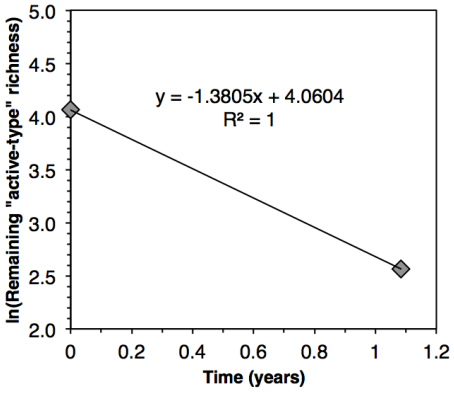


Zone Descriptions

Zone-1: Richness lost after 13 months of seep quiescence, never to be re-gained.
Zone-2: Richness retained after 13 months of seep quiescence, later to be lost permanently.
Zone-3: Richness temporarily gained after 13 months of seep quiescence, later to be lost permanently.
Zone-4: Richness gained after 13 months of seep quiescence, to be permanently kept.
Zone-5: Richness not gained yet after 13 months of seep quiescence, but later to be permanently acquired.
Zone-6: Richness loss after 13 months of seep quiescence, but later to be regained permanently.
Zone-7: Richness never lost nor gained; cosmopolitan richness.

Modeling Richness Loss

At t=0.00 yr, all richness that will be lost is represented by Zones 1, 2, and 6 (e.g., n=58).
At t=1.08 yr, Zones 1 and 6 are already lost and only Zone 2 remains (e.g., n=13).
Assuming the amount of "active-type" richness remaining is a log-linear function with time, an exponential model of decay results as defined by two data points: (0.00, ln(58)) and (1.08, ln(13)).



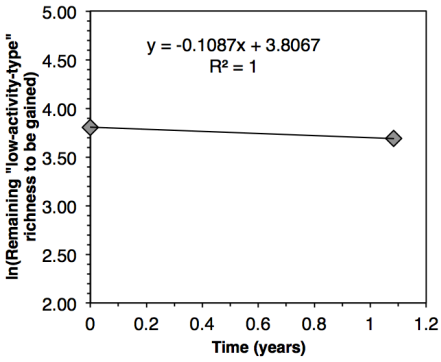
Modeling Richness Gain

Richness gain is modeled according to the amount of richness remaining to be gained.

At t=0.00, the richness that remains to be gained is represented by Zones 4, 5 and 6 (e.g., n=45).

At t=1.08, the richness that remains to be gained is represented by Zones 5 and 6 (e.g., n=40).

Assuming the amount of "low-activity-type" richness remaining to be gained is a log-linear function with time, an exponential model of decay results as defined by two data points: (0.00, ln(45)) and (1.08, ln(40)).



Putting the Model Together

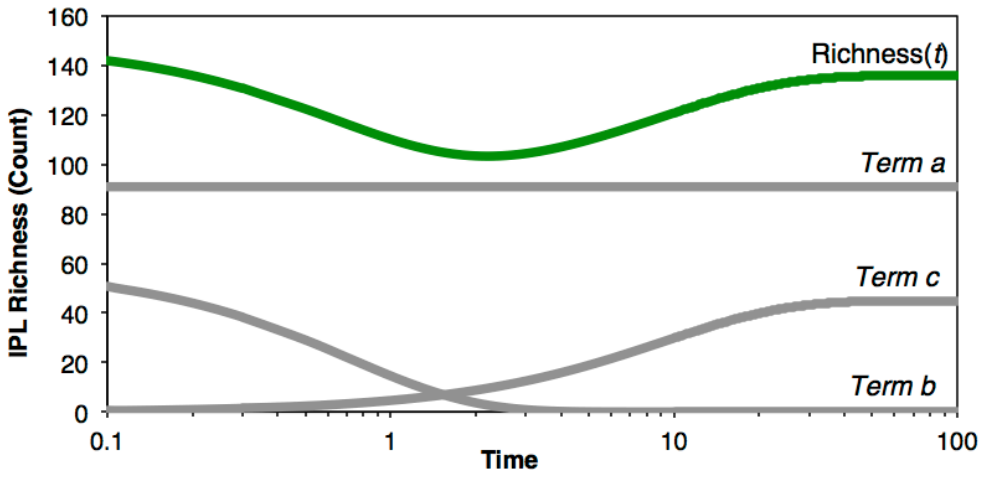
At any given time *t*, the amount of richness can be defined as the sum of three terms:

a) The constant richness never lost nor gained (Zone 7; "cosmopolitan richness").
b) The richness remaining from the "active-type" environment.
c) The richness gained from the "low-activity-type" environment.

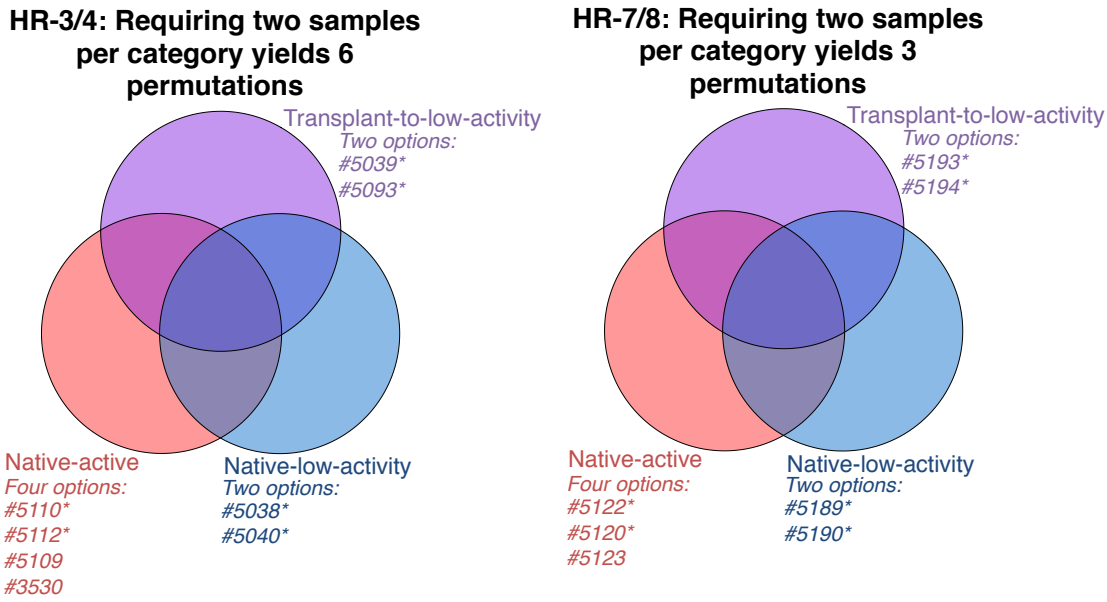
Term a, in this example, has a value of 91. *Term b* is a function of *t* defined in the first plot above. *Term c* is not quite the equation defined in the second plot above. The plot defines the amount of richness remaining to be gained, while we want to sum the amount of richness that has already been gained. This is achieved by subtracting the amount of richness remaining to be gained from the ultimate "low-activity-type" richness, in this example, 45.

The equation is thus:

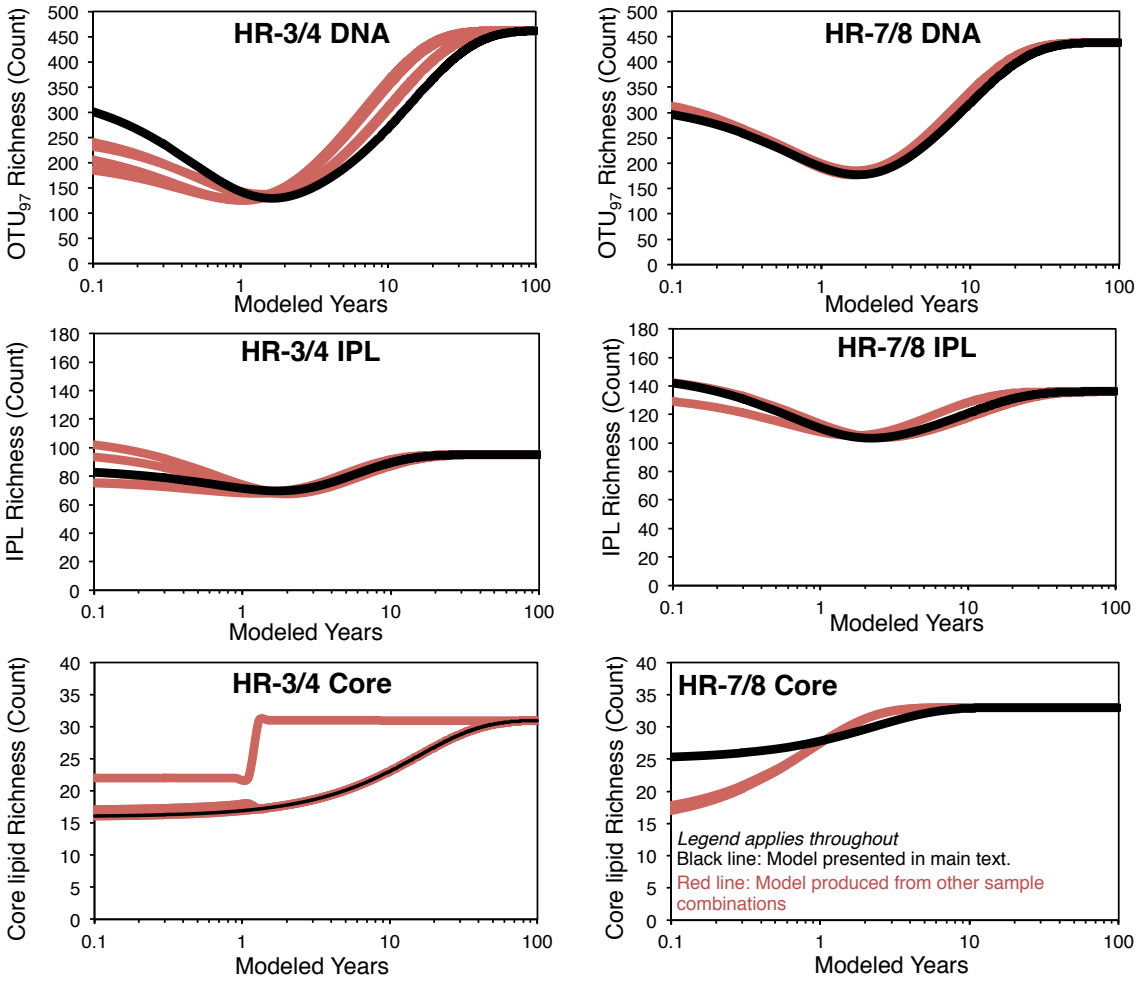
	Term a		*Term b*		*Term c*
Richness(t) =	91	+	$\exp(-1.3805t + 4.0604)$	+	$(45 - \exp(-0.1087t + 3.8067))$



Because richness as defined here (presence/absence count) is highly sensitive to the number of samples examined (discovery opportunity), we require the number of samples representing each category to be equal (in Fig. 5 of the main text, two samples are chosen for each of the native-active, transplant-to-low-activity, and native-low-activity categories). In order to test whether sample choice had a significant impact on model output, we performed the model with all possible combinations of samples representing each category (See figure on next page). Overall the model behavior is independent of the specific samples chosen, especially for 16S rRNA gene and IPL biomarkers. For core lipids, in particular at the paired HR-3/4 stations, the specific sample-sample combinations in some instances made a significant impact on the model results. This was due to some sample-sample pairs indicating "active-site-specific" core lipids, while other sample-sample pairs indicated zero "active-site-specific" core lipids (Zone 1 in the Venn diagram described above). Thus, the core lipid modeling results are the most tentative although we note that between HR-3/4 and HR-7/8, only 2 out of 9 permutations yielded the odd model structure. Seven out of nine were congruent with the "richness increase" trend presented in Fig. 6 of the main text.



*Samples with * were used for presentation in the main text and were chosen to match the samples chosen in Case et al. (2015) mBio, which were chosen randomly.*



3.9.3 CONVERTING ng/g TO cells/cm^3

IPL concentrations in this study were well characterized in ng_{IPL} per $g_{dry\ carbonate}$. In order to compare to other studies (e.g., Marlow et al. 2014a), this value is best converted into cells per cubic centimeter of carbonate. In order to perform this conversion, we applied the following assumed values:

Constants:

$\rho_{aragonite}$ = 2.93 g/cc	
$\rho_{calcite}$ = 2.71 g/cc	
$\rho_{dolomite}$ = 2.83 g/cc	
Porosity of seep carbonates = 16.3%	[average value from Marlow et al. 2014a]
Cellular IPL weight A = 18.3 fg_{IPL}/cell	[Zink et al. 2008; derived for marine subsurface bacteria]
Cellular IPL weight B = 4.9 fgIPL/cell	[Simon and Azam 1989; theoretical calculation for marine bacteria]
Cellular IPL weight C = 0.58 fg_{IPL}/cell	[Meador et al. 2014; empirical from archaeal *Thermococcus kodakarensis* pure culture]
Cellular IPL weight D = 1.4 fgIPL/cell	[Lipp et al. 2008; theoretical calculation for archaeal cells]

Based upon these values, four calculations were performed in order to generate estimates of cell density (cells/cm^3) from IPL concentrations (ng/g). The density of each carbonate was inferred by applying the densities of each $CaCO_3$ morphology with the weighted mineralogical composition of each rock, and accounting for the average porosity of seep carbonates. Conversion of IPL abundance (ng) to cellular abundance was completed separately for archaeal and bacterial IPLs, using the empirical and theoretical values given above. Thus, four estimates of cell concentration were generated:

Estimate 1: Bacterial IPLs converted with Zink et al., 2008 value; archaeal IPLs converted with Meador et al., 2014 value.

Estimate 2: Bacterial IPLs converted with Zink et al., 2008 value; archaeal IPLs converted with Lipp et al., 2008 value.

Estimate 3: Bacterial IPLs converted with Simon & Azam, 1989 value; archaeal IPLs converted with Meador et al., 2014 value.

Estimate 4: Bacterial IPLs converted with Simon & Azam, 1989 value; archaeal IPLs converted with Lipp et al., 2008 value.

These calculations result in highly similar estimates of seep carbonate cell density as a function of recoverable IPL concentrations. Therefore, the results are reported as average and standard deviation of the four calculations in Fig. 1b.

3.10 SUPPLEMENTAL MATERIAL: TABLES

Table S1. Concentrations of polar lipids were corrected using response factors based on commercial or purified standards (as listed below) to account for differences in ionization during HPLC-ESI-MS.

#	Standard	Source	Polar Lipids
1	G-GDGT 0	Zhu et al. (2013)	G-GDGT(0-5)
2	2G-GDGT 0	Zhu et al. (2013)	2G-GDGT(0-5)
3	G-GDGT-PG	Matreya LLC, PA, USA	PG-GDGT(0-4)-G/2G/PG
4	G-AR	Zhu et al. (2013)	G-AR/OH-AR
5	2G-AR	Zhu et al. (2013)	2GAR/OH-AR
6	PE-AR	Avanti Polar Lipids Inc., USA	PG/PI/PE/PS-AR*
7	$C_{21:0}/C_{21:0}$-PC	Avanti Polar Lipids Inc., USA	PC/PI (DAG/AEG/DEG)
8	Glucosylceramide	Avanti Polar Lipids Inc., USA	PE/PG/PI - Sphingolipids
9	$C_{16:0}/C_{16:0}$-DGTS	Avanti Polar Lipids Inc., USA	BL/OL
10	$C_{16:0}/C_{16:0}$-PE	Avanti Polar Lipids Inc., USA	PE (DAG/AEG/DEG)
11	$C_{16:0}/C_{16:0}$-PDME	Avanti Polar Lipids Inc., USA	PME (DAG/AEG), PDME (DAG)
12	$2C_{18:1}/2C_{18:1}$-CL	Avanti Polar Lipids Inc., USA	CL/monolysyl CL
13	$C_{16:0}/C_{16:0}$-PG	Avanti Polar Lipids Inc., USA	PG (AEG/DEG)

1. monoglycosyl glycerol dibiphytanyl glycerol tetraether; 2. diglycosyl glycerol dibiphytanyl glycerol tetraether; 3. monoglycosyl-phosphatidylglycerol glycerol dibiphytanyl glycerol tetraether; 4. Monoglycosyl archaeol; 5. diglycosyl archaeol; 6. phosphatidylethanolamine archaeol; 7. 1,2-dihenarachidoyl-sn-glycero-3-phosphatidylcholine (PC), DAG=diacylglycerol, AEG=acyletherglycerol, DEG=dietherglycerol; 8. D-glucosyl-1,1'-N-stearoyl-D-erythro-sphingosine, PI=phosphatidylinositol; 9. 1,2-dipalmitoyl-sn-glycero-3-O-4'-(N,N,N-trimethyl)-homoserine, BL=betaine lipids, OL=ornithine lipids; 10. 1,2-dipalmitoyl-sn-glycero-3-phosphatidylethanolamine (PE); 11. 1,2-dipalmitoyl-sn-glycero-3-phosphoethanolamine-N,N-dimethyl (PDME), PME=phosphatidylmethylethanolamine; 12. 1',3'-bis[1,2-dioleoyl-sn-glycero-3-phospho]-sn-glycerol or cardiolipin (CL); 13. 1,2-dipalmitoyl-sn-glycero-3-phospho-(1'-rac-glycerol). *archaeols included AR, OH-AR, 2OH-AR and Ext-OH-AR (for abbreviations of archaeal lipids please see Table SXX).

Table S2. Compilation of raw lipid data for this study (both the IPL (bacteria & archaea) and ASL (archaeal-specific IPL & core lipid) datasets).

Supplementary Table 2 can be found in .xlsx format in the Caltech online repository along with this thesis.

Table S3. Possible sources of archaeal polar lipids in carbonates according to Rossel et al. (2011), Yoshinaga et al. (2015), and Kellermann et al. (2016).

Polar Lipids*	Potential Source
G-GDGT(0-5)	ANME-1, Thaumarchaeota GDGT(5)
2G-GDGT(0-5)	ANME-1, Thaumarchaeota GDGT(5)
PG-GDGT(0-4)-G/2G/PG	ANME-1
G/2G-AR/OH-AR	ANME-1 (AR), ANME-2 (AR+OH-AR)
PE-AR/OH-AR	ANME-1? (see Wegener et al., 2016)
PG/PI/PE/PS-AR/OH-AR	ANME-1 (AR), ANME-2 (AR+OH-AR)
PG/PS-2OH-AR	ANME-2
PI-Ext-OH-AR	ANME-2

*G=monoglycosyl; 2G=diglycosyl; PG=phosphatidylglycerol; PI=phosphatidylinositol; PS=phosphatidylserine; PE=phosphatidylethanolamine; GDGT=glycerol-dibiphytanyl-glycerol-tetraether (note that the numbers in parenthesis correspond to numbers of rings, being 5 the crenarchaeol); AR=archaeol; OH-AR=hydroxylated AR; 2OH-AR=dihydroxylated AR; Ext-OH-AR=extended (C_{25}-C_{20}) OH-AR.

Table S4. Table listing Pearson R values for correlations between16S rRNA gene OTUs and each of the lipids from the IPL and ASL datasets. Values are only given if the Pearson R value is >0.8 and the p-value is <0.05. Correlations were determined from all native-active and native-low-activity carbonates.

Supplementary Table 4 can be found in .xlsx format in the Caltech online repository along with this thesis.

Table S5. Biomarker data used to generate Venn diagrams in this study (Fig. 5; Fig. S1).

Supplementary Table 5 can be found in .xlsx format in the Caltech online repository along with this thesis.

3.11 SUPPLEMENTAL MATERIAL: FIGURES

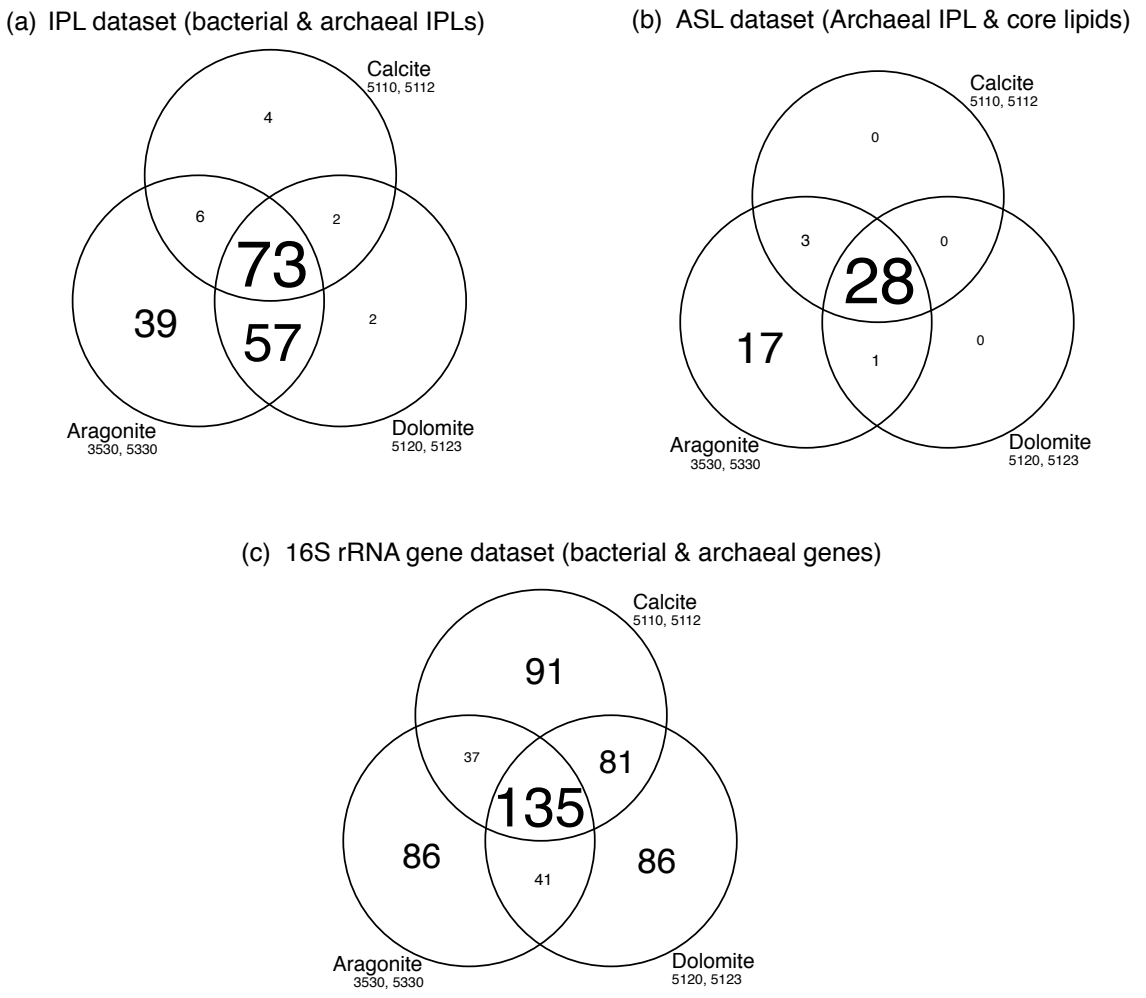


Fig. S1: Venn diagrams of the (a) IPL, (b) ASL, and (c) 16S rRNA gene datasets, binned by mineralogy. In order to avoid sample bias for any particular mineralogy, two carbonates were chosen to represent each of aragonite, calcite, and dolomite from the set of native-active carbonates. Only native carbonates from active seep sites were considered in order to remove seep activity as a factor affecting taxa distribution. The two carbonates with highest proportions of each morphology were chosen, based on XRD data. In order for a biomarker to count as "present" for this presence/absences analysis, it must have been observed in both replicates. In each field of the Venn diagrams, text size is proportional to value to help guide the readers' eye. The two IPLs in (a) identified as diagnostic of dolomite are bacterial: C65:3 cardiolipin-DAG and C31:0 PE-DAG.

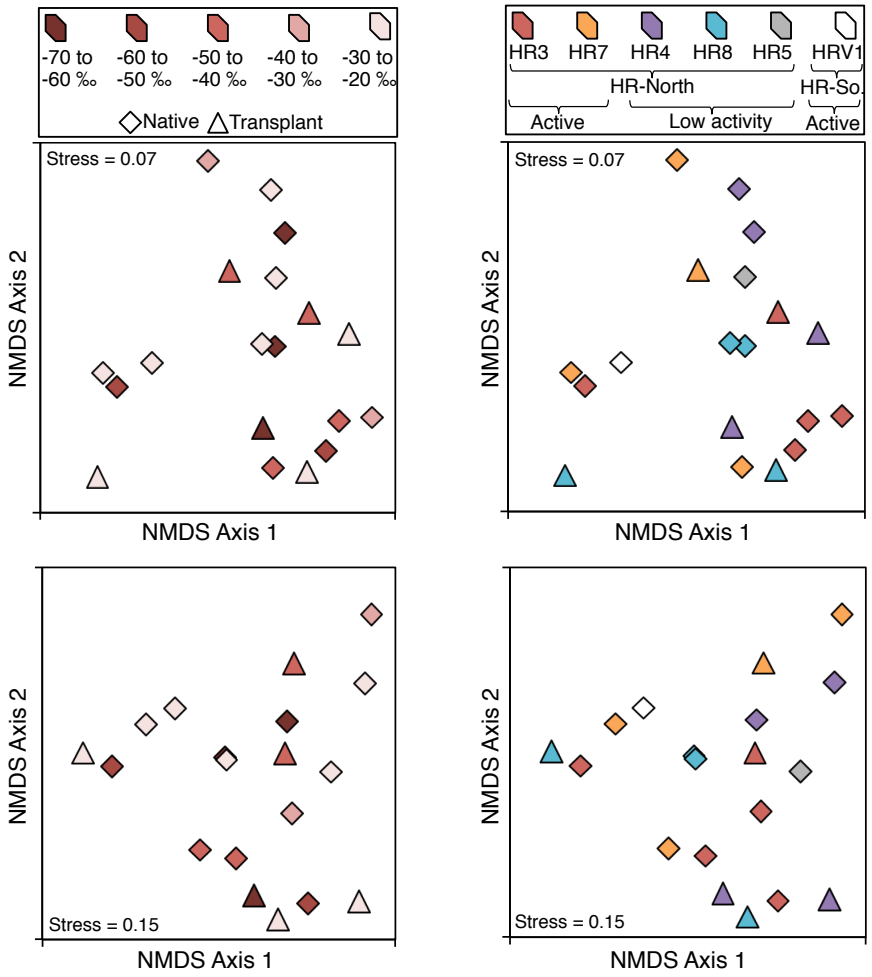


Fig S2: NMDS ordination of IPL relative abundance data. (a,c) Samples grouped by $\delta^{13}C_{org}$ value, (b,d) samples grouped by seafloor station. (a,b) Absolute abundance profiles of the IPL dataset. (c,d) Relative abundance profiles of the IPL data set. Neither $\delta^{13}C_{org}$ nor seafloor station significantly differentiate the lipid profiles.

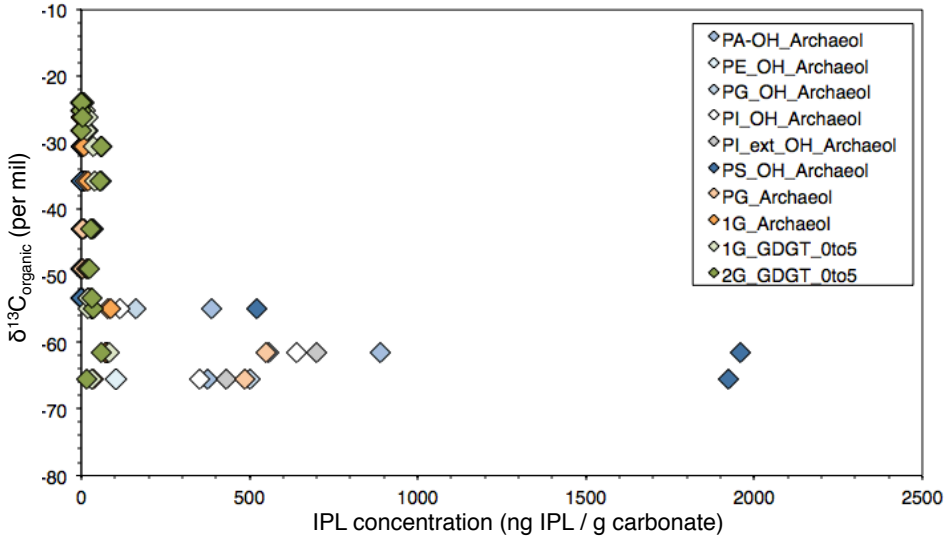


Fig. S3: Cross-plot of $\delta^{13}C_{org}$ *vs* concentration of various archaeal IPLs. Samples with the highest concentrations of OH-AR also exhibit the most depleted $\delta^{13}C_{org}$ values.

3.12 REFERENCES

Aloisi, G., I. Bouloubassi, S. K. Heijs, R. D. Pancost, C. Pierre, J. S. Sinninghe Damsté, J. C. Gottschal, L. J. Forney, and J.-M. Rouchy. 2002. CH4-consuming microorganisms and the formation of carbonate crusts at cold seeps. Earth and Planetary Science Letters **203**: 195–203.

Bayon, G., G. M. Henderson, and M. Bohn. 2009. U–Th stratigraphy of a cold seep carbonate crust. Chemical Geology **260**: 47–56.

Benincà, E., J. Huisman, R. Heerkloss, K. D. Jöhnk, P. Branco, E. H. Van Nes, M. Scheffer, and S. P. Ellner. 2008. Chaos in a long-term experiment with a plankton community. Nature **451**: 822–825.

Bergmann, K. 2013. Constraints on the carbon cycle and climate during the early evolution of animals.

Berner, R. A. 1980. Earth Diagenesis, Princeton University Press.

Birgel, D., D. Feng, H. H. Roberts, and J. Peckmann. 2011. Changing redox conditions at cold seeps as revealed by authigenic carbonates from Alaminos Canyon, northern Gulf of Mexico. Chemical Geology **285**: 82–96.

Birgel, D., J. Peckmann, S. Klautzsch, V. Thiel, and J. Reitner. 2006a. Anaerobic and Aerobic Oxidation of Methane at Late Cretaceous Seeps in the Western Interior Seaway, USA. Geomicrobiology Journal **23**: 565-577.

Birgel, D., M. Elvert, X. Han, and J. Peckmann. 2008a. ^{13}C-depleted biphytanic diacids as tracers of past anaerobic oxidation of methane. Organic geochemistry **39**: 152–156.

Birgel, D., T. Himmler, A. Freiwald, and J. Peckmann. 2008b. A new constraint on the antiquity of anaerobic oxidation of methane: Late Pennsylvanian seep limestones from southern Namibia. Geology **36**: 543–546.

Birgel, D., V. Thiel, Hinrichs, M. Elvert, K. A. Campbell, J. Reitner, J. D. Farmer, and J. Peckmann. 2006b. Lipid biomarker patterns of methane-seep microbialites from the Mesozoic convergent margin of California. Organic geochemistry **37**: 1289–1302.

Blättler, C.L., N.R. Miller, and J.A. Higgins. 2015. Mg and Ca isotope signatures of authigenic dolomite in siliceous deep-sea sediments. Earth and Planetary Science Letters **419**: 32-42. doi:10.1016/j.epsl.2015.03.006.

Blumenberg, M., E.-O. Walliser, M. Taviani, R. Seifert, and J. Reitner. 2015. Authigenic carbonate formation and its impact on the biomarker inventory at hydrocarbon seeps – A case study from the Holocene Black Sea and the Plio-Pleistocene Northern Apennines (Italy). Marine and Petroleum Geology **66**: 532–541.

Blumenberg, M., R. Seifert, J. Reitner, T. Pape, and W. Michaelis. 2004. Membrane lipid patterns typify distinct anaerobic methanotrophic consortia. Proceedings of the National Academy of Sciences **101**: 11111–11116.

Boetius, A., and E. Suess. 2004. Hydrate Ridge: a natural laboratory for the study of microbial life fueled by methane from near-surface gas hydrates. Chemical Geology **205**: 291–310.

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

Bristow, T.F., M. Bonifacie, A. Derkowski, J.M. Eiler, and J.P. Grotzinger. 2011. A hydrothermal origin for isotopically anomalous cap dolostone cements from south China. Nature **474**: 68-71. doi:10.1038/nature10096.

Burton, E. A. 1993. Controls on marine carbonate cement mineralogy: review and reassessment. Chemical Geology **105**: 163–179.

Case, D. H., A. L. Pasulka, J. J. Marlow, B. M. Grupe, L. A. Levin, and V. J. Orphan. 2015.

Methane Seep Carbonates Host Distinct, Diverse, and Dynamic Microbial Assemblages. mBio **6**: e01348–15.

Dell'Anno, A., M. Fabiano, G. C. A. Duineveld, A. Kok, and R. Danovaro. 1998. Nucleic Acid (DNA, RNA) Quantification and RNA/DNA Ratio Determination in Marine Sediments: Comparison of Spectrophotometric, Fluorometric, and HighPerformance Liquid Chromatography Methods and Estimation of Detrital DNA. Applied and Environmental Microbiology **64**: 3238–3245.

Elvert, M., A. Boetius, K. Knittel, and B. B. Jørgensen. 2003. Characterization of Specific Membrane Fatty Acids as Chemotaxonomic Markers for Sulfate-Reducing Bacteria Involved in Anaerobic Oxidation of Methane. Geomicrobiology Journal **20**: 403–419.

Faust, K., L. Lahti, D. Gonze, W. M. de Vos, and J. Raes. 2015. Metagenomics meets time series analysis: unraveling microbial community dynamics. Current Opinion in Microbiology **25**: 56–66.

Gieskes, J., C. Mahn, S. Day, J. Martin, J. Greinert, T. Rathburn, and B. McAdoo. 2005. A study of the chemistry of pore fluids and authigenic carbonates in methane seep environments: Kodiak Trench, Hydrate Ridge, Monterey Bay, and Eel River Basin. Chemical Geology **220**: 329–345.

Goedert, J. L., V. Thiel, O. Schmale, W. W. Rau, W. Michaelis, and J. Peckmann. 2003. The Late Eocene "Whiskey Creek" methane-seep deposit (western Washington State). Facies **48**: 223–239.

Greinert, J., G. Bohrmann, and E. Suess. 2001. Gas Hydrate-Associated Carbonates and Methane-Venting at Hydrate Ridge: Classification, Distribution, and Origin of Authigenic Lithologies, p. 99–113. *In* C.K. Paull and W.P. Dillon [eds.], Natural Gas Hydrates. American Geophysical Union.

Hagemann, A., T. Leefmann, J. Peckmann, V.-E. Hoffmann, and V. Thiel. 2012. Biomarkers from individual carbonate phases of an Oligocene cold-seep deposit, Washington State, USA. Lethaia **46**: 7–18.

Heijs, S. K., G. Aloisi, I. Bouloubassi, R. D. Pancost, C. Pierre, J. S. Sinninghe Damsté, J. C. Gottschal, J. D. Elsas, and L. J. Forney. 2006. Microbial Community Structure in Three Deep-Sea Carbonate Crusts. Microbial Ecology **52**: 451–462.

Hinrichs. 2002. Microbial fixation of methane carbon at 2.7 Ga: Was an anaerobic mechanism possible? Geochemistry Geophysics Geosystems **3**: 1042.

Hinrichs, R. E. Summons, V. Orphan, S. P. Sylva, and J. M. Hayes. 2000. Molecular and isotopic analysis of anaerobic methane-oxidizing communities in marine sediments. Organic Geochemistry **31**: 1685–1701.

Hoffmann-Sell, L., D. Birgel, E. T. Arning, K. B. Föllmi, and J. Peckmann. 2011. Archaeal lipids in Neogene dolomites (Monterey and Sisquoc Formations, California) – Planktic versus benthic archaeal sources. Organic geochemistry **42**: 593–604.

Jiang, G., M. J. Kennedy, and N. Christie-Blick. 2003. Stable isotopic evidence for methane seeps in Neoproterozoic postglacial cap carbonates. Nature **426**: 822–826.

Katz, M. E., D. K. Pak, G. R. Dickens, and K. G. Miller. 1999. The Source and Fate of Massive Carbon Input During the Latest Paleocene Thermal Maximum. Science **286**: 1531–1533.

Kellermann, M. Y., M. Y. Yoshinaga, G. Wegener, V. Krukenberg, and Hinrichs. 2016. Tracing the production and fate of individual archaeal intact polar lipids using stable isotope probing. Organic geochemistry **95**: 13–20.

Kiel, S., D. Birgel, K. A. Campbell, J. S. Crampton, P. Schiøler, and J. Peckmann. 2013. Cretaceous methane-seep deposits from New Zealand and their fauna. Palaeogeography, Palaeoclimatology, Palaeoecology **390**: 17–34.

Knittel, K., T. Lösekann, A. Boetius, R. Kort, and R. Amann. 2005. Diversity and Distribution of Methanotrophic Archaea at Cold Seeps. Applied and Environmental Microbiology **71**: 467–479.

Korhonen, J., J. Soininen, and H. Hillebrand. 2010. A quantitative analysis of tempoeral turnover in aquatic species assemblages across ecosystems. Ecology **91**: 508–517.

Kulm, L. D., and E. Suess. 1990. Relationship between carbonate deposits and fluid venting: Oregon Accretionary Prism. Journal of Geophysical Research: Solid Earth **95**: 8899–8915.

Leefmann, T., J. Bauermeister, A. Kronz, V. Liebetrau, J. Reitner, and V. Thiel. 2008. Miniaturized biosignature analysis reveals implications for the formation of cold seep carbonates at Hydrate Ridge (off Oregon, USA). Biogeosciences **5**: 731–738.

Levin, L. A., G. F. Mendoza, J. P. Gonzalez, A. R. Thurber, and E. E. Cordes. 2010. Diversity of bathyal macrofauna on the northeastern Pacific margin: the influence of methane seeps and oxygen minimum zones. Marine Ecology Progress Series **31**: 94–110.

Levy-Booth, D. J., R. G. Campbell, and R. H. Gulden. 2007. Cycling of extracellular DNA in the soil environment. Soil Biology & Biochemistry **39**: 2977-2991.

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

Little, C. T. S., D. Birgel, A. J. Boyce, J. A. Crame, J. E. Francis, S. Kiel, J. Peckmann, D. Pirrie, G. K. Rollinson, and J. D. Witts. 2015. Late Cretaceous (Maastrichtian) shallow water hydrocarbon seeps from Snow Hill and Seymour Islands, James Ross Basin, Antarctica. Palaeogeography, Palaeoclimatology, Palaeoecology **418**: 213–228.

Lloyd, K.G., D.B. Albert, J.F. Biddle, J.P. Chanton, O. Pizarro, and A. Teske. 2010. Spatial Structure and Activity of Sedimentary Microbial Communitites Underlying a *Beggiatoa* spp. Mat in the Gulf of Mexico Hydrocarbon Seep. PLoS ONE **5**: e8737. doi:10.1371/journal.pone.0008738.

Logemann, J., J. Graue, J. Köster, B. Engelen, J. Rullkötter, and H. Cypionka. 2011. A laboratory experiment of intact polar lipid degradation in sandy sediments. Biogeosciences **8**: 2547–2560.

Luff, R., K. Wallmann, and G. Aloisi. 2004. Numerical modeling of carbonate crust formation at cold vent sites: significance for fluid and methane budgets and chemosynthetic biological communities. Earth and Planetary Science Letters **221**: 337–353.

Marlow, J. J., J. A. Steele, W. Ziebis, A. R. Thurber, L. A. Levin, and V. J. Orphan. 2014a. Carbonate-hosted methanotrophy represents an unrecognized methane sink in the deep sea. Nature Communications **5**: 1–12.

Marlow, J., J. A. Steele, D. Case, S. A. Connon, L. A. Levin, and V. J. Orphan. 2014b. Microbial abundance and diversity patterns associated with sediments and carbonates from the methane seep environments of Hydrate Ridge, OR. Frontiers in Marine Science **1**: 1–16.

Meador, T. B., E. J. Gagen, M. E. Loscar, and T. Goldhammer. 2014. Thermococcus kodakarensis modulates its polar membrane lipids and elemental composition according to growth stage and phosphate availability. Frontiers in Microbiology **5**: 1-13.

Moore, J. C., D. Orange, and L. D. Kulm. 1990. Interrelationship of fluid venting and structural evolution: Alvin observations from the frontal accretionary prism, Oregon. Journal of Geophysical Research: Solid Earth **95**: 8795–8808.

Naehr, T.H., P. Eichhubl, V.J. Orphan, M. Hovland, C.K. Paull, W. Ussler III, T.D. Lorenson, and H.G. Greene. 2007. Authigenic carbonate formation at hydrocarbon seeps in continental margin sediments: A comparative study. Deep-Sea Research II **54**: 1268-1291. doi:10.1016/j.dsr2.2007.04.010.

Natalicchio, M., J. Peckmann, D. Birgel, and S. Kiel. 2015. Seep deposits from northern Istria, Croatia: a first glimpse into the Eocene seep fauna of the Tethys region. Geol. Mag. **152**: 444–459.

Niemann, H., and M. Elvert. 2008. Diagnostic lipid biomarker and stable carbon isotope signatures of microbial communities mediating the anaerobic oxidation of methane with sulphate. Organic geochemistry **39**: 1668–1677.

Oksanen, J., F. G. Blanchet, R. Kindt, P. Legendre, P. Minchin, R. B. OHara, G. Simpson, P.

Solymos, M. H. H. Stevens, and H. Wagner. 2013. vegan: Community Ecology Package.

Orphan, V. J., Hinrichs, W. Ussler, C. K. Paull, L. T. Taylor, S. P. Sylva, J. M. Hayes, and E. F. Delong. 2001. Comparative Analysis of Methane-Oxidizing Archaea and Sulfate-Reducing Bacteria in Anoxic Marine Sediments. Applied and Environmental Microbiology **67**: 1922–1934.

Orphan, V., W. Ussler, T. H. Naehr, C. H. House, Hinrichs, and C. K. Paull. 2004. Geological, geochemical, and microbiological heterogeneity of the seafloor around methane vents in the Eel River Basin, offshore California. Chemical Geology **205**: 265–289.

Pancost, R. D., I. Bouloubassi, G. Aloisi, J. S. Sinninghe Damsté, and T. M. S. Scientific Party. 2001. Three series of non-isoprenoidal dialkyl glycerol diethers in cold-seep carbonate crusts. Organic geochemistry **32**: 695–707.

Pasulka, A. L., L. A. Levin, J. A. Steele, D. H. Case, M. R. Landry, and V. J. Orphan. 2015. Microbial eukaryotic distributions and diversity patterns in a deep-sea methane seep ecosystem. Environmental microbiology doi:10.1111/1462-2920.13185.

Pearson, A., S. J. Hurley, S. R. S. Walter, S. Kusch, S. Lichtin, and Y. G. Zhang. 2016. Stable carbon isotope ratios of intact GDGTs indicate heterogeneous sources to marine sediments. Geochimica et Cosmochimica Acta **181**: 18–35.

Peckmann, J., and J. L. Goedert. 2005. Geobiology of ancient and modern methane-seeps. Palaeogeography, Palaeoclimatology, Palaeoecology **227**: 1–5.

Peckmann, J., D. Birgel, and S. Kiel. 2009. Molecular fossils reveal fluid composition and flow intensity at a Cretaceous seep. Geology **37**: 847–850.

Peckmann, J., J. L. Goedert, V. Thiel, W. Michaelis, and J. Reitner. 2002. A comprehensive approach to the study of methane-seep deposits from the Lincoln Creek Formation, western Washington State, USA. Sedimentology **49**: 855–873.

Peckmann, J., V. Thiel, W. Michaelis, P. Clari, C. Gaillard, L. Martire, and J. Reitner. 1999. Cold seep deposits of Beauvoisin (Oxfordian; southeastern France) and Marmorito (Miocene; northern Italy): microbially induced authigenic carbonates. Int J Earth Sci (Geol Rundsch) **88**: 60–75.

Pop Ristova, P., F. Wenzhöfer, A. Ramette, J. Felden, and A. Boetius. 2015. Spatial scales of bacterial community diversity at cold seeps (Eastern Mediterranean Sea). The ISME Journal **9**: 1306–1318.

R Core Team. 2014. R: A language and environment for statistical computing.

Rasmussen, B. 2000. Filamentous microfossils in a 3,235-million-year-old volcanogenic massive sulphide deposit. Nature **405**: 676–679.

Reeburgh, W. S. 2007. Oceanic Methane Biogeochemistry. Chemical Reviews **107**: 486–513.

Reitner, J., J. Peckmann, A. Reimer, G. Schumann, and V. Thiel. 2005. Methane-derived carbonate build-ups and associated microbial communities at cold seeps on the lower Crimean shelf (Black Sea). Facies **51**: 66–79.

Ritger, S., B. Carson, and E. Suess. 1987. Methane-derived authigenic carbonates formed by subduction-induced pore-water expulsion along the Oregon/Washington margin. Geological Society of America Bulletin **98**: 147–156.

Rossel, P. E., J. S. Lipp, H. F. Fredricks, J. Arnds, A. Boetius, M. Elvert, and Hinrichs. 2008. Intact polar lipids of anaerobic methanotrophic archaea and associated bacteria. Organic geochemistry **39**: 992–999.

Rossel, P. E., M. Elvert, A. Ramette, A. Boetius, and Hinrichs. 2011. Factors controlling the distribution of anaerobic methanotrophic communities in marine environments: Evidence from intact polar membrane lipids. Geochimica et Cosmochimica Acta **75**: 164–184.

Ruff, S. E., J. F. Biddle, A. P. Teske, K. Knittel, A. Boetius, and A. Ramette. 2015. Global dispersion and local diversification of the methane seep microbiome. Proceedings of the National Academy of Sciences 1–6.

Rütters, H., H. Sass, H. Cypionka, and J. Rullkötter. 2001. Monoalkylether phospholipids in the

sulfate-reducing bacteria Desulfosarcina variabilis and Desulforhabdus amnigenus. Archives of microbiology **176**: 435–442.

Sahling, H., D. Rickert, R. W. Lee, P. Linke, and E. Suess. 2002. Macrofaunal community structure and sulfide flux at gas hydrate deposits from the Cascadia convergent margin, NE Pacific. Marine Ecology Progress Series **231**: 121–138.

Schrag, D. P., J. A. Higgins, F. A. Macdonald, and D. T. Johnston. 2013. Authigenic Carbonate and the History of the Global Carbon Cycle. Science **339**: 540–543.

Schubotz, F., J. S. Lipp, M. Elvert, and Hinrichs. 2011. Stable carbon isotopic compositions of intact polar lipids reveal complex carbon flow patterns among hydrocarbon degrading microbial communities at the Chapopote asphalt volcano. Geochimica et Cosmochimica Acta **75**: 4399–4415.

Shade, A., J. G. Caporaso, J. Handelsman, R. Knight, and N. Fierer. 2013. A meta-analysis of changes in bacterial and archaeal communities with time. The ISME Journal **7**: 1493–1506.

Simon, M., and F. Azam. 1989. Protein content and protein synthesis rates of planktonic marine bacteria. Marine ecology progress series **51**: 201-213.

Slotznick, S.P., and W.W. Fischer. 2016. Examining Archean methanotrophpy. Earth and Planetary Science Letters **441**: 52-59. doi:10.1016./j.epsl.2016.02.013.

Stadnitskaia, A., D. Nadezhkin, B. Abbas, V. Blinova, M. K. Ivanov, and J. S. Sinninghe Damsté. 2008. Carbonate formation by anaerobic oxidation of methane: Evidence from lipid biomarker and fossil 16S rDNA. Geochimica et Cosmochimica Acta **72**: 1824–1836.

Sturt, H. F., R. E. Summons, K. Smith, M. Elvert, and Hinrichs. 2004. Intact polar membrane lipids in prokaryotes and sediments deciphered by high-performance liquid chromatography/electrospray ionization multistage mass spectrometry—new biomarkers for biogeochemistry and microbial ecology. Rapid Commun. Mass Spectrom. **18**: 617–628.

Suess, E., B. Carson, S. D. Ritger, J. C. Moore, M. L. Jones, L. D. Kulm, and G. R. Cochrane. 1985. Biological communities at vent sites along the subduction zone off Oregon. Bulletin of the Biological Society of Washington **6**: 475–484.

Teichert, B., A. Eisenhauer, G. Bohrmann, A. Haase-Schramm, B. Bock, and P. Linke. 2003. U/Th systematics and ages of authigenic carbonates from Hydrate Ridge, Cascadia Margin: Recorders of fluid flow variations. Geochimica et Cosmochimica Acta **67**: 3845–3857.

Teichert, B., G. Bohrmann, and E. Suess. 2005. Chemoherms on Hydrate Ridge--Unique microbially-mediated carbonate build-ups growing into the water column. Palaeogeography, Palaeoclimatology, Palaeoecology **227**: 67–85.

Tennant, C. B., and R. W. Berger. 1957. X-ray determination of dolomite-calcite ratio of a carbonate rock. American Mineralogist **42**: 23–29.

Thiel, V., J. Peckmann, R. Seifert, P. Wehrung, J. Reitner, and W. Michaelis. 1999. Highly isotopically depleted isoprenoids: molecular markers for ancient methane venting. Geochimica et Cosmochimica Acta **63**: 3959–3966.

Torti, A., M. A. Lever, and B. B. Jørgensen. 2015. Origin, dynamics, and implications of extracellular DNA pools in marine sediments. Marine Genomics **24**: 185–196.

Trembath-Reichert, E., D. H. Case, and V. J. Orphan. 2016. Characterization of microbial associations with methanotrophic archaea and sulfate-reducing bacteria through statistical comparison of nested Magneto-FISH enrichments. PeerJ **4**: e1913–31.

Treude, T., A. Boetius, K. Knittel, K. Wallmann, and B. B. Jørgensen. 2003. Anaerobic oxidation of methane above gas hydrates at Hydrate Ridge, NE Pacific Ocean. Marine Ecology Progress Series **264**: 1–14.

Tryon, M. D., K. M. Brown, and M. E. Torres. 2002. Fluid and chemical flux in and out of sediments hosting methane hydrate deposits on Hydrate Ridge, OR, II: Hydrological processes. Earth and Planetary Science Letters **201**: 541–557.

Tryon, M. D., K. M. Brown, M. E. Torres, A. M. Tréhu, J. McManus, and R. W. Collier. 1999. Measurements of transience and downward fluid flow near episodic methane gas vents,

Hydrate Ridge, Cascadia. Geology **27**: 1075–1078.

Van Lith, Y., R. Warthmann, C. Vasconcelos, and J. A. Mckenzie. 2003. Microbial fossilization in carbonate sediments: a result of the bacterial surface involvement in dolomite precipitation. Sedimentology **50**: 237–245.

Vasconcelos, C., J. A. Mckenzie, S. Bernasconi, D. Grujic, and A. J. Tiens. 1995. Microbial mediation as a possible mechanism for natural dolomite formation at low temperatures. Nature **377**: 220–222.

Warthmann, R., C. Vasconcelos, H. Sass, and J. A. Mckenzie. 2005. Desulfovibrio brasiliensis sp. nov., a moderate halophilic sulfate-reducing bacterium from Lagoa Vermelha (Brazil) mediating dolomite formation. Extremophiles **9**: 255–261.

Wegener, G., and A. Boetius. 2009. Short-term changes in anaerobic oxidation of methane in response to varying methane and sulfate fluxes. Biogeosciences **6**: 867–876.

White, D., D. B. Ringelberg, S. J. Macnaughton, S. Alugupalli, and D. Schram. 1997. Signature Lipid Biomarker Analysis for Quantitative Assessment In Situ of Environmental Microbial Ecology, p. 22–34. *In* R.P. Eganhouse [ed.], Molecular Markers in Environmental Geochemistry. American Chemical Society.

White, D., W. Davis, J. Nickels, J. King, and R. Bobbie. 1979. Determination of the sedimentary microbial biomass by extractible lipid phosphate. Oecologia **40**: 51–62.

Wörmer, L., J. S. Lipp, J. M. Schröder, and Hinrichs. 2013. Application of two new LC–ESI–MS methods for improved detection of intact polar lipids (IPLs) in environmental samples. Organic geochemistry **59**: 10–21.

Xie, S., J. S. Lipp, G. Wegener, T. G. Ferdelman, and Hinrichs. 2013. Turnover of microbial lipids in the deep biosphere and growth of benthic archaeal populations. Proceedings of the National Academy of Sciences **110**: 6010–6014.

Yoshinaga, M. Y., C. S. Lazar, M. Elvert, Y.-S. Lin, C. Zhu, V. B. Heuer, A. Teske, and Hinrichs. 2015. Possible roles of uncultured archaea in carbon cycling in methane-seep sediments. Geochimica et Cosmochimica Acta **164**: 35–52.

Yoshinaga, M. Y., M. Y. Kellermann, P. E. Rossel, F. Schubotz, J. S. Lipp, and Hinrichs. 2011. Systematic fragmentation patterns of archaeal intact polar lipids by high-performance liquid chromatography/electrospray ionization ion-trap mass spectrometry. Rapid Commun. Mass Spectrom. **25**: 3563–3574.

Zink, K.-G., H. Wilkes, U. Disko, M. Elvert, and B. Horsfield. 2003. Intact phospholipids—microbial "life markers" in marine deep subsurface sediments. Organic geochemistry **34**: 755–769.

Zink, K.-G., K. Mangelsdorf, L. Granina, and B. Horsfield. 2008. Estimation of bacterial biomass in subsurface sediments by quantifying intact membrane phospholipids. Anal Bioanal Chem **390**: 885–896.

C h a p t e r F o u r

DEVELOPMENT OF AN IN-HOUSE PREPARATION, PROCESSING, AND ANALYSIS WORKFLOW FOR NEXT GENERATION SEQUENCING DATA

David H. Case[1]

in collaboration with,

Alexis Pasulka[1], Elizabeth Trembath-Reichert[1], Stephanie Connon[1], Katherine Dawson[1], and Victoria Orphan[1]

[1]Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA, USA

This chapter will be made publically available through the Caltech thesis repository, as well as by request to the authors.

4.0 ABSTRACT

Next-generation sequencing ("iTag", "NGS") has revolutionized microbial surveys in the last decade, enabling orders of magnitude advances in sample throughput and per-sequence cost. As expected, many aspects of the upstream (i.e., sample preparation) and downstream (i.e., data processing) employment of iTag have required development in order to fully and accurately utilize new technologies. In this chapter I discuss the Orphan lab's development of a workflow for upstream and downstream iTag methods (Fig. 1). As a result of our methods testing, we are able to make estimates of iTag precision (i.e., reproducibility, estimated at 0.77-1.85% relative abundance) and accuracy (variable depending on taxa and polymerase enzyme; range: 20x underrepresentation to 7x overrepresentation). The field of next-generation sequencing remains in rapid development (and may be moving into so-called "next-next-generation sequencing" or "3rd generation sequencing"), such that workflows and analysis techniques are in constant need of development, improvement, and updating. What follows is in many ways a "state of the lab" snapshot that will change quickly as time passes. For example, as of this writing the Orphan lab is considering switching to a newer, more flexible approach of sample preparation that would allow simultaneous sequencing of multiple gene targets; this approach was not feasible even a year or two ago but will likely herald a new period of sequencing moving forward. At the end of this chapter, a primer is included to familiarize the reader with some of the quantitative ecological techniques most frequently employed in this thesis: Nonmetric Muldi-Dimensional Scaling (NMDS), Analysis of Similarity (ANOSIM), and Similarity Percentage (SIMPER).

4.1 INTRODUCTION AND PREPARING SAMPLES FOR SEQUENCING

The process of DNA extraction by definition isolates genomic material ("gDNA") from microbial cells (although not without biases depending on extraction method – see Morono et al. 2014). While the emergent fields of genomics and meta-genomics, enabled by next-generation sequencing technologies, interpret entire genomic content, in many cases a particular experiment only calls for analysis of specific genes. The most common assay is analysis of the 16S rRNA gene, a core ribosomal gene (Woese and Fox 1977). The 16S rRNA gene is conserved across all microbial life, is vertically inherited, and mutates at a slow but steady rate across evolutionary time. These characteristics make the 16S rRNA gene a high-quality marker of inter-organism relatedness, and therefore a metric of phylogenetic identity and whole-community diversity. In order to isolate the 16S rRNA gene from genomic content, a polymerase chain reaction (PCR) is employed (Kleppe et al. 1971; Mullis and Faloona 1987; Saiki et al. 1988). PCR is a flexible approach in which any region of genomic interest can be amplified in a series of repeating reactions ("cycles"). In principle, each cycle of a PCR doubles the concentration of the gene of interest; PCR is regularly performed for 20-35 cycles, thereby amplifying the gene of interest by many orders of magnitude over the initial gDNA template.

The entire 16S rRNA gene is ~1,500 base pairs (bp) in length and includes nine hypervariable regions interspersed within conserved regions. Sequencing of the hypervariable regions enables robust inter-species resolution, and ideally an experiment will include sequencing of all nine regions. Sequencing technology has evolved over the last four decades since the "molecular revolution" began with the advent of "Sanger sequencing" (Sanger and Coulson 1975; Sanger et al. 1977a; b). Sanger sequencing offers robust, high-quality, and most notably, nearly full-length sequencing of genes of interest, including hypervariable regions of the 16S rRNA gene. However, on a per-sequence basis Sanger sequencing is expensive and due to cost and time constraints is often limited to ~100 sequences per sample. Nonetheless, Sanger

sequencing remained the gold standard of DNA sequencing technology for three decades, until the advent of "next-generation sequencing", a new wave of DNA sequencing technologies which offered orders of magnitude more sequences per analysis as well as the ability to "multiplex" – to sequence hundreds of samples simultaneously. The development of next-generation sequencing is marked by numerous corporate acquisitions and rapid invention and obsolescence of technologies on a year-to-year timescale. Roughly, next-generation sequencers began to be commercially viable with the advent of 454/Roche Sequencing (Margulies et al. 2005), which offered thousands of sequences per sample, albeit at truncated length (a maximum length of several hundred bp – not enough for the entire 16S rRNA gene). Several years later, Illumina, Inc. released a commercially available sequencing platform that increased the number of sequences per sample by an order of magnitude and increased base calling accuracy, but again at the cost of shorter sequences – often limited to one hypervariable region of the 16S rRNA gene (Bentley et al. 2008). Illumina sequencing has remained the dominant method for massively parallel single-gene sequencing, enabling microbial ecology studies that span many environments and time points. The technology has additionally improved to allow longer sequences without sacrificing quality. However, a new "next-next-generation" of DNA sequencing is in development, incorporating technologies recently released by Ion Torrent Systems Inc. and Pacific Biosciences (Schadt et al. 2010).[1]

In order to prepare gDNA for sequencing of the 16S rRNA gene (the Orphan lab uses an Illumina MiSeq platform operated at Laragen, Inc.), two PCR steps are employed (see sections below for more details on the development of sample preparation protocols). First, duplicate PCRs are performed to amplify the V4 region of the 16S rRNA gene for each sample. These products, after being checked for quality by gel electrophoresis, are pooled and transferred to a second PCR reaction in which unique barcodes are appended to amplicons, as well as

[1] These paragraphs describing a brief overview of DNA sequencing are not comprehensive, do not describe all corporate and academic contributions to the development of the field, and are not an endorsement of any one particular DNA sequencing technology.

oligonucleotide adapters which bind the amplicon to the Illumina MiSeq flow cell upon sequencing. At this point, the barcoded amplicons from each sample are uniquely tagged and therefore may be combined into a single mixture. Prior to combination, each sample's amplicon pool is quantified by fluorescence assay. Samples are then combined by adding an equi-molar amount of each sample to one batch tube, so that no single sample swamps the signal from all others. This single aliquot, containing uniquely tagged 16S rRNA gene amplicons from hundreds of different samples, is passed through a PCR cleanup kit (Qiagen, Inc) and shipped to Laragen, Inc. for sequencing.

4.2 PROCESSING RAW DATA

The following section will describe the Orphan lab's workflow for processing raw Illumina MiSeq data, followed by a section detailing tests which were run in order to determine the best PCR practices for preparing environmental iTag samples. Briefly, this section will include discussion of three sample types: negative controls, plasmid mock communities, and genomic mock communities. Negative controls are PCR reactions which were run with zero gDNA template added (1 µL of PCR-grade water was used as a volume substitute) and amplified for enough cycles to produce a product which could then be sequenced and processed like any other sample. The plasmid mock communities (n=4) were generated by mixing known ratios of plasmids from uncultured methane seep organisms (Table 1). The genomic mock communitites (n=4) were generated by mixing known ratios of gDNA extract from cultured organismsm grown in the laboratory (Table 2). Both types of mock communities are employed to test iTag precision; the plasmid mock communities are additionally used to test iTag accuracy.

4.2.1 JOINING & QUALITY CHECKING, CHIMERA DETECTION, & SINGLETON REMOVAL

When sequences are first generated on an Illumina platform, they are produced as "paired-end reads". Paired-end reads consist of two separate sequences ("R1" and "R2") which represent sequencing from opposite ends of a single amplicon, respectively[2]. If the amplicon is short enough, the forward and reverse sequencing overlap one another. This allows the two reads to be "joined" into a single contig, in which they are overlapped and base calls are checked against one another. There are many algorithms for joining, and in practice our lab has settled on a commonly used software package, fastq-join (Aronesty 2011). Our implementation of this software requires a minimum overlap of 50 bp between the forward and reverse reads, with no more than an 8% difference in base calls within the overlapping region (therefore, no more than 4 mismatched base calls in a 50 bp overlap region). If a paired R1 and R2 read fail to meet these two criteria, they are removed from the dataset. In instances where 4 or fewer mismatches are identified, the contig is assigned at that position the base call which corresponds to the higher quality value from the R1 or R2 read[3]. In our datasets, there is wide range in the proportion of sequences removed during joining (17%±9%; Fig. 2).

Joining the R1 and R2 reads inherently provides initial quality filtering of the raw sequence data, by virtue of checking the forward read against the reverse read. However, further quality filtering is needed in order to assess non-overlapping regions of the contig. Therefore, a quality-filtering step is performed in which contigs must meet two criteria: first, a contig is not

---

[2] For an excellent description of Illumina MiSeq sequencing, see the following video introduction produced by Illumina, Inc.: https://www.youtube.com/watch?v=womKfikWlxM

[3] Quality values, termed "Q-scores" or "Phred scores", are a value indicating the confidence that an assigned base call is correct. Q-scores and probability are related by the following equation:

$$P = 10^{(-Q/10)}$$

For example, a Q-score of 30 corresponds to 10^{-3}, or a 1-in-1,000 probably the base call is incorrect (99.9% certainty). In practice, the most frequently used cutoff values for "acceptable" Q-scores are 20, 25, or 30.

allowed to have any "N" base calls in which a base is unidentified. Given even just one "N" base call, a contig is removed from the dataset. Second, every base call of a contig must have at least a Q-score of 30, meaning that after quality filtering every base of every contig is known to greater than or equal to 99.9% certainty. In practice only a small proportion of sequences are removed at this step (2%±1%, Fig. 2), but these quality filtering steps, applied with the QIIME software (Caporaso et al. 2010), improve our confidence in downstream interpretation of the sequencing profiles of our samples.

Quality filtering is an effective method of improving the dataset's veracity, but further steps are necessary before proceeding to data interpretation. After quality filtering, we apply a chimera checking step. Chimeras are oligonucleotide fragments which arise during a PCR reaction, in which two 16S rRNA gene fragments are erroneously combined into a single amplicon. These amplicons may be relatively abundant in the dataset and during sequencing may be assigned high quality scores; they are not, however, meaningful in the context of a given study and therefore ought to be removed. We achieve this by employing the UCHIME algorithm within the USEARCH software package (Edgar et al. 2011), which checks whether one half of a contig aligns strongly to a database sequence (Seq_A) while the other half of a contig aligns more strongly to a different database sequence (Seq_B). If so, the contig is identified as chimeric and removed from the dataset. If not (e.g., both halves of the contig align best to Seq_A), then the contig is deemed non-chimeric and retained in the dataset. For samples amplified with the 5-PRIME polymerase enzyme, this step usually removes a moderate proportion of sequences (5%±3%; Fig. 2); for samples amplified with the NEB Q5 polymerase, the removal rate is higher (~10-20%, data not shown). The specific parameters of chimera checking in UCHIME are highly modifiable; we apply the default stringencies for chimera detection, but future users could tune the parameters as desired. In particular, chimera detection on short Illumina sequences (in our case, <300 bp in length and covering only one hypervariable region of the 16S rRNA gene) presents novel challenges not faced when detecting chimeras on full-length 16S rRNA gene sequences. With

only one hypervariable region to analyze, is may not be feasible to identify two spliced 16S rRNA genes in a single contig (Nelson et al. 2014; Ruiz-Calderon et al. 2016), and some recent studies have not included chimera detection (Metcalf et al. 2016; Ruiz-Calderon et al. 2016). As iTag processing methodologies continue to develop, this question ought to be further addressed. Nonetheless, it remains common (arguably standard) in the community to apply chimera detection to short iTag sequences (Kozich et al. 2013; Nelson et al. 2014; Dominguez-Bello et al. 2016), as it has been for the published studies in this thesis (Mason et al. 2015; Case et al. 2015). Furthermore, as iTag technology improves and longer 16S rRNA sequences become feasible, covering multiple hypervariable regions, chimera detection will once more become unquestionably relevant and applicable, and therefore will remain a step in iTag processing workflows in the future.

It is possible at this point that hundreds or thousands of individual sequences are highly similar (or even 100% identical) to one another, and therefore these sequences may be grouped into clusters in order to save computational time. This is done by creating operational taxonomic units (OTUs): bins of highly similar sequences (Edgar, 2010). Most often, these OTUs are created by requiring 97% or 99% similarity of all sequences within the group. Because a fundamental interest of microbial ecologists is to identify the many species present in their dataset, a taxonomy is matched to each OTU (Wang et al., 2007), wherein as representative of each OTU the most frequently occurring oligonucleotide sequence is chosen. These steps are performed in the QIIME software package.

Even after these steps, "spurious" OTUs will remain which pass all quantitative quality thresholds but nonetheless are unlikely to represent genuine microbial community information. At this point a singleton-removal cutoff is applied, in which OTUs that occur one time in one sample in the entire dataset (i.e., a lone sequence dissimilar to any other sequence among the entire dataset). This step does not eliminate many sequences from the dataset (~1%; Fig. 2), but is nonetheless important for filtering out spurious sequencing data.

4.2.2 NEGATIVE CONTROLS

As with any method, it is critical to address contamination through analysis of negative controls. In iTag sequencing, a negative control is particularly important because the depth of sequencing allows unprecedented views into the "rare biosphere", those OTUs which are present at roughly less than 1-in-100 abundance in the microbial community, and thus would have been missed in the majority of clone library studies employing Sanger sequencing. However, iTag sequencing is fundamentally dependent on a PCR reaction at the very first step of sample preparation. By definition PCR reactions amplify low amounts of DNA into higher concentrations; while this is generally a benefit that has enabled the "molecular revolution" over the last ~40 years, it also presents challenges for iTag sequencing. Not even the most pure PCR-grade water is truly clean of genomic content, nor are the other components of a PCR reaction, i.e., dNTP mixes, polymerase enzymes, and oligonucleotide primers. This is not generally a problem, as the contaminant genomic content present in PCR reagents is usually swamped by the (relatively) high concentration of one's sample gDNA introduced as template. To the extent that contaminant genomic material is amplified, it is often in low enough relative bundance as to not be captured in clone libraries with low sequencing depth. However, if template gDNA concentration is low, approaching parity with contaminant DNA, then the resulting amplified product can be a mixture of genuine, sample-derived gene amplicons and a significant fraction of contaminant gene amplicons. Even in PCR reactions where the overall template gDNA concentration is orders of magnitude higher than contaminant genomic content, the 16S rRNA gene copies of a particular organism (e.g., a species present at 1-in-10,000 abundance in the bulk microbial community) may still approach parity with contaminant genomic content. In that case, it becomes critical to define which sequences are trusted as genuine and which are suspected to be

contaminants, as well as the threshold below which a user no longer trusts any iTag sequences to reflect genuine signal from low-abundance community members.

In order to address negative control contamination in iTag data, we have employed two approaches: one "pre-PCR" laboratory practice and one bioinformatics technique. The "pre-PCR" approach, while obvious, is unfortunately not always practiced in microbiology laboratories. Simply put, during iTag PCR preparation in our laboratory, we take extra care to maintain a working environment clean of possible exogenous sources of genomic contamination. When setting up PCR reactions with high-concentration template and for the purpose of generating clone libraries (where only community members greater than ~1% of the population will likely be observed), it is not uncommon for users to set up the PCR reaction on the bench-top after a wipe-down with ethanol and/or bleach and perhaps an open flame to sterilize surrounding air. However, these approaches are insufficient for iTag PCR setup. Instead, we perform all PCR setup in a Purifier Class II Biosafety Cabinet (Labconco) whose airflow is HEPA filtered to greater than >99.99% particulate purity. The cabinet itself is wiped before and after each use with RNAse AWAY (Molecular Bio-Products, Inc.), a surfactant specifically designed to clean PCR equipment. A dedicated set of pipets and pipet tips is left in the cabinet at all times, and these equipment are also wiped with RNAse AWAY before and after use. Furthermore, the pipets, pipet tips, plastic PCR strips, and the interior of the cabinet itself are all subjected to ultraviolet light sterilization before and after every use in order to break down contaminant gDNA. While none of these techniques are able to address genomic contamination in PCR reagents themselves (e.g., dNTPs, polymerase enzymes), they do at least minimize the chances of exogenous genomic contamination during PCR setup. As a part of this "pre-PCR" practice, with every iTag sequencing run we include at least one negative control PCR reaction. This is a reaction in which PCR-grade water is used as the "template", and amplification is run for enough cycles that an amplicon band is observed on electrophoresis gel (environmental samples are generally amplified for 30 cycles in the first PCR, while the negative control usually requires 35-37 cycles in order to

generate an observable product). From there on, the sample is treated like any regular environmental sample in the preparation and sequencing pipeline.

Including a negative control in the sequencing enables our bioinformatics approache to addressing contamination. This approach is simply to identify which OTUs are present in the sequences from the negative control and remove them from the samples, not unlike a blank subtraction (Fig. 3). Since our lab began iTag sequencing November 2013, we have performed six runs (2013-11, 2014-05, 2014-11, 2015-03, 2015-09, 2015-12) and have used two polymerase enzymes (5-PRIME Hot Master Mix and NEB Q5, see discussion below). Over those runs and with both enzymes, we have observed contaminant OTUs associated with a range of taxonomies. Negative controls amplified with the 5-PRIME enzyme have more consistent composition than the negative controls amplified with NEB Q5. 5-PRIME negative controls are rich in a Gammaproteobacterial OTU (#57 in Fig. 3) as well as OTUs associated with Betaproteobacteria, Firmicutes, and Bacteroidetes. Negative controls amplified with the NEB Q5 enzyme, while also occasionally containing OTUs associated with those taxa, also have exhibited OTUs associated with Acidobacteria, Actinobacteria, Planctomycetes, and Alphaproteobacteria. This wide range of OTUs associated with negative controls demonstrates the importance of including a negative control in every iTag run. This enables our bioinformatics approach to addressing contamination, which is not fundamentally different than a "blank correction" applied across a variety of disciplines: for each run, we identify the OTUs present in the corresponding negative control and subtract those OTUs from all environmental samples in that run's dataset. This is bioinformatically simple to do, and can be achieved by two methods. The first method, which is more conservative, is to identify the taxonomies associated with negative control OTUs for a given iTag run (e.g., in November 2014 the 5-PRIME negative control contained Clostridia-associated OTUs; Fig. 3). Then, all OTUs associated with that taxonomy can be removed from the environmental dataset. While this method conservatively removes data which might be contaminant-related, it has the potential to remove OTUs from the environmental dataset which

are genuine (e.g., Clostridia OTUs which are genuinely amplified from the environmental template, not from contamination). A more refined method, then, is to identify only the specific OTUs associated with the negative control for a given run and remove solely those OTUs from the environmental dataset. In a sample set of our environmental data, this second method removed 4%±2% of the dataset's sequences (Fig. 2). In earlier publications (Mason et al. 2015; Case et al. 2015), the conservative "taxa-removal" method was applied, but in future studies I recommend the more precise "OTU-removal" method as it still accurately addresses contamination while retaining as much genuine environmental data as possible. From a practical standpoint, the two methods do not make a significant difference in studies from the methane seep environment. This is because the methane seep microbial ecosystem is dominantly populated by taxa (e.g., ANME archaea, Deltaproteobacteria, Epsilonproteobacteria) which are rarely observed in the negative controls. Therefore removing, for example, an entire clade observed in the negative controls has a limited impact on the methane seep environmental samples. However, in other environments negative control contaminant taxa may be closely related to organisms of experimental interest. For example, human microbiome samples are often rich in Bacteroidetes- and Firmicutes-related organisms (Faith et al. 2013; Rosenbaum et al. 2015). Therefore it is not reasonable to remove all OTUs associated with these taxa, even if some Bacteroidetes- and Firmicutes-associated OTUs are observed in the negative control. In such a case, an "OTU-removal" method would be best suited to address the experimental goals.

At this stage, it is also logical to remove OTUs from the dataset which are undesirable for other reasons besides their presence in the negative controls. For example, OTUs which are "unassigned" a taxonomy (often these are sequences which have high enough quality to pass joining and quality criteria but have zero matches in the reference database – and therefore cannot by definition be identified as chimeric) are not useful and are often removed. Also, depending on a study's objectives, certain clades may simply not be desired for the scientific question at hand. For example, if only Bacteria are of interest for the study, all OTUs associated

with Archaea could be intentionally removed at this step. It is also not uncommon for a few OTUs to be identified as Eukaryotic; these are generally removed at this point.

4.2.3 TRESHOLD FILTERING

In addition to removing OTUs linked to contamination, it is also necessary to consider spurious OTUs which may have been generated either during the preparation PCR reactions or during sequencing itself (Fig. 4). This was already partially addressed by the removal of singletons (OTUs which occur once, in one sample, in the entire dataset; Fig. 2). However, further applying a threshold cutoff, below which OTUs are summarily removed, can minimize the impact of spurious OTUs and increase confidence in the remaining very-low-abundance sequences (Bokulich et al. 2013). To address this, we take advantage of our plasmid mock communities which have known composition. We know that 12 plasmids were introduced; therefore, in a perfect dataset we would expect to only see 12 OTUs recovered in the sequencing data. However, even after removing singleton OTUs and subtracting contaminant taxa as determined by the relevant negative control, we still observe hundreds of OTUs present in the plasmid mock communities (Fig. 4). This observation is not improved by using either the 5-PRIME or NEB Q5 polymerase. Ultimately, this suggests that interpretation of very low abundance OTUs in iTag data may be fraught, since many OTUs appear to be spurious. Thus, it is desirable to choose a threshold relative abundance below which OTUs are removed from the dataset. Clearly a higher threshold will remove more OTUs, but at the cost of eliminating data which might represent genuine, low abundance 16S rRNA gene copies. In our mock communities, a threshold cutoff of 0.001 (0.1% relative abundance) appears to be appropriate; such a cutoff would come very close to limiting our sequencing data to the 12 OTUs expected to be observed (Fig. 4). However, in practice environmental samples uniformly host greater 16S rRNA gene diversity than mock communities; therefore, removing data less abundant than 0.1% may eliminate OTUs which

genuinely represent microbial diversity. Furthermore, in practice the threshold cutoff is often applied to the entire dataset; that is, if the threshold value is 0.001, then of all the sequences across all samples, the sequences in OTU_x must be present at least above 0.1% of the entire dataset. It is easy to conceive of a scenario in which OTU_x is genuinely highly abundant in $Sample_x$, but not in other samples in the dataset. Therefore, OTU_x might fail the entire dataset threshold and be removed from the sample set. In such a scenario, genuine and significant information will have been lost from the sequencing run, and the resulting sequencing composition of $Sample_x$ will not be reflective of its actual 16S rRNA gene profile. An obvious solution to this issue would be to apply cutoff thresholds on a per-sample basis, rather than a per-dataset basis. Because the cutoff threshold is often applied on a whole-dataset basis, a lower cutoff threshold minimizes the likelihood of accidentally removing important, but not well distributed, OTUs. In order to strike a balance between addressing spurious OTUs and retaining genuine biologic information, we have settled on 0.0001 (0.01% relative abundance; horizontal dashed lines in Fig. 4) as an appropriate threshold cutoff in our iTag data (16%±8% of sequences are removed with this cutoff; Fig. 2). This somewhat liberal cutoff value is itself more conservative than a previously recommended value of 0.0005 (0.005% relative abundance; Bokulich et al. 2013). Of course, this value may be varied by each user during bioinformatic processing, and depending on experimental details a value between 0.0001 and 0.001 may be appropriate.

4.3 iTag SEQUENCING PRECISION

By applying the above set of processing steps to our mock community data over multiple sequencing runs, we have been able over time to generate sequencing data of the same samples repeatedly. This enables comparison across runs to estimate precision of iTag sequencing as applied in the Orphan lab. These results do not represent the precision of any individual step (e.g.

PCR, or MiSeq sequencing), but are rather a cumulative result of the entire workflow from template amplification to sequencing to data processing.

Precision is overall quite good, with little variation between taxa, mock communities, or amplification enzyme across runs (Fig. 5). For plasmid mock communities amplified with the 5-PRIME or NEB Q5 enzyme, precision is 0.77% or 0.84%, respectively. For gDNA mock communities amplified with the 5-PRIME or NEB Q5 enzyme, precision is 1.85% or 0.88%, respectively. Since environmental samples are amplified from natural mixes of gDNA, rather than mixed plasmids, the gDNA mock communities are likely more relevant for determining precision of iTag sequencing data. Depending on the enzyme used to amplify environmental samples, then, a precision of between 1% and 2% is conservative and reasonable.

Within the gDNA mock communities, precision is worst for OTUs of Streptococcus spp. when amplified with the 5-PRIME enzyme. This is due to an oddity, that in the September 2015 iTag run, zero sequences were recovered in any of the four gDNA mock communities for Streptococcus spp. amplified with the 5-PRIME enzyme. Strangely, sequences of Streptococcus spp. were recovered at high and consistent relative abundance in all gDNA mock communities from all other iTag runs, and even in the September 2015 run of gDNA mock communities amplified with the NEB Q5 enzyme. The lack of Streptococcus spp. sequences remains perplexing, but appears to have been an isolated incident.

4.4 iTag SEQUENCING ACCURACY

In addition to the precision of iTag sequencing, it is important to consider the accuracy of sequencing results. All methods have inherent bias, including PCR which can yield variable results depending on the primers employed, the polymerase enzyme used, the annealing temperature, the gDNA template concentration, and other factors. It is obviously critical to employ, as much as possible, the exact some conditions within a given study. However, the effects

of PCR bias become especially concerning when attempting to compare results across multiple microbial ecology studies. Some community efforts have attempted to address this. For example, the Earth Microbiome Project (EMP) advocates for one set of universal 16S rRNA gene primers to be employed in all surveys of environmental microbiology (Gilbert et al. 2011). However, these results are hampered by several factors. First, the EMP primers are known to have certain taxonomic biases (Parada et al. 2015; Trembath-Reichert et al. 2016). Second, as next-generation sequencing technology rapidly develops, it becomes necessary (and beneficial) to develop new primers targeting larger segments of the 16S rRNA gene, thereby compromising efforts to apply consistent primers across datasets and across time (the EMP has partially addressed this by recommending changes to their primer sets over time). Third, the EMP primers are theoretically universal in their coverage but in some studies only specific microbial community members are of experimental interest (e.g., Archaea but not Bacteria). In that case, the user must decide whether to employ domain-specific primers (at the cost of comparability to other studies) or employ universal EMP primers (at the cost of throwing away a large proportion of the expensive dataset). Ultimately the Orphan lab, given the diverse range of environments studied (e.g. marine methane seeps and deep terrestrial boreholes) has moved forward with applying the EMP primers with the hope of generating comparable data across studies both within the lab and outside of the lab.

With our plasmid mock communities, whose composition is very well characterized, we are able to evaluate the accuracy of iTag sequencing on taxa which are particularly relevant in marine methane seep ecosystems (Fig. 6). We find some taxa are faithfully recovered in sequencing data regardless of polymerase enzyme employed (e.g., Desulfococcus; Fig. 6a-b), although the NEB Q5 enzyme better represents more taxa than 5-PRIME (e.g., ANME-1b, Desulfobulbus, Sulfurovum; Fig. 6a-b). Of particular note are the biases for and against certain taxa. Sequencing results from both enzymes indicate underrepresentation of some SEEP-SRB1 bacteria as well as ANME-2a, -2b, and -2c (Fig. 6a-b). The 5-PRIME samples are especially biased against ANME-2c, an important taxon on marine methane seep settings, yielding ANME-

2c sequences at ~25% of the expected relative abundance (or, a 4x bias). The NEB Q5 enzyme does moderately better at representing ANME-2c, although still underrepresents ANME-2c by half (a 2x bias). The biases against ANME-2 archaea have been explored in a recent publication, and do not appear to be related to primer mismatch (Trembath-Reichert et al. 2016). Correspondingly, plasmid mock communities amplified with the 5-PRIME and NEB Q5 enzymes also overrepresent some taxa (this is not surprising, given that relative abundance is a zero-sum metric – if some taxa are underrepresented, some must by definition be overrepresented). Within our plasmid mock communities, the ANME-1 archaea are consistently overrepresented by a factor of ~2 (Fig. 6a-b).

One robust method for countering bias, common in macrofaunal ecological literature (e.g., Levin et al. 2015) but only recently applied to microbial datasets, is to apply down-weighting transformations to the relative abundance data. A commonly applied function is to take the square root of the relative abundance data. Because the square root function is non-linear with increasing values, taxa at low relative abundance will be relatively unimpacted by the square root function while taxa at high relative abundance will be more greatly impacted. This acts to "smear" the relative abundance data, reducing bias effects. Indeed, when applying the square root transformation to our plasmid mock communities, the representation is overall improved as compared to non-transformed data (Fig. 6c-d). In fact, a transformation can be applied as strongly as desired. A 4^{th} root transformation of our plasmid mock community data results in even better representation of the data (Fig. 6e-f).

The most severe transformation would be to simply count the presence/absence of OTUs in the dataset, forgoing relative abundance completely. In such a transformation, all the data in our plasmid mock communities would be "perfectly" represented. However, such a severe transformation sacrifices genuine differences in relative abundance between microbial community members, potentially curtailing ecological interpretations which might otherwise contribute to a study. Ultimately the square root transformation is a good compromise between addressing bias

while retaining relative abundance information. Of course, it is not necessary to limit a dataset to one single analysis. If broad inter-sample trends are consistent across multiple transformations, it strengthens the conclusions to be able to note that the data is consistent across data treatments (Pasulka et al. 2015). Moreover, if the inter-sample trends do differ upon variable transformation of the data, it does not imply that any particular transformation is "right" or "wrong", *per se*. Although care must be taken with regard to biases, if the user is confident that the dataset is inherently unbiased, then transformations can be a powerful tool to examine the contribution of high-relative-abundance or low-relative-abundance community members to inter-sample trends. Untransformed data will naturally be dominated by the high-relative-abundance community members, and thus conclusions drawn from untransformed data will represent those members. If the scientific question at hand is with regard to the rare or low-relative-abundance members of a biological community, then a moderate or severe transformation will draw out the effect of those species upon data interpretation.

4.5 TESTING PREPARATION METHODS ON ENVIRONMETANL SAMPLES

The above sections of this chapter have focused on developing a data processing pipeline for iTag data, with an emphasis on what could be learned from negative controls and mock communities in terms of data quality, precision, and accuracy. Having used those samples to inform a robust, high-quality processing workflow, it is possible to now examine data from environmental samples (i.e., complex microbial communities) as a function of various methods tests applied during the initial phase of iTag preparation: PCR. Below, I summarize results from five tests:

1) Performing PCR in one step vs two steps.
2) Pooling singlet, duplicate, or triplicate PCR products.
3) Template concentration across three orders of magnitude.
4) 5-PRIME vs NEB Q5 enzyme amplification.
5) Annealing temperature for the NEB Q5 enzyme.

4.5.1 1-STEP vs 2-STEP PCR

The Orphan lab employs three deviations from the standard EMP preparation protocol. Firstly, the original EMP protocol calls for only one PCR reaction to be employed per sample. During this reaction, which is recommended for 35 PCR cycles, primers are employed which contain, in addition to the primer itself (19 bp-long for the 515f primer, 20 bp-long for the 806r primer), a 24- to 29 bp-long Illumina adapter, a 10 bp-long primer pad, a 2 bp-long primer linker, and, for the 806r primer, a 12 bp-long barcode sequence:

```
515f_EMP: AATGATACGGCGACCACCGAGATCTACAC              TATGGTAATT GT GTGCCAGCMGCCGCGGTAA
806r_EMP: CAAGCAGAAGACGGCATACGAGAT      XXXXXXXXXXXX AGTCAGTCAG CC GGACTACHVGGGTWTCTAAT
          |----------adapter----------| |-barcode--| |pad--linker| |------primer------|
```

Thus, the 515f primer is 60 bp in length and the 806r primer is 68 bp in length. Due to evidence that employing long primers over many cycles may lead to enhanced PCR bias (Berry et al. 2011), we decided to test a modification to the EMP protocol: rather than one amplification for 35 cycles with long primers, to instead perform first a 30 cycle amplification with only the raw primers ("PCR#1"):

```
515f: GTGCCAGCMGCCGCGGTAA
806r: GGACTACHVGGGTWTCTAAT
```

Followed by a second PCR for 5 cycles in which the full EMP primers are used in order to attach the pads, linkers, barcodes, and adapters ("PCR#2"). This test was performed on sediment #2687, from Eel River Basin, a marine methane seep offshore California. Besides this one protocol modification, the sample was treated identically in all other respects of the sample preparation and data processing workflow.

Results from this test indicated some differences in the recovered 16S rRNA gene profile (Fig. 7). Although the majority of OTUs were recovered in both the 1-step and 2-step treatments (the small number of OTUs not shared between the two treatments were very minor (<0.2% relative abundance) constituents of the gene profiles), the 2-Step PCR had decreased relative abundance of major OTUs and increased relative abundance of minor OTUs (Fig. 7b,d). Thus, a cross plot of OTU relative abundance between the two treatments resulted in a slope of just 0.69. It is hypothesized that 1-step PCR using long barcoded primers causes PCR bias because the overhanging barcode and/or adapter interact with template gDNA in a manner which varies based on template oligonucleotide sequences (Berry et al. 2011). Our data suggests that performing a 1-step PCR procedure decreases evenness across resulting sequence data as compared to a 2-step procedure. Our lab has moved forward with the 2-step approach, minimizing the chances for biased amplification of 16S rRNA genes from our environmental samples.

4.5.2 POOLING SINGLET, DUPLICATE, OR TRIPLICATE PCR PRODUCTS

The Orphan lab's second modification from the EMP protocol involves the pooling of replicated PCR products. The original EMP protocol calls for performing triplicate PCR reactions, which are all pooled before sequencing. Ostensibly, this is to buffer out the possible effect of PCR bias occurring in any single reaction series. However, performing triplicate PCR reactions is time consuming and costly, and so we tested the effect on sequencing data of performing just a single a PCR reaction, pooling duplicates, or pooling triplicate products. This test was performed on four marine methane seep samples representing a variety of environmental substrates: #2687 (sediment; the same sample on which we tested 1-step vs 2-step PCR), #5036 (a wood block colonized on the seafloor), #5193 (a carbonate), and #5472 (bottom water).

Results from all four samples were similar, indicating very little effect of performing and pooling one, two, or three PCR reactions (Fig. 8-11). In all cases the majority of OTUs, including all OTUs in all samples greater than 0.3% relative abundance, were recovered in all three treatments (Fig. 8-11, sub-plots a). Not only were the OTUs consistently recovered, they were represented in the sequencing data at very similar relative abundances regardless of pooling treatment. Cross plots of relative abundances therefore exhibited slopes very close to 1.0 with R^2 values also near 1.0 (Fig. 8-11, sub-plots c). Furthermore, there was no measurable effect on evenness; the major and minor OTUs were represented nearly equally across all pooling treatments (Fig. 8-11, sub-plot d). Not surprisingly, then, a non-metric multidimensional scaling analysis of the 12 sequencing profiles (4 samples x 3 pooling treatments each) revealed the samples to be well differentiated by substrate and not well differentiated by pooling treatment (Fig. 12). While these results suggest single PCR reactions would be sufficient to represent the data, our lab has taken the conservative approach of pooling duplicate PCR preparations for each sample; this saves time and cost from the originally recommended triplicate PCR reactions, while still minimizing the chances of spurious bias introduced in any single PCR reaction.

4.5.3 TEMPLATE CONCENTRATION ACROSS THREE ORDERS OF MAGNITUDE

As described above, PCR reactions can be compromised if gDNA template is low enough to be on par with the concentration of contaminant genetic material present in the PCR reactants or enzymes. In addition, it is possible for template concentration itself to affect the amplicon profile of a PCR reaction, likely due to an effect on the efficiency of each PCR cycle (Chandler et al. 1997). In order to test this affect, we applied our iTag pipeline to one sample (#5122, a carbonate) in which the gDNA template was undiluted ("1x"), diluted 10-fold ("10x"), and diluted 100-fold ("100x"). Besides these variations in template concentration, all other aspects of PCR preparation, sequencing, and data analysis were identical for the three samples.

It appeared from our data that template dilution had a measureable effect on the resulting sequence profiles (Fig. 13). Once again the majority of OTUs were shared across all treatments (Fig. 13a), but the relative abundances of the constituent OTUs was different depending on the dilution factor. The most dilute (100x) sample displayed the most even profile, with major OTUs exhibiting lower relative abundance and minor OTUs exhibiting higher relative abundance as compared to the 10x and 1x treatments (Fig. 13b,d). Despite these differences, a cross plot of relative abundance of all OTUs suggests only a moderately magnitude of effect on the sequence profile, with slope of 1.2 and R^2=0.92. Thus, while the effect of gDNA template concentration is measurable, it does not seem to impact the recovered sequence profile too adversely. While it is certainly preferable to perform PCR reactions with an equimolar amount of gDNA template in each reaction, it also appears that in cases where this is impracticable the resulting sequencing data may still be cross-compared between samples. An example of such a scenario includes when template concentrations are so low as to be immeasurable by even high-sensitivity fluorescence assays (Trembath-Reichert et al. 2016).

4.5.4 5-PRIME vs NEB Q5 enzyme amplification

The specific polymerase enzyme employed during a PCR reaction can have a dramatic effect on the resulting amplicon profile (Brandariz-Fontes et al. 2015). High-fidelity polymerase enzymes are less prone to accidentally synthesize an incorrect base during a PCR cycle, an error which can be propogated to high concentration over the course of many cycles in a PCR reaction. Errors introduced during PCR cycles affect the sequencing results and, depending on the extent of the error, may influence taxonomy assignments downstream.

The third and final deviation the Orphan lab has explored from the EMP protocol has been the choice of polymerase enzyme. The original protocol suggests using 5-PRIME HotMasterMix (catalog #2200410), but we have also explored use of the high-fidelity New

England Biolabs Q5 enzyme (catalog #M0491L). This enzyme is similar to the Finnzymes Phusion High Fidelity DNA Polymerase (S. Connon, personal communication from New England Biolabs), which has been independently shown to produce highly accurate PCR products in next-generation sequencing data (Brandariz-Fontes et al. 2015). In our tests, we employed the 5-PRIME HotMasterMix and NEB Q5 enzymes on five separate environmental samples in order to compare sequencing results (note, we also applied both enzymes to our plasmid mock communities and found the NEB Q5 enzyme to produce more accurate results – see above sections). These samples included four carbonates (#3622, #3624, #5104d, #5122) and one sediment (#5133).

As was the case with our plasmid mock communities, we observed a measurable difference in the resulting sequence profiles when amplified with 5-PRIME or NEB Q5 enzymes (Fig. 5,6,14). Although a large proportion of OTUs were reproducibly recovered in both treatments, the slopes and R^2 values of cross plots varied substantially from values of 1.0 (slope range: 0.78-1.31; R^2 range: 0.52-0.85). This suggests the choice of polymerase enzyme is important when preparing samples for next-generation sequencing; given that our plasmid mock communities were more accurately represented by the NEB Q5 enzyme, we made the decision to employ that enzyme moving forward from November 2014 (Orphan lab runs in November 2013 and May 2014 exclusively used the 5-PRIME enzyme and some continuing projects continue to employ the enzyme for continuity across datasets).

4.5.5 ANNEALING TEMPERATURE FOR THE NEB Q5 ENZYME

One of our observations from sequencing of plasmid mock communities was the bias against ANME-2 archaea which appeared to be improved by switching to the NEB Q5 polymerase enzyme. In order to further test if we could improve representation of ANME-2 archaea in the dataset, we applied the NEB Q5 enzyme at two different annealing temperatures

to an environmental sample (#5133, sediment) which was known to be rich in ANME-2 from previous studies (Trembath-Reichert et al. 2013). In a PCR reaction, annealing is the step at which the primer adheres to the template, after which extension (i.e., amplification) will occur. Only those genomic fragments will be amplified which successfully adhere with a primer. In general, specificity is a good quality at this step, as a user does not want PCR primers to incorrectly adhere to genomic fragments for which they were not intended. Specificity is improved by raising the temperature of the annealing reaction, requiring a better match between the primer and the template in order for successful adhesion (otherwise the primer will denature off the template). However, in the case that a user wishes to decrease specificity (e.g., to increase amplification of ANME-2, if primer adhesion were an issue), the annealing temperature can be lowered. This lowers the energetic requirements for adhesion, helping the primers adhere to genomic template that may not be a perfect nucleotide match. We tested the NEB Q5 enzyme on sample #5133 at annealing temperatures of 50˚C and 54˚C, in addition to 5-PRIME enzyme amplification at 50˚C. We note that the NEB Q5 enzyme is optimized to operate at 57-64˚C, but we employed lower temperatures in order to decrease specificity.

As hypothesized, the NEB Q5 enzyme at 50˚C did provide better representation of ANME-2 archaea than the NEB Q5 enzyme at 54˚C (and much better than the 5-PRIME enzyme; Fig. 15). However, overall the differences in NEB Q5 annealing at 50˚C or 54˚C were relatively small (Fig. 15a,d), and we chose to employ the NEB Q5 enzyme at an annealing temperature of 54˚C as it strikes a balance between the optimized temperature for the enzyme (57-64˚C) and a cooler temperature in order to increase adhesion with a wide range of genomic templates.

4.5.6 SUMMARY OF PCR TEST RESULTS

We observed that variations in the sample preparation procedure induced a range of differences in the resulting sequence profiles for environmental samples. Some variations

produced relatively large differences (e.g., 5-PRIME vs NEB Q5 polymerase enzyme) while others made little difference (e.g., pooling one, two, or three replicate PCR reactions). In order to visualize these differences at a broad level, we included all preparation tests in a non-metric multidimensional scaling ordination with a variety of other environmental samples from marine methane seep worldwide (Fig. 16). These results showed that although intra-sample differences were observed as a function of preparation method, the differences were uniformly small as compared to inter-sample difference in 16S rRNA gene profiles. Therefore, we are confident that the ecological interpretations made from iTag datasets are not compromised by small differences in sample preparation protocols. Of course, we recommend run-to-run consistency and following the best possible protocol at all times (e.g., more replicates is better than fewer, and high-fidelity polymerase enzymes are better than lower-fidelity ones), but we find these effects to be secondary to genuine microbial community differences as a function of environmental factors such as habitat substrate.

4.6 PRIMER ON QUANTITATIVE ECOLOGICAL TOOLS EMPLOYED IN DOWNSTREAM DATA INTERPRETATION

The field of ecology employs a number of mathematical and statistical tools in order to probe the relationships between biological communities and the environment. Here I will briefly describe the principle techniques used in this thesis, acknowledging that the full array of methods available to ecologists would (and does) encompass entire textbooks (e.g., Clarke and Warwick 2001; Legendre and Legendre 2012). When parsing ecological datasets, the investigator's goals are often divided into two aims: 1) to visualize inter-sample relationships by translating biological similarity into spatial distance, producing a graphic, or ordination, which enables hypothesis development, and 2) to apply statistical metrics to assess both whether apparent inter-sample or inter-group differences are robust and, if so, which biological community members contribute to

inter-sample or inter-group differences. In this thesis, the primary ordination method employed is Nonmetric Multi-Dimensional Scaling (NMDS). Inter-group differences are probed with the Analysis of Similarity (ANOSIM) statistical test, and biological community members contributing to inter-group differences are identified by the Similarity Percentage (SIMPER) routine.

NMDS is a more appropriate ordination technique for biological datasets, and in particular large next-generation sequencing datasets, than other familiar techniques such as Principal Components Analysis (PCA) or Principal Coordinates Analysis (PCoA). PCA, perhaps the most widely familiar ordination technique, plots sample data from *m* species onto *m* axes (aka, components), and then projects the resulting multidimensional field onto a lower dimensional graph. As a simple analogy, imagine holding up a ball-and-stick crystal lattice structure, then shining a flashlight on it and viewing the resulting projection against the wall. In this example, a 3-dimensional structure is collapsed to a 2-dimensional structure. In next-generation sequencing datasets, however, it is common to have hundreds or thousands of taxa represented. Genuine inter-sample relationships therefore exist in a multidimensional space with hundreds or thousands of axes, and collapsing the ordination down to two or three dimensions has the potential to drastically distort inter-sample relationships. Furthermore, PCA is highly sensitive to "joint absences" – zero values in a species abundance matrix. Imagine, for example, collecting species data from a forest, a desert, and a beach. Several dolphins are counted at the beach site, but zero dolphins are counted in the forest or in the desert. Intuitively, an ecologist would not claim that because the forest and the desert lack dolphins, they are more similar to one another. (Similarity is better defined as *joint presence* than *joint absence*. For example the forest and the desert are more similar because they both contain rabbits, not because they both lack dolphins.) PCoA is better at addressing joint absences than PCA, but suffers from the same issue in collapsing multidimensional ordination into lower dimensions by projection.

The algorithm supporting NDMS is fundamentally different than PCA or PCoA. First introduced in the 1960s in the field of Psychology (Shepard 1962; Kruskal 1964a; b), NMDS

avoids the constraint of attempting to project Euclidian distance onto lower dimensions (the crystal lattice and flashlight example above). Instead, NMDS is an iterative algorithm in which the number of dimensions is set *a priori* (usually 2 or 3) and the locations of samples in the ordination space are repeatedly adjusted in order to maximize the accuracy of the depicted inter-sample relationships (functionally, this involves minimizing a stress function that measures how accurately inter-sample relationships are represented). Importantly, because NMDS attempts to preserve the rank ordering of sample similarities, rather than Euclidian distance, the graphical depiction of inter-sample differences will never be perfect, but in practice the aggregate representation of the samples' data is often more accurate than in a PCA or PCoA plot. Furthermore, NMDS is not constrained by the problem of joint absences as was PCA. This is because the NDMS algorithm does not work on a raw sample-by-species table of abundance (or relative abundance), but rather on a triangular matrix of inter-sample similarity (this is the same workaround that PCoA uses to avoid the joint absence problem, although PCoA still attempts to project high-dimensional Euclidian distance onto lower dimensional space).

A common inter-sample similarity metric applied, and the one employed in this thesis, is Bray-Curtis similarity (S_{jk}), which is immune to joint absences and which results in intuitive values of 0 (if two samples share no species) and 100 (if two samples share all the same species at the same abundances or relative abundances; Bray and Curtis 1957; Oksanen et al. 2013). Bray-Curtis dissimilarity is also frequently employed in ecological approaches, and is simply equivalent to 100-S_{jk}, denoted as δ_{jk}:

$$S_{jk} = 100 \bullet (1 - (\Sigma^{p}_{i=1} \, |y_{ij} - y_{ik}|) \, / \, (\Sigma^{p}_{i=1} \, (y_{ij} + y_{ik}))) \qquad \text{Eq. 1}$$

$$\delta_{jk} = 100 \bullet ((\Sigma^{p}_{i=1} \, |y_{ij} - y_{ik}|) \, / \, (\Sigma^{p}_{i=1} \, (y_{ij} + y_{ik}))) \qquad \text{Eq. 2}$$

where p is the number of the species observed in the dataset and y is the relative abundance of species i in sample k or j. For a sample set with n samples, the number of sample-

sample combinations (i.e., the number of calculated S_{jk} values) will be equal to $n(n\text{-}1)/2$. As an example, take the synthetic relative abundance data described in Table 3. The Bray-Curtis similarities for the five samples are calculated and presented in Table 4. As expected from the raw relative abundance data, Samples 1 and 2 are highly similar. The other samples demonstrate a range of similarity values from 44 to 75. A 2-dimensional NMDS ordination would begin by placing five points on a plane (representing the five samples), and checking how close on the plane each sample is to each other sample. Sample 1 and 2 ought to be closer to each other than to any other samples, for example, and Sample 4 ought to be closest to samples 5, 2, 1, and 3 in that order. In other words, in a perfect ordination the distance between any two samples in low-dimensional space would increase monotonically with dissimilarity (defined as $100\text{-}S_{jk}$, or δ_{jk}). A plot of distance between two points in ordination space vs distance predicted from a monotonic regression line is a *Shepard Diagram* (Fig. 17B,D). The *stress* of an NMDS plot is defined as the extent to which ordination distance does not increase monotonically with $100\text{-}S_{jk}$:

$$\text{Stress} = \Sigma(d_{jk} - \hat{d}_{jk})^2 \,/\, \Sigma(d^2_{jk})` \qquad \text{Eq. 3}$$

where d is the actual distance between samples j and k on the ordination plot, and $\hat{d}$ is the predicted distance between samples j and k from the monotonic regression line (Kruskal 1964b; Clarke and Warwick 2001; Oksanen et al. 2013; Fig. 17). In the second iteration of the NMDS algorithm, the five data points will be relocated in order to minimize the stress value. Once the stress value is minimized to a pre-set threshold, or once the stress value no longer decreases, the ordination algorithm terminates and the resulting ordination is saved. NMDS, by virtue of handling Bray-Curtis similarity matrices rather than raw data (and therefore being immune to joint absences), and by applying a stress minimization function to achieve low-dimensional ordination rather than attempting to project high dimensional space onto lower dimensions, is considered a robust ordination technique for representing inter-sample relationships in biological

(including microbiological) datasets (Clarke and Warwick 2001; Ramette 2007). However, NMDS is not more than a visualization technique in order to generate hypotheses.

It is conceivable that once a study's samples are ordinated, some apparent trends will emerge. For example, in Fig. 17C it is apparent that the 16S rRNA gene profiles of carbonate and non-carbonate habitat substrates may differ. In order to test this hypothesis, the ANOSIM test is applied. This tests whether the inter-sample similarities are higher among samples of a defined group vs among samples of different groups. In this example, the test will be whether the 16S rRNA gene profiles between two samples of the same habitat substrate (sediment, nodule, bottom water, or carbonate) display, on average, higher Bray-Curtis similarity than the 16S rRNA gene profiles between two samples of differing habitat substrate (e.g., a sediment vs a carbonate). In order to discern this, the triangular Bray-Curtis similarity matrix is converted into a triangular rank similarity matrix; that is, every similarity is ranked with #1 being the two samples which have the highest Bray-Curtis similarity and #$n(n\text{-}1)/2$ being the two samples which have the lowest Bray-Curtis similarity. An example of this conversion is provided in Table 5. Then, the calculation of the ANOSIM test is straightforward. A metric, R, is calculated as described in Clarke and Warwick 2001 and Oksanen et al. 2013:

$$R = (r_{between} - r_{within}) \,/\, (n(n\text{-}1)/4) \qquad \text{Eq. 4}$$

Where r_{within} is the average rank similarity of all like-type sample-sample combinations (e.g., sediment-sediment or carbonate-carbonate) and $r_{between}$ is the average rank similarity of all sample-sample combinations which are of differing type (e.g., sediment-carbonate). As before, n is the number of samples in the study. If within-group similarity is higher than between-group similarity, an R value greater than 0 is calculated (R=1 would indicate every possible sample-sample pair within groups is more similar to every possible sample-sample pair between groups). By recalculating R many hundreds of times using randomly rearranged sample assignments to be

within or between groups, a significance value, *p*, can be calculated. This tests whether observed inter-group differences are likely to be encountered by chance. In the example from Fig. 17C, R and p values of 0.49 and <0.001, respectively, confirm that the 16S rRNA gene profiles are significantly distinguished by habitat substrate; this was explored further in Chapter Two of this thesis.

Once it has been determined that two (or more) sample groups differ in their biological communities, it is logical to probe which species contribute to inter-group differences. This problem can be computationally approached using the SIMPER routine (Clarke and Warwick 2001; Oksanen et al. 2013). The SIMPER algorithm is conceptually simple, but also susceptible to yielding misleading results in cases when standard deviations of species distributions are high. For this reason, SIMPER results must be interpreted with a critical eye. The safest use of SIMPER is to use the routine for identifying *possible* species contributing to inter-group differences, and then to always return to the raw data to confirm the trends. Depending on the hypotheses being tested, in some cases it is possible to discern the most important species in a dataset simply by examining the raw species-sample abundance table, and foregoing the SIMPER routine altogether.

SIMPER works by deconstructing Bray-Curtis dissimilarity (recall dissimilarity is defined as δ_{jk}, equal to $100\text{-}S_{jk}$) into contributions from each individual species. Recall from Eq. 2 that for each pair of samples *j* and *k*, each species *i* represents one term in the numerator summation. Therefore, in order to calculate each species' individual contribution to the Bray-Curtis dissimilarity between a sample-sample pair, the dissimilarity equation is applied without the numerator summation (Clarke and Warwick 2001; Oksanen et al. 2013):

$$\delta_{jk}(i) = 100 \bullet ((|y_{ij} - y_{ik}|) / (\Sigma^{p}_{i=1} (y_{ij} + y_{ik}))) \qquad \text{Eq. 5}$$

This calculation is applied to every species *i* for every pair of *j* and *k*, where *j* and *k* are samples between the two groups being tested (e.g., sediments *vs* carbonates). For each species, the resulting values across all sample pairs of $\delta_{jk}(i)$ are averaged, resulting in an average contribution of species *i* to inter-group differences in the dataset. The susceptibility of SIMPER to misleading results comes from the fact that over all possible combinations of *j* and *k*, there is sometimes very wide standard deviation among the calculated values of $\delta_{jk}(i)$. For this reason, in addition to an average contribution of each species *i* to inter-group dissimilarity, the standard deviation of the contribution is also always reported. This standard deviation is critically important for assessing the significance, or consistency, of species *i* contributing to inter-group differences. It is for this reason that I also strongly recommend using SIMPER only as a tool to identify *possible* species contributing to inter-group differences, and then always following up by double checking the raw species abundance (or relative abundance) data to confirm the inter-group differences are robust. Similar caution is recommended in Clarke and Warwick 2001.

A BRIEF SUMMARY OF HIGHLIGHTED QUANTITATIVE ECOLOGICAL TOOLS

An NMDS plot is a low dimension graph (in this thesis, always 2 dimensions) in which each point represents the entire microbial community profile (all species observed, including their relative abundances). Distances between points indicate sample-sample similarity, with closer points being having more similar biomarker profiles to one another. Since by nature an NMDS plot only attempts to preserve relative inter-sample differences (i.e., rank ordered differences), the units of the x and y axes (as well as orientation, rotation, and scaling) are arbitrary and therefore not generally reported. Lower stress values indicate better representation of the cumulate data, and stress values of <0.20 are generally considered sufficient for interpretation (<0.1 is ideal but often not achieved in large environmental datasets; Clarke and Warwick 2001).

The ANOSIM test evaluates whether groups of samples are statistically distinct from other groups of samples. Furthermore, the R value produced from the ANOSIM test is a measure of how strongly the defined groups differentiate the samples being interrogated. In practice R values >0.60 are rarely computed in diverse environmental 16S rRNA gene datasets presented in this thesis, but statistically significant R values in the range of 0.30 to 0.60 are not uncommon.

The SIMPER test identified species that contribute to differences in biological communities between sample groups. The SIMPER algorithm parses the Bray-Curtis similarity calculation to assess the contribution of each species, but is susceptible to wide standard deviations due to the large number of possible sample combinations. SIMPER should be carefully evaluated and always double checked against the raw abundance (or relative abundance) data.

Many more details of these tests, including variable implementations, can be found in numerous literature resources (Clarke and Warwick 2001; Legendre and Legendre 2012). This section is intended to briefly familiarize the reader with some of the types of ecological techniques which were frequently employed in this thesis, especially in Chapters One through Three. With the exception of Chapter One, all ecological calculations in this thesis were performed with the 'vegan' (v2.0.10) package of the R environment (Oksanen et al. 2013; R Core Team 2014), with frequent conceptual clarity gleaned from the methods manual produced in conjunction with the Primer-E software package (Clarke and Warwick 2001). Analyses in Chapter One were performed with the Primer-E software package (Clarke and Warwick 2001).

Table 1. Relative abundances of taxa included in the plasmid mock communities. Mock communities were generated by mixing quantified amounts of 16S rRNA plasmids clone libraries in lab. These mock communities were used to assess both precision and accuracy of iTag sequencing.

Phylogeny	Comm. 1	Comm. 2	Comm. 3	Comm. 4
Archaea/Euryarchaeota/Methanomicrobia/ANME-1/ANME-1a	0.01	0.04	0.05	0.20
Archaea/Euryarchaeota/Methanomicrobia/ANME-1/ANME-1b	0.01	0.04	0.05	0.20
Archaea/Euryarchaeota/Methanomicrobia/Methanosarcinales/ANME-2a-2b/ANME-2b	0.04	0.01	0.15	0.05
Archaea/Euryarchaeota/Methanomicrobia/Methanosarcinales/ANME-2c	0.04	0.01	0.15	0.05
Archaea/Euryarchaeota/Methanomicrobia/Methanosarcinales/ANME-2a-2b/ANME-2a	0.03	0.01	0.15	0.05
Bacteria/Proteobacteria/Deltaproteobacteria/Desulfobacterales/Desulfobulbaceae/Desulfobulbus	0.03	0.03	0.11	0.11
Bacteria/Proteobacteria/Epsilonproteobacteria/Campylobacterales/Helicobacteraceae/Sulfurovum	0.25	0.25	0.01	0.01
Bacteria/Proteobacteria/Deltaproteobacteria/Desulfobacterales/Desulfobacteraceae/Desulfococcus	0.03	0.03	0.11	0.11
Bacteria/Proteobacteria/Deltaproteobacteria/Desulfobacterales/Desulfobacteraceae/SEEP-SRB1	0.03	0.03	0.11	0.10
Archaea/Thaumarchaeota/Miscellaneous Crenarchaeotic Group	0.10	0.10	0.01	0.01
Archaea/Euryarchaeota/Thermoplasmata/Thermoplasmatales/Marine Benthic Group D and DHVEG-1	0.42	0.42	0.01	0.01
Bacteria/Proteobacteria/Deltaproteobacteria/Desulfobacterales/Desulfobacteraceae/SEEP-SRB1	0.03	0.03	0.09	0.10

Table 2. Relative abundances of taxa included in the genomic mock communities. Mock communities were generated by mixing known amounts of genomic extract (gDNA) from cultures grown in lab. Compositions are corrected to include copy number of the 16S rRNA gene for each organism.

Phylogeny	Comm. 1	Comm. 2	Comm. 3	Comm. 4
Bacteria/Firmicutes/Clostridia/Clostridiales/Eubacteriaceae/Acetobacterium/Woodii	0.18	0.13	0.08	0.02
Bacteria/Bacteroidetes/Bacteroidia/Bacteroidales/Bacteroidaceae/Bacteroides/Coprosuis	0.11	0.08	0.05	0.03
Bacteria/Firmicutes/Bacilli/Lactobacillales/Streptococcaceae/Streptococcus/spp.	0.39	0.29	0.17	0.05
Bacteria/Proteobacteria/Deltaproteobacteria/Desulfobacterales/Desulfobacteraceae/Desulfococcus/Multivorans	0.04	0.05	0.04	0.02
Bacteria/Proteobacteria/Deltaproteobacteria/Desulfobacterales/Desulfobulbaceae/Desulfobulbus/Propionicus	0.07	0.14	0.22	0.13
Archaea/Euryarchaeota/Methanomicrobia/Methanosarcinales/Methanosarcinaceae/Methanosarcina/Acetivorans	0.11	0.21	0.39	0.61
Archaea/Euryarchaeota/Methanomicrobia/Methanosarcinales/Methanosarcinaceae/Methanolobus/Zinderi	0.04	0.03	0.02	0.11
Archaea/Euryarchaeota/Thermoplasmata/Thermoplasmatales/Thermoplasmataceae/Thermoplasma/Acidophilum	0.04	0.03	0.02	0.01
Bacteria/Proteobacteria/Gammaproteobacteria/Methylococcales/Methylococcaceae/Methyloprofundus/Sedimenti	0.04	0.03	0.02	0.02

Table 3. Synthetic relative abundance data of five species recovered from five samples.

	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5
Species A	0.23	0.20	0.62	0.06	0.31
Species B	0.14	0.11	0.22	0.43	0.29
Species C	0.34	0.30	0.01	0.19	0.16
Species D	0.20	0.24	0.02	0.13	0.12
Species E	0.09	0.15	0.13	0.19	0.12

Table 4. Bray-Curtis similarity values (S_{jk}) as calculated from Eq. 1 for samples 1 through 5 in Table 3.

	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5
Sample 1					
Sample 2	90				
Sample 3	49	47			
Sample 4	61	64	44		
Sample 5	74	71	68	75	

Table 5. Ranked similarities as determined from Table 4.

	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5
Sample 1					
Sample 2	#1				
Sample 3	#8	#9			
Sample 4	#7	#6	#10		
Sample 5	#3	#4	#5	#2	

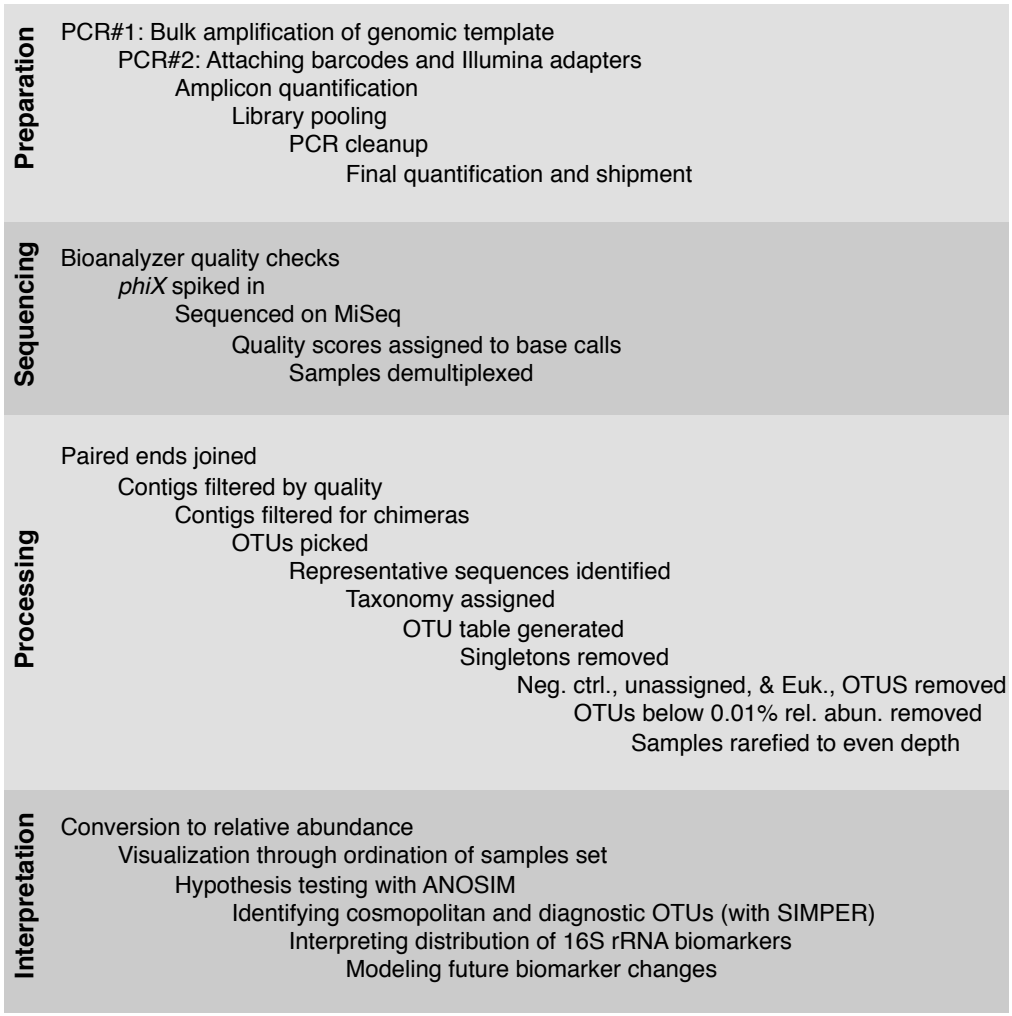


Figure 1. Flow chart of procedures from amplification of bulk genomic template to interpretation of Illumina sequencing data. The workflow is divided into four categories. Preparation, Processing, and Interpretation and performed at Caltech; Sequencing is performed at Laragen, Inc. At every step of the workflow, options are variations are available and sometimes appropriate; each step requires advanced knowledge by the user in order to make informed decisions about how to prepare, process, and interpret the data.

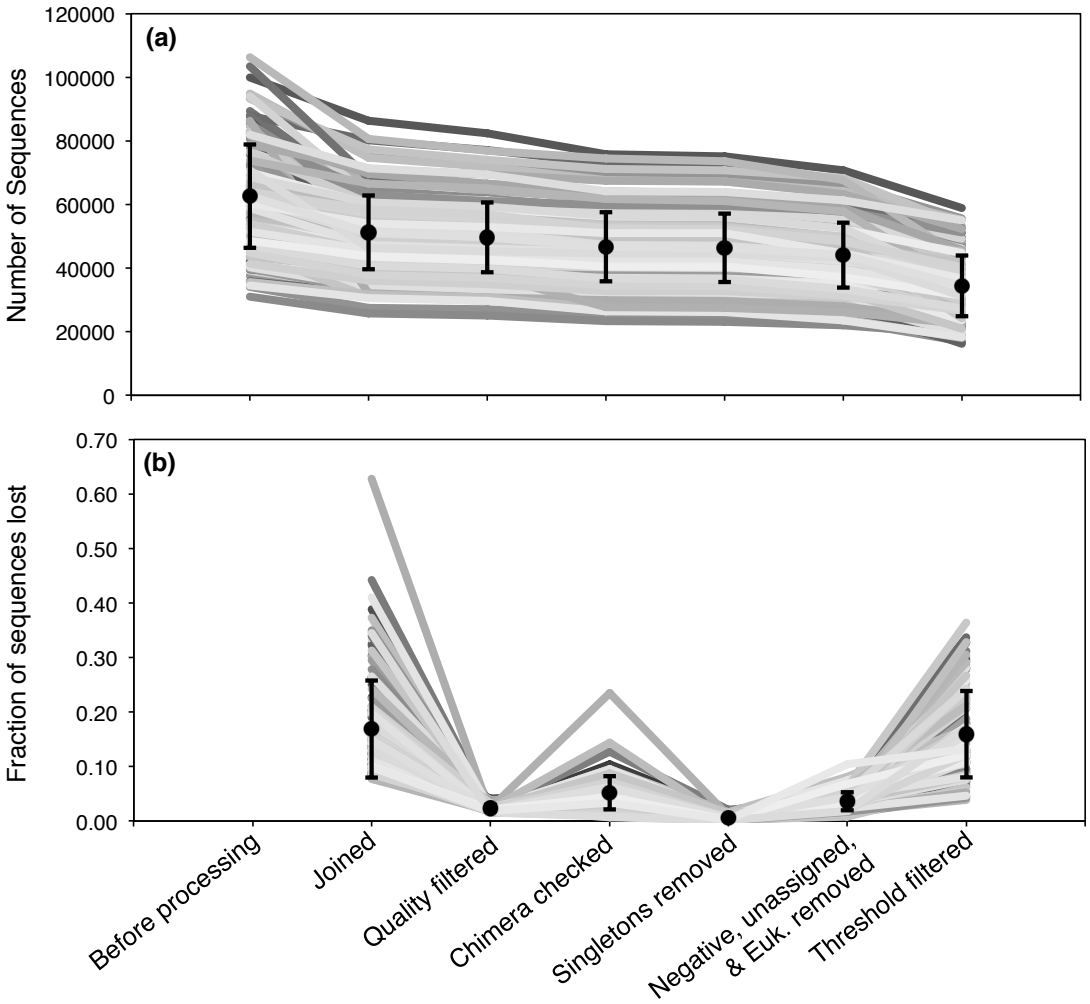


Figure 2. Detailed examination of the loss of sequences at each step of data processing. Panel (a) gives the raw number of sequences at each step, i.e. the value given for "Joined" is the number of remaining sequences after joining. Panel (b) gives the fraction of initial sequences which are lost at each processing step. In both panels, average values for each step are denoted with a black circle (plus/minus one standard deviation). Most sequences are lost at the joining step or at the threshold filtering step. Trimming, singleton removal, and filtering of negative control, unassigned, or eukaryotic sequences have a relatively minor effect on the dataset.

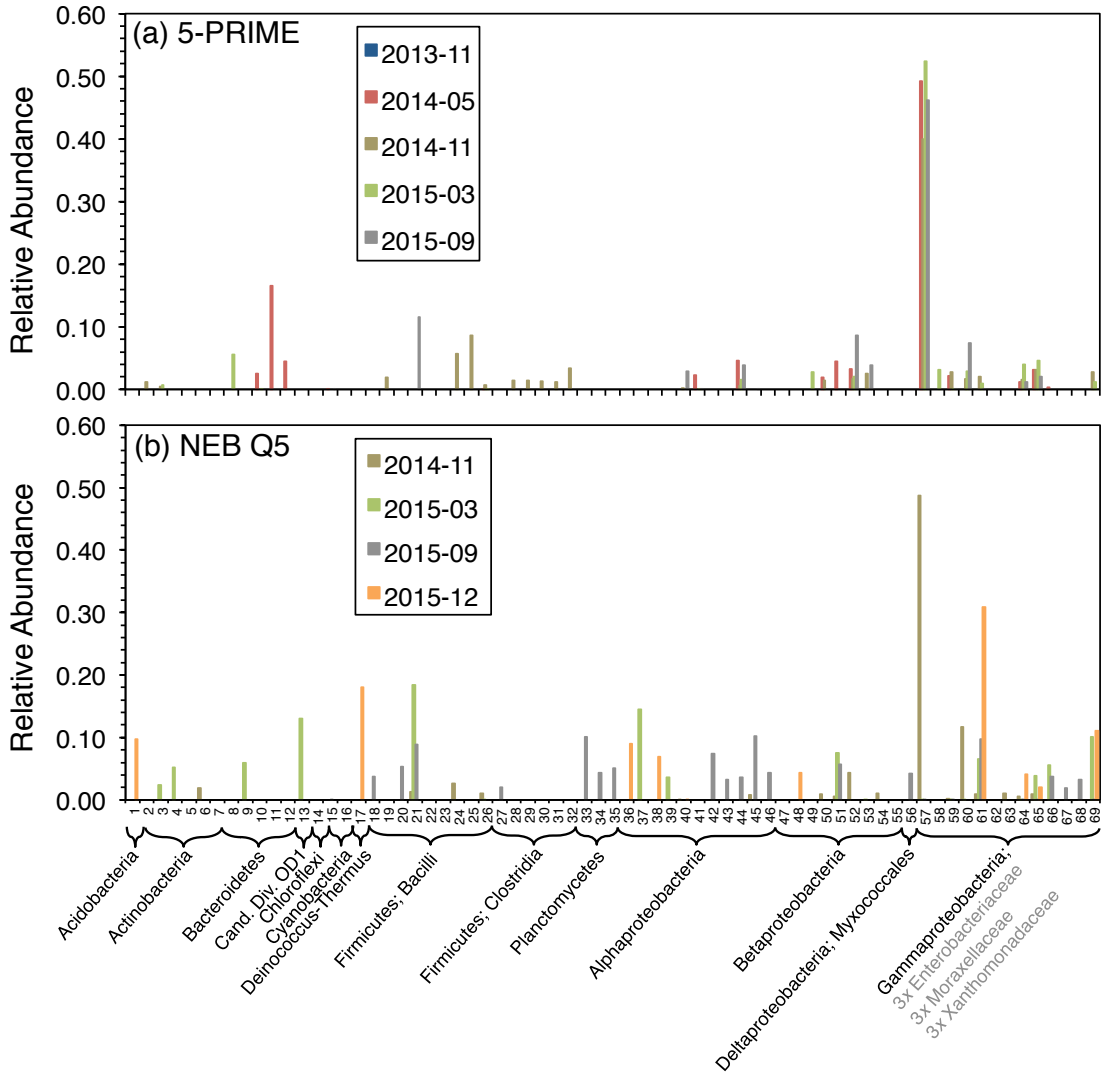


Figure 3. Relative abundance of taxa in negative controls. (a) Negative controls amplified with the 5-PRIME Hot Master Mix *Taq* (Item# 2200410). (b) Negative controls amplified with the Q5 Hot Start High-Fidelity 2X Master Mix (Item# M0494s). Negative controls are included in each sequencing batch (here sequencing batches are identified in the legends). In samples amplified with the 5-PRIME enzyme, an OTU identified generally as *Gammaprotoebacteria* is highly abundant in all negative controls. Other OTUs are also present at up to ~20% relative abundance, representing a range of taxonomies. These OTUs and their associated taxonomies vary within the negative control from each sequencing batch, emphasizing the importance of including negative controls in iTag sequencing batches. Variability is even greater in the Q5-amplified negative controls, with no particular OTU or taxa being consistently observed across all negative controls.

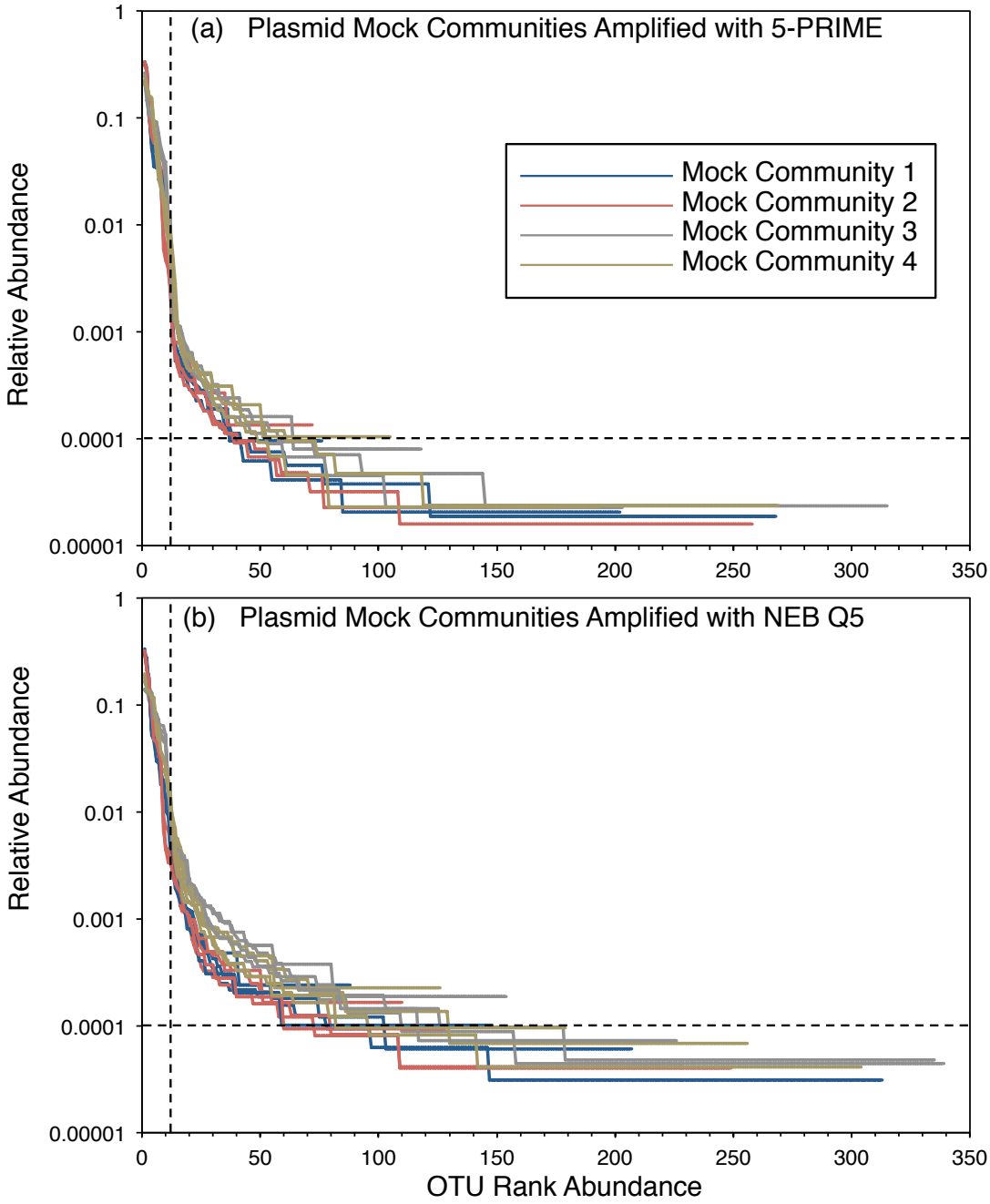


Figure 4. Relative abundance of OTUs detected in the plasmid mock communities, ordered by rank abundance of OTUs. Legend (a) applies also to (b). In both panels, the vertical dashed line indicates the expected number of OTUs, 12, which was the number of plasmids mixed into the mock communities. In both panels, the horizontal dashed line denotes a relative abundance of 0.0001 (0.01%), the threshold cutoff value applied to environmental samples. Although a threshold cutoff of 0.001 (0.1%) would have been more conservative, in practice this removed an undesirable fraction of the total dataset: environmental samples host richer OTU diversity than mock communities by several orders of magnitude, making us hesitant to remove too many OTUs which might be genuine signals from environmental samples. Furthermore, our environmental analyses exclusively focus on OTUs which are relatively highly abundant in the datasets (generally >0.1%), so our conclusions regarding microbial distribution and ecology are rarely dependent on the veracity of OTUs present at only ~0.01% (i.e., our conclusions are not impacted by the retention of OTUs present at 0.1%-0.01% relative abundance). Finally, when the threshold cutoff is applied to environmental, it is applied simultaneously to the *whole dataset*. That is, among *all sequences* from all samples in the dataset, an OTU must be present at 0.01% or greater relative abundance among *all sequences* in order to be retained. This could be achieved by a single sample genuinely hosting very many sequences of an OTU, while all other samples lack the OTU. In that case, the OTU ought to be retained since it represents real microbial presence. Setting a threshold cutoff at 0.01% relative abundance therefore minimizes our chances of accidentally removing real and relevant OTUs.

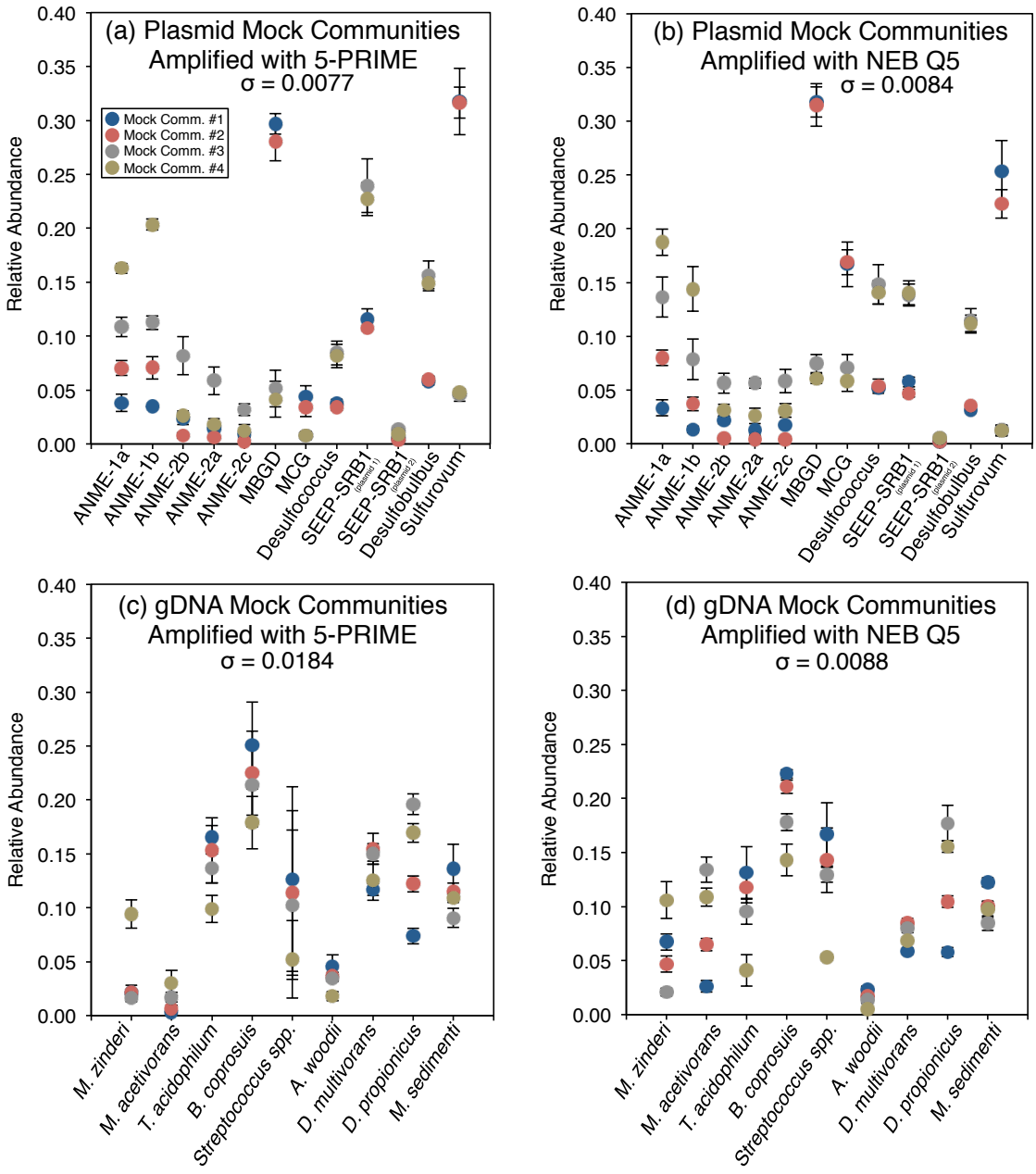


Figure 5. Reproducibility (precision) of iTag sequencing of mock communities. Data of plasmid mock communities is given in (a) and (b). Data of gDNA mock communities is given in (c) and (d). Reproducibility is better for the plasmid mock communities than the gDNA mock communities, possibly due to better-prepared template. For plasmid mock communities, 1-sigma precision is 0.77% (5-PRIME) and 0.84% (NEB Q5). For gDNA mock communities, 1-sigma precision is 1.85% (5-PRIME) and 0.88% (NEB Q5). Thus, it appears that iTag sequencing is precise to ~1-2%, depending on specific of sample preparation such as amplification enzyme. Legend in (a) applies to all panels.

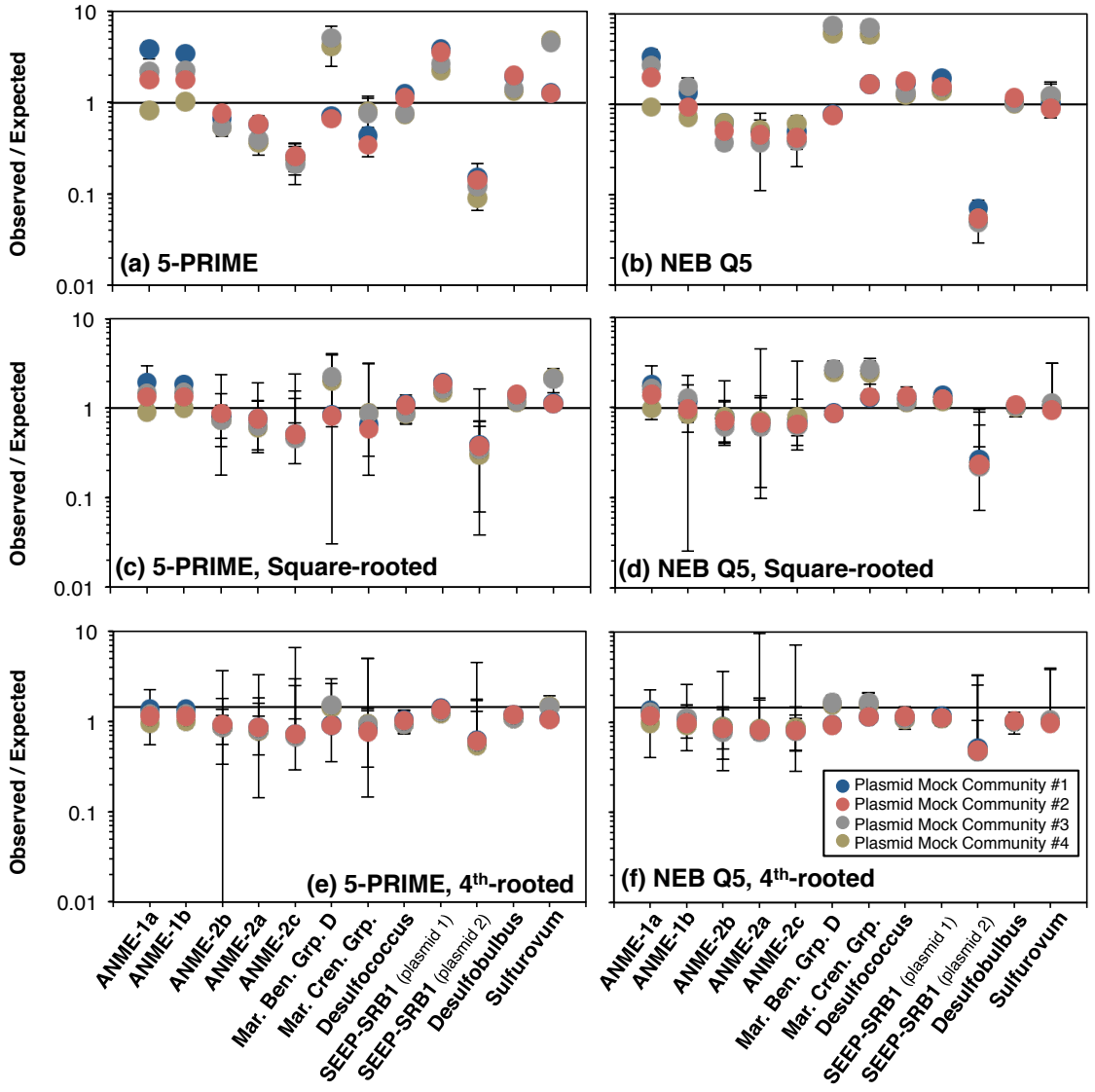


Figure 6. Accuracy of iTag sequencing of plasmid mock communities. In the raw relative abundance data (a, b), the accuracy of representation of OTUs varies widely. Some OTUs are well represented, especially after amplification with NEB Q5 (c.f. Desulfobulbus and Sulfurovum in (b)), while others are overrepresented by as much as ~8*x* (c.f. MBGD and MCG in (b)) or underrepresented by as much as ~10*x* (c.f. one SEEP-SRB1 plamid in (a) and (b)). Notably, the ANME-1 archaea are slightly overrepresented by a factor of ~2-3*x* while the ANME-2 archaea are generally underrepresented by a factor of ~2-3*x*. This is faiirly consistent between 5-PRIME and NEB Q5, with the exception that ANME-1b and ANME-2c are markedly better represented in NEB Q5 data than 5-PRIME data. In general, all OTUs are better represented when a square-root normalization is applied to the relative abundance data. The reason for this is that the square-root function mitigates PCR bias by preferentially down-weighting the OTUs which appear at high relative abundance compared to those at relative abundance. A more severe correction, such as a 4th root normalization (e,f), further mitigates PCR bias, but at the cost of lost information regarding which OTUs are genuinely more or less abundant in the dataset (the most severe transformation possible is toexamine only presence/absence). Ultimately, the square-root transformation is a good compromise between addressing PCR bias while still retaining valuable information about the relative abundance of OTUs.

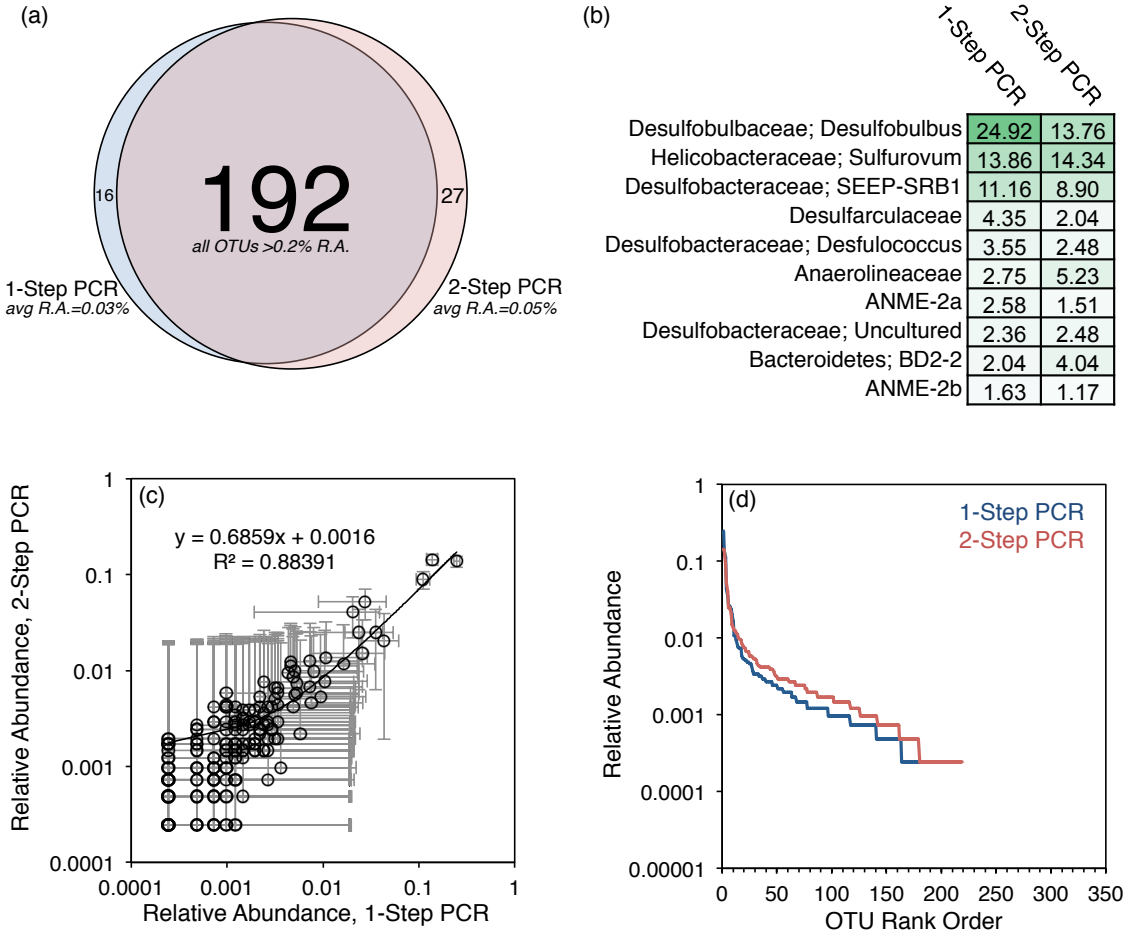


Figure 7. Comparison of amplification approach: a single PCR reaction employing long primers containing Illumina adapters and barcodes ("1-Step PCR") vs two PCR reactions, in which the first employs short primers and the second attaches the adapters and barcodes ("2-Step PCR"). This test was applied to a single methane seep sediment sample, #2687, in the November 2013 Illumina run. (a) shows the OTU overlap between the two samples. The majority of OTUs are shared, including all the major OTUs (any OTU >0.2% relative abundance in either sample is shared between the two samples). Only a small number of OTUs, with very low relative abundance, are unique to either the 1-Step or 2-Step PCR sample. (b) tabulates the relative abundances of the top 10 most abundant OTUs between the two samples, while (c) is a cross-plot of the relative abundance of the 192 OTUs shared between the two samples. Error bars in (c) are 1.85% relative abundance (c.f. precision for samples amplified with 5-PRIME *Taq*). (d) gives the OTU rank abundance curve for both samples. Overall the curves are quite similar. The fact that low-abundance OTUs (rank: ~25-250) are at slightly higher relative abundance in the 2-Step PCR suggests that rare community members are better represented by applying a 2-Step PCR preparation protocol than a 1-Step PCR preparation protocol.

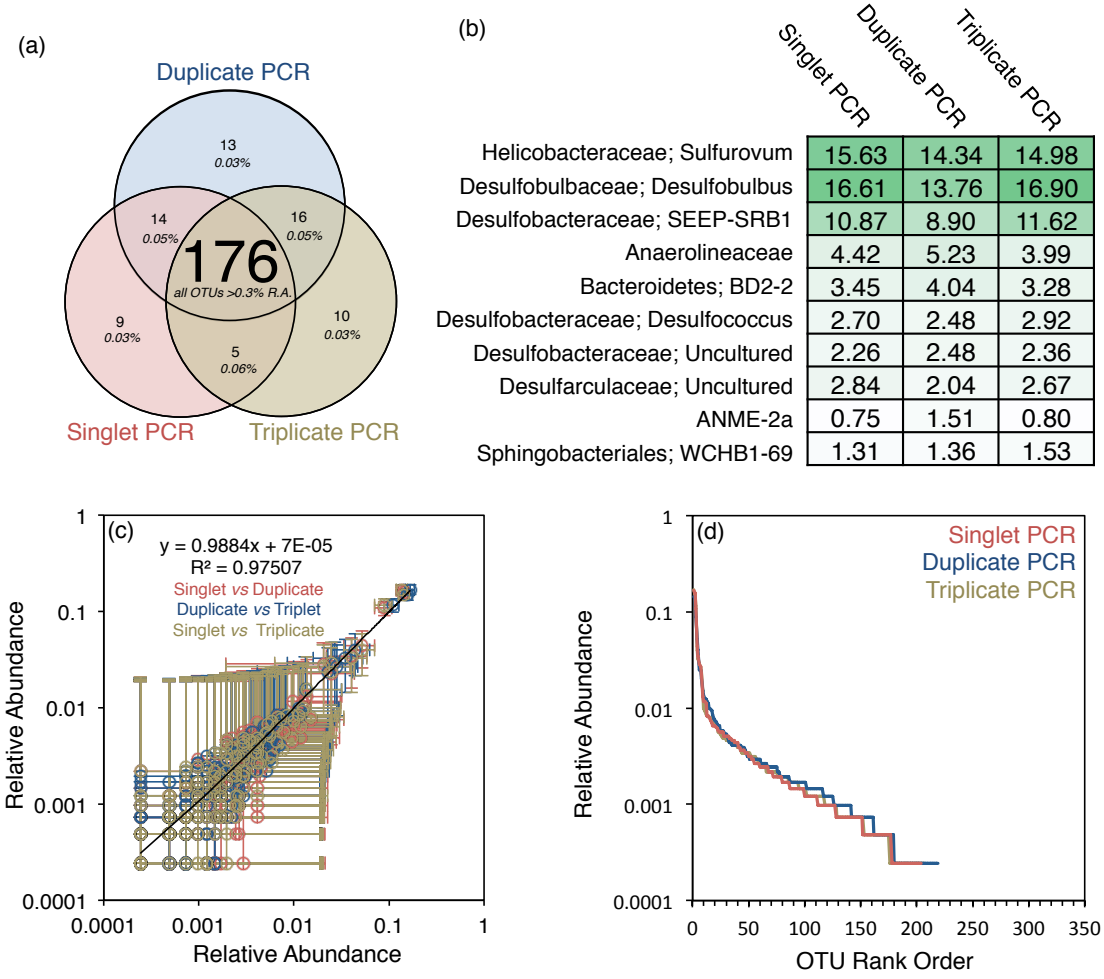


Figure 8. Comparison of amplification approach: pooling single, double, or triple PCR products during preparation of sample #2687 (sediment) for iTag sequencing. (a) shows the OTU overlap between results from the three preparations. The majority of OTUs are shared, including all the major OTUs (any OTU >0.3% relative abundance in either sample is shared between the two samples). Only a small number of OTUs, with very low relative abundance, are not shared between preparations. (b) tabulates the relative abundances of the top 10 most abundant OTUs between the preparations, while (c) is a cross-plot of the relative abundance of the OTUs shared between the three preparations. Error bars in (c) are 1.85% relative abundance (c.f. precision for samples amplified with 5-PRIME *Taq*). (d) gives the OTU rank abundance curve for all three preparations.

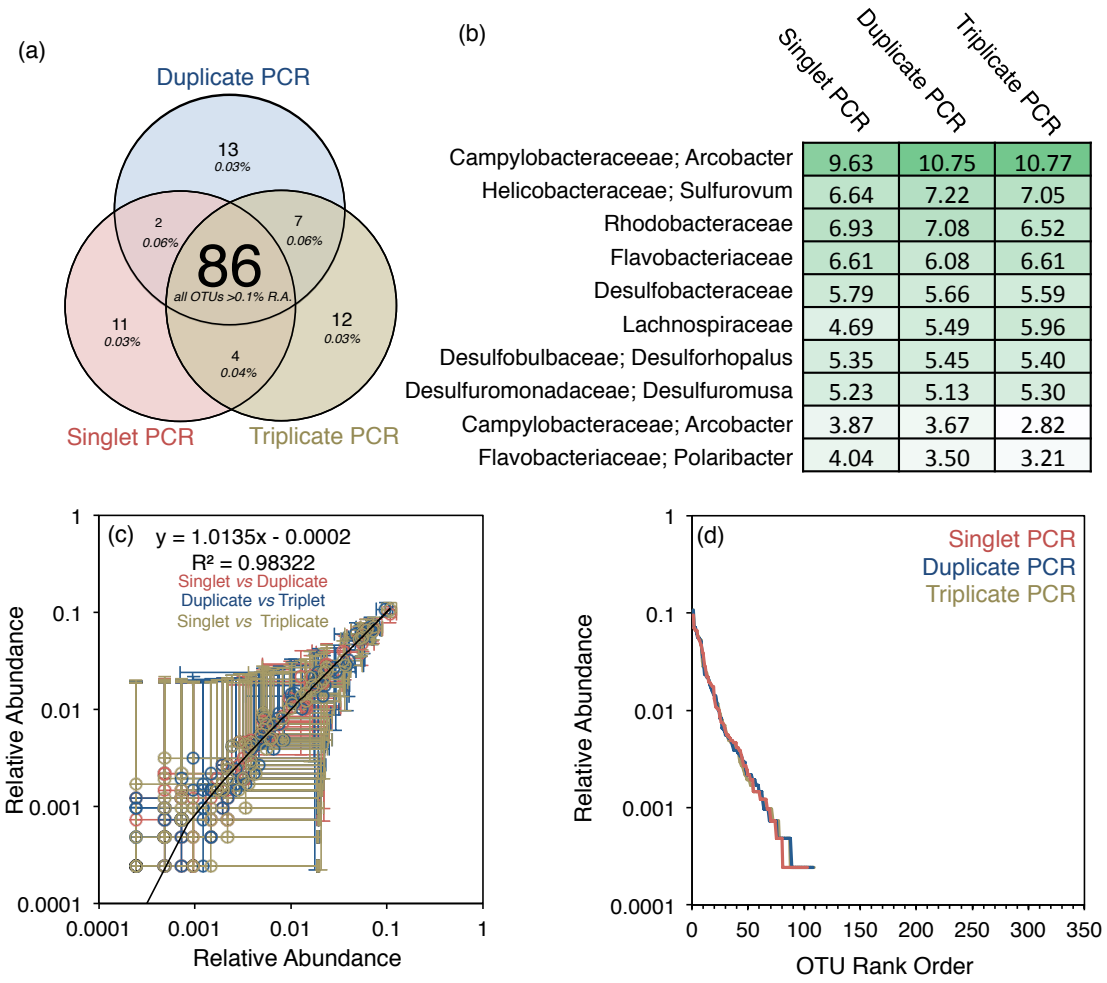


Figure 9. Comparison of amplification approach: pooling single, double, or triple PCR products during preparation of sample #5036 (colonized wood) for iTag sequencing. (a) shows the OTU overlap between results from the three preparations. The majority of OTUs are shared, including all the major OTUs (any OTU >0.1% relative abundance in either sample is shared between the two samples). Only a small number of OTUs, with very low relative abundance, are not shared between preparations. (b) tabulates the relative abundances of the top 10 most abundant OTUs between the preparations, while (c) is a cross-plot of the relative abundance of the OTUs shared between the three preparations. Error bars in (c) are 1.85% relative abundance (c.f. precision for samples amplified with 5-PRIME *Taq*). (d) gives the OTU rank abundance curve for all three preparations.

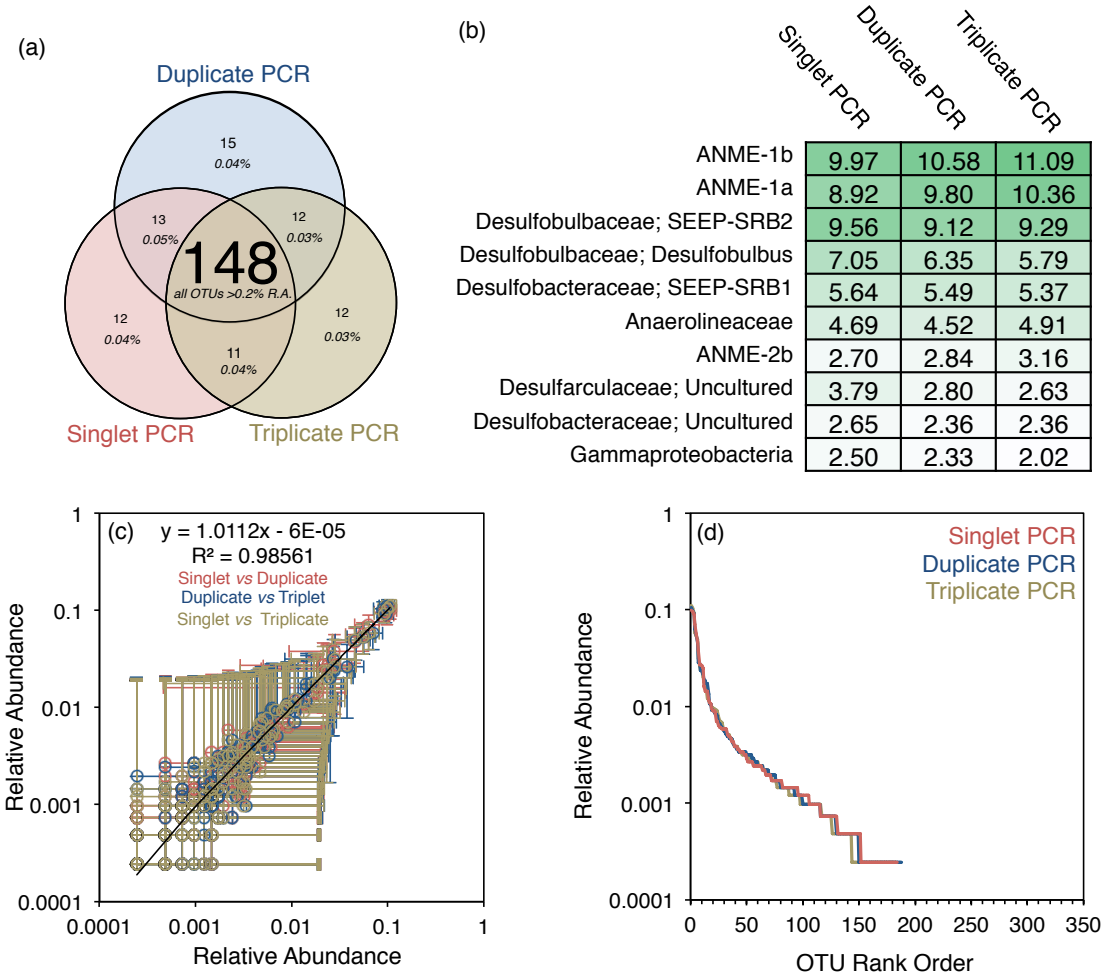


Figure 10. Comparison of amplification approach: pooling single, double, or triple PCR products during preparation of sample #5193 (transplanted carbonate) for iTag sequencing. (a) shows the OTU overlap between results from the three preparations. The majority of OTUs are shared, including all the major OTUs (any OTU >0.2% relative abundance in either sample is shared between the two samples). Only a small number of OTUs, with very low relative abundance, are not shared between preparations. (b) tabulates the relative abundances of the top 10 most abundant OTUs between the preparations, while (c) is a cross-plot of the relative abundance of the OTUs shared between the three preparations. Error bars in (c) are 1.85% relative abundance (c.f. precision for samples amplified with 5-PRIME *Taq*). (d) gives the OTU rank abundance curve for all three preparations.

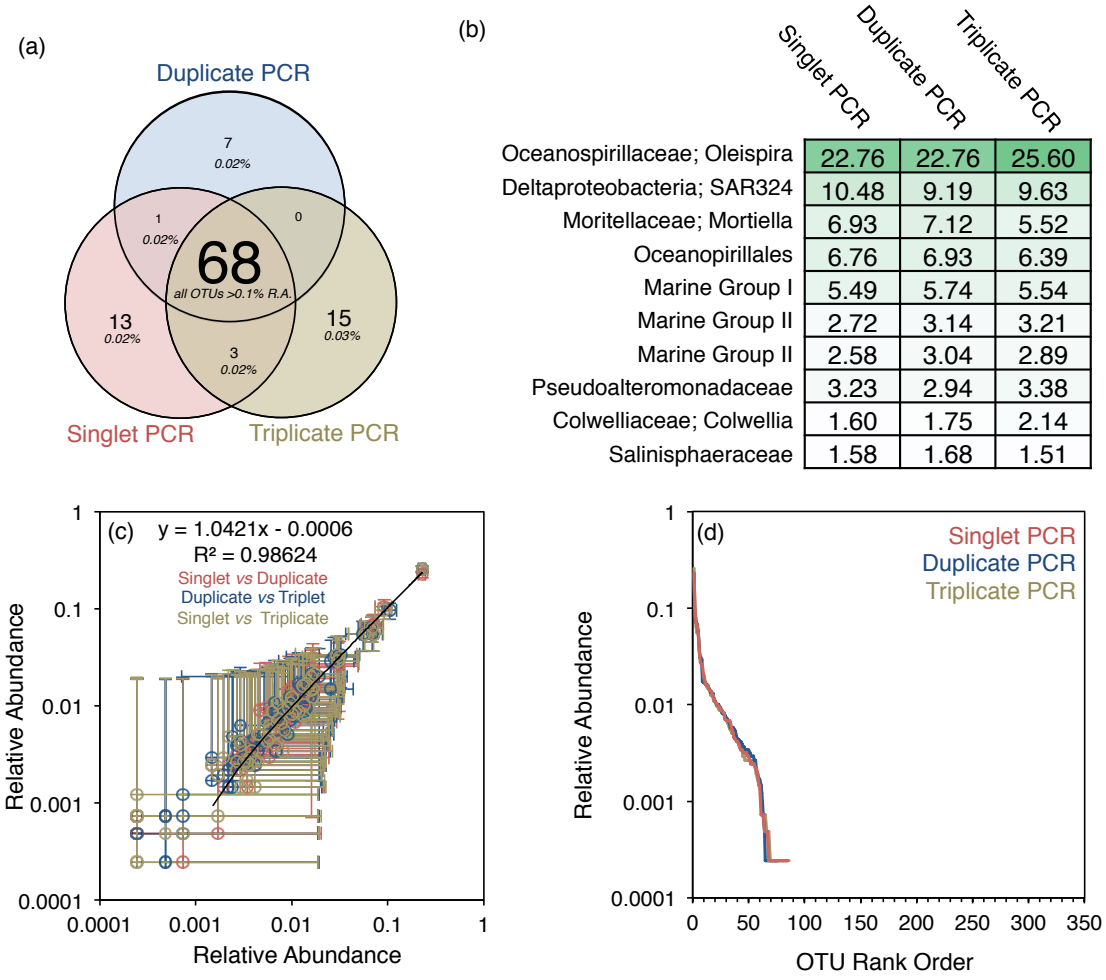


Figure 11. Comparison of amplification approach: pooling single, double, or triple PCR products during preparation of sample #5472 (bottom water) for iTag sequencing. (a) shows the OTU overlap between results from the three preparations. The majority of OTUs are shared, including all the major OTUs (any OTU >0.1% relative abundance in either sample is shared between the two samples). Only a small number of OTUs, with very low relative abundance, are not shared between preparations. (b) tabulates the relative abundances of the top 10 most abundant OTUs between the preparations, while (c) is a cross-plot of the relative abundance of the OTUs shared between the three preparations. Error bars in (c) are 1.85% relative abundance (c.f. precision for samples amplified with 5-PRIME *Taq*). (d) gives the OTU rank abundance curve for all three preparations.

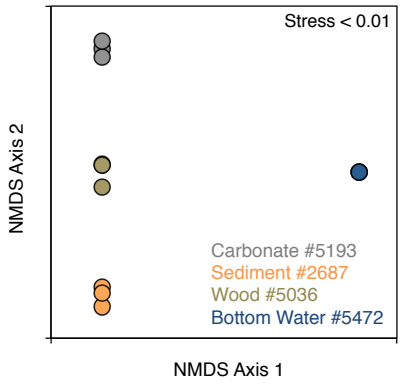


Figure 12. Non-metric multidimensional scaling analysis of PCR pooling treatments (single, double, triple) of four marine methane seep samples representing various substrates. The samples are well-differentiated by substrate but not by pooling treatment, indicating pooling treatment is not likely to alter ecological interpretations of datasets.

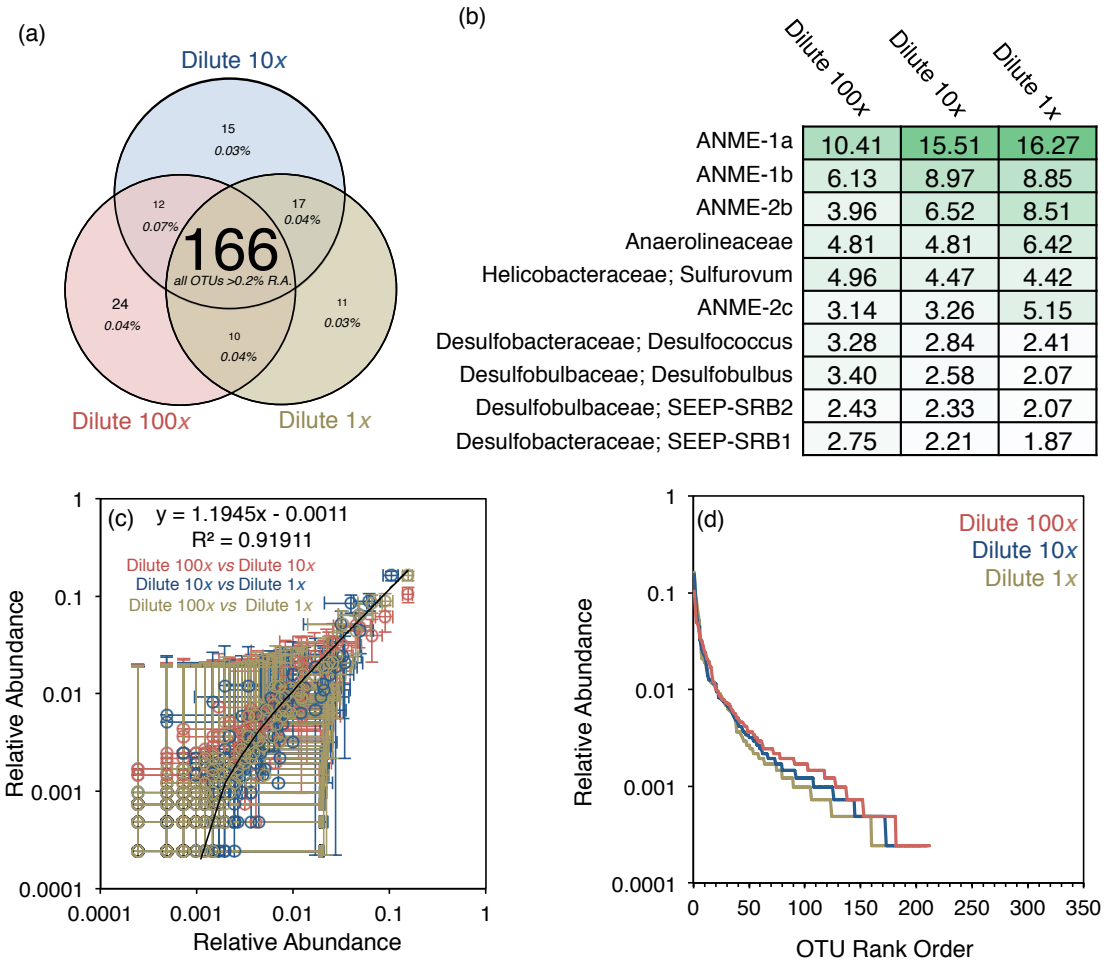


Figure 13. Comparison of amplification approach: gDNA template diluted 1X, 10X, or 100X prior to PCR amplification of sample #5122 (carbonate) for iTag sequencing. (a) shows the OTU overlap between results from the three preparations. The majority of OTUs are shared, including all the major OTUs (any OTU >0.2% relative abundance in either sample is shared between the two samples). Only a small number of OTUs, with very low relative abundance, are not shared between preparations. (b) tabulates the relative abundances of the top 10 most abundant OTUs between the preparations, while (c) is a cross-plot of the relative abundance of the OTUs shared between the three preparations. Error bars in (c) are 1.85% relative abundance (c.f. precision for samples amplified with 5-PRIME *Taq*). (d) gives the OTU rank abundance curve for all three preparations.

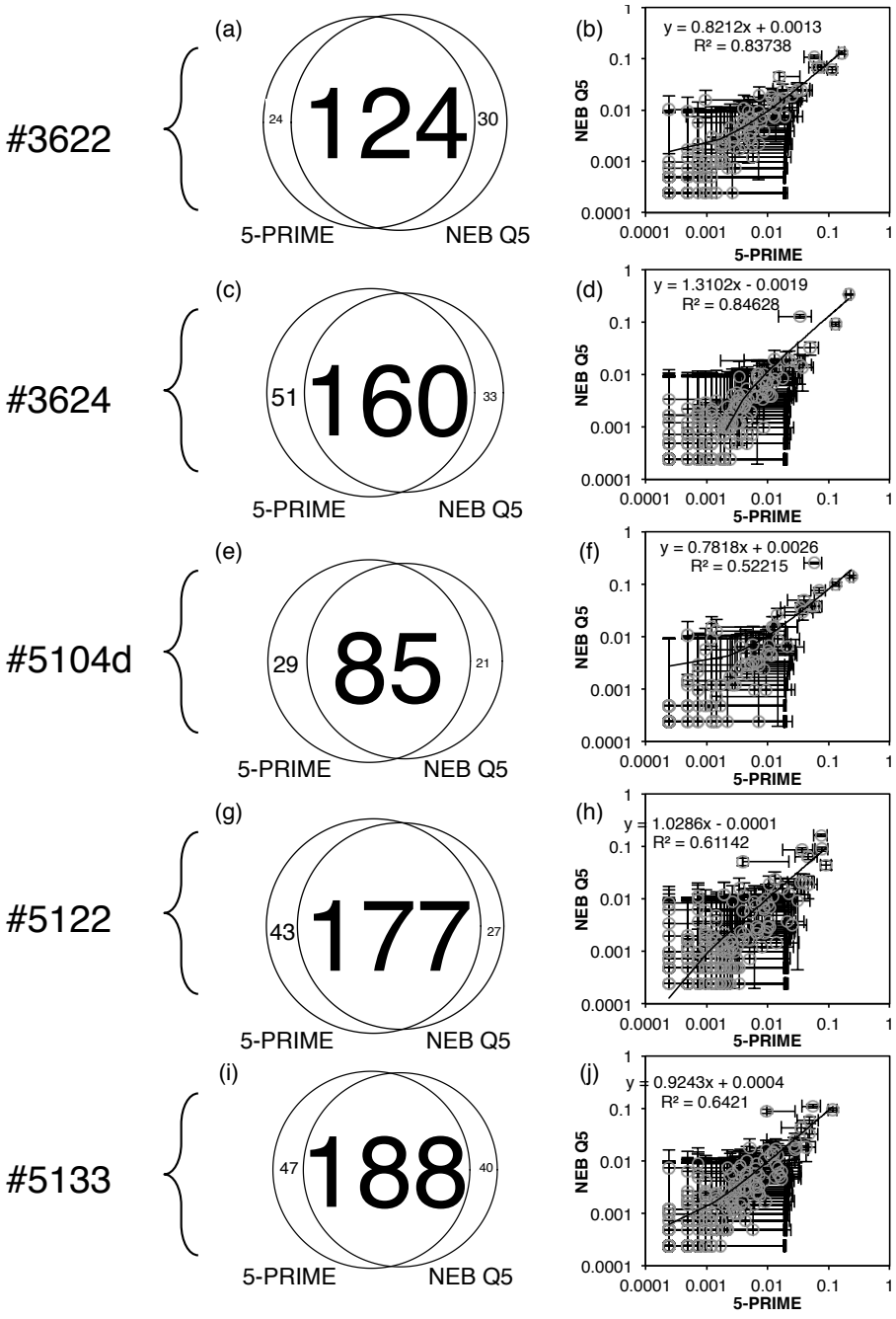


Figure 14. Comparison of amplification approach: 5-PRIME Hot Master Mix vs NEB Q5. (a,c,e,g,i) show the OTU overlap between the two amplifications. (b,d,f,g,j) show a cross plot of the shared OTUs. Although the majority of OTUs are shared, due to known biases between the enzymes (c.f. Mock Communities), the R^2 values in the cross plots relatively low with slopes of the regression line deviating from a value of 1.

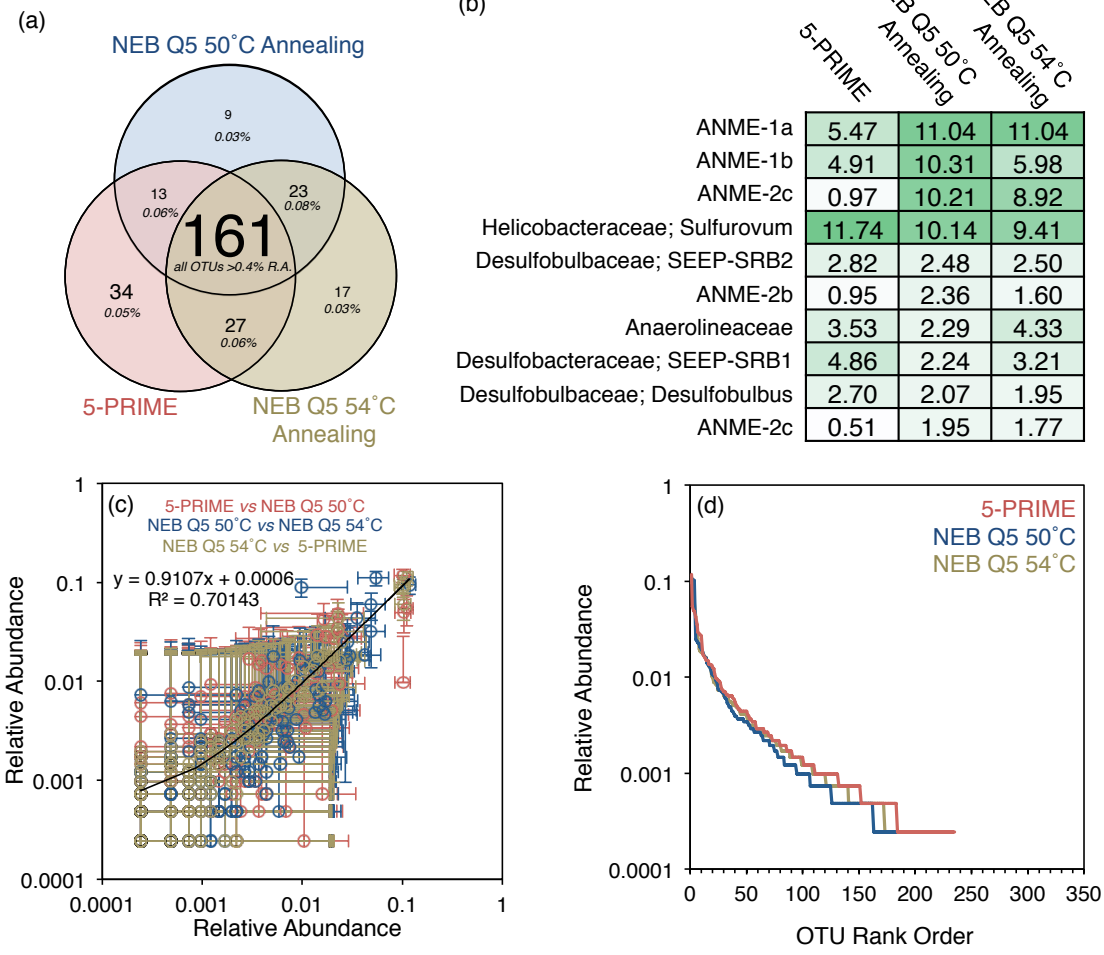


Figure 15. Comparison of amplification approach: amplification using 5-PRIME *Taq*, NEB Q5 annealing at 50˚C, or NEB Q5 annealing at 54˚C for sample #5133 (sediment) for iTag sequencing. (a) shows the OTU overlap between results from the three preparations. The majority of OTUs are shared, including all the major OTUs (any OTU >0.4% relative abundance in either sample is shared between the two samples). Only a small number of OTUs, with very low relative abundance, are not shared between preparations. (b) tabulates the relative abundances of the top 10 most abundant OTUs between the preparations, while (c) is a cross-plot of the relative abundance of the OTUs shared between the three preparations. Error bars in (c) are 1.85% relative abundance (c.f. precision for samples amplified with 5-PRIME *Taq*). (d) gives the OTU rank abundance curve for all three preparations.

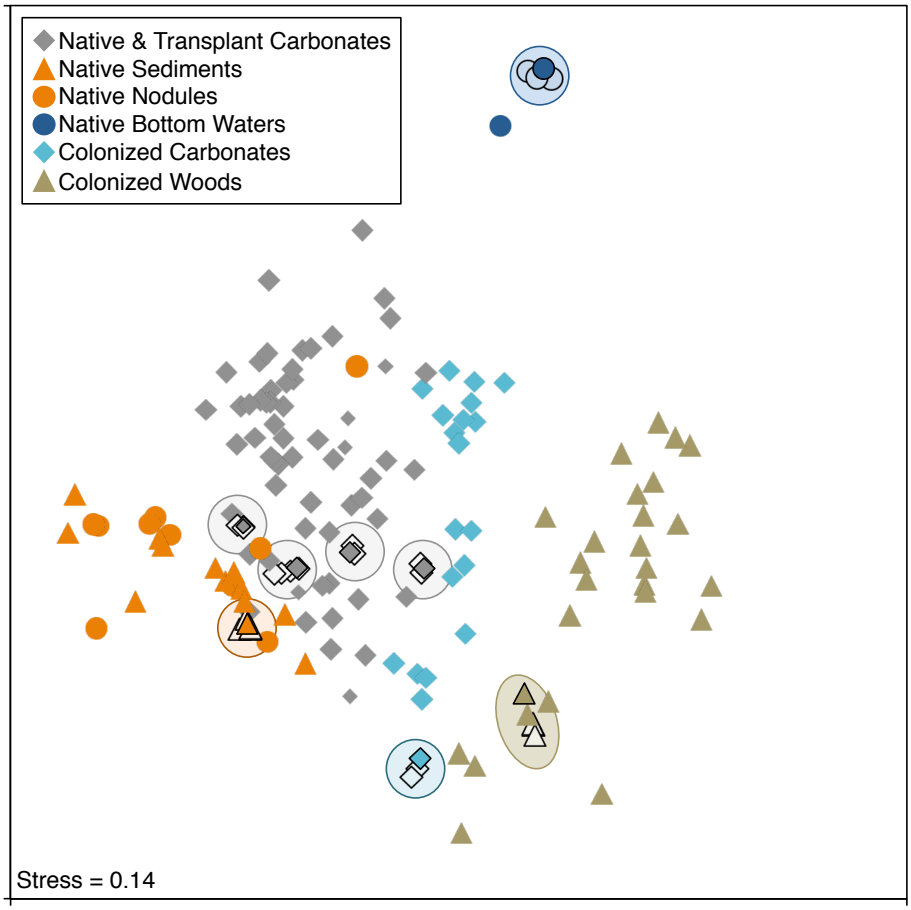


Figure 16. Nonmetric multidimensional scaling of 134 samples published in Case *et al.,* 2015, along with samples which were subjected to various methodology tests. Overall the methodology tests, despite differences in recovered 16S rRNA gene profiles, do not exhibit a large difference when compared to other samples in a large environmental dataset. Samples published in Case *et al.*, 2015 (using the "default" preparation and processing methodology) are given symbols with bold colors and a black border. Their corresponding samples which were subjected to methodology tests are identified by symbols with pale colors and a black border. Colored ovals are drawn by hand to guide the reader's eye to these groupings.

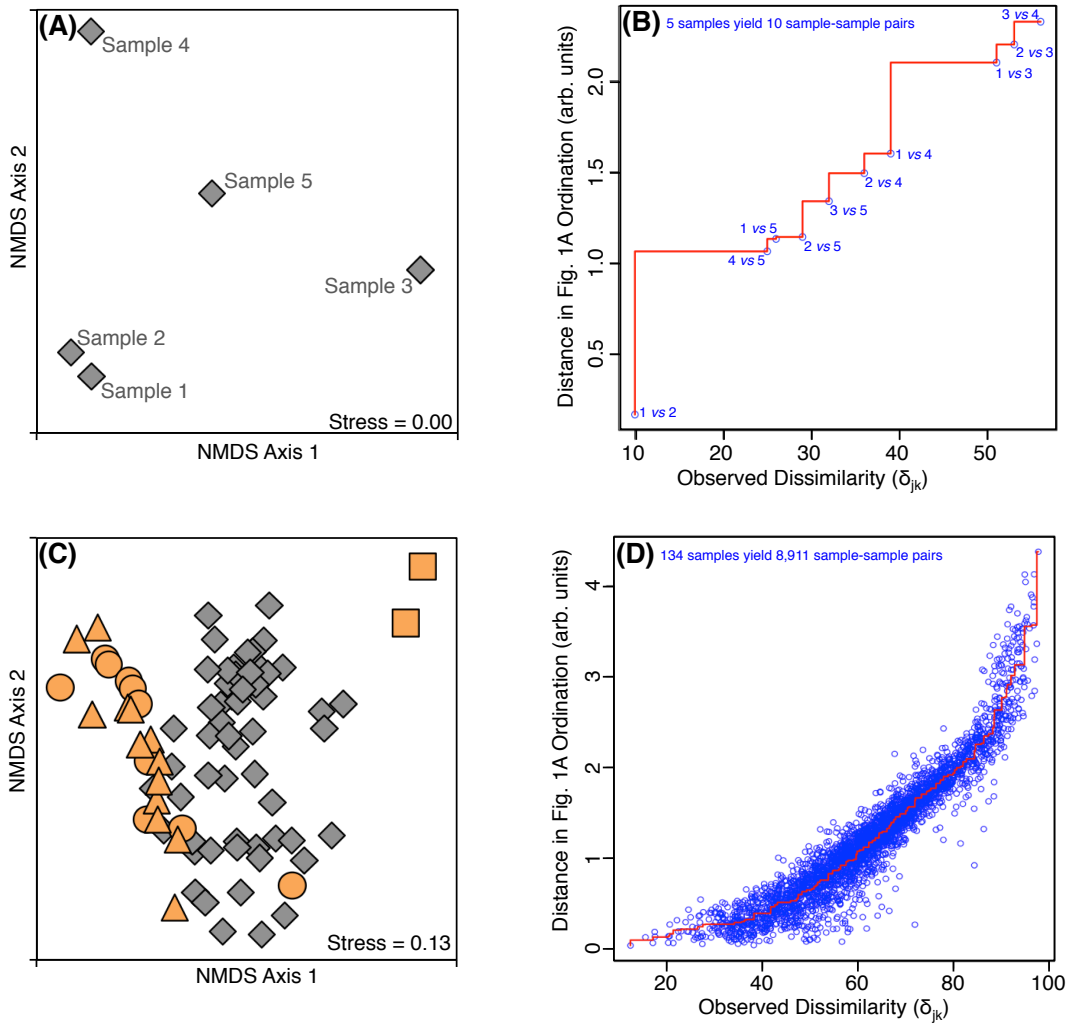


Figure 17. NMDS and Shepard plots. Panels (A) and (B) are calculated from synthetic data of five samples given in Tables 3-5. Panels (C) and (D) are calculated from 16S rRNA gene data from 134 methane seep samples (c.f., Chapter Two of this thesis; Case et al., 2015). In NMDS plots (panels (A) and (C)), each point represents the entire microbial assemblage from one sample. Data points closer to one another are more biologically similar. In (C), gray and orange indicate carbonate and non-carbonate habitats, respectively. Also in (C), circles, triangles, squares, and diamonds represent nodules, circles, bottom waters, and carbonates, respectively. In Shepard plots (panels (B) and (D)), the x-axis is calculated using Equation 2 of Chapter 4 and the y-axis is calculated by Euclidian distance on the accompanying NMDS plots. Every blue circle represents the dissimilarity and ordination distance between one pair of samples in the dataset. The red line is a monotonic regression to the blue circles. Stress, reported in the lower right corners of (A) and (C), is calculated as in Equation 3 of Chapter 4 by summing the differences between blue data points and the red regression line.

4.9 REFERENCES

Aronesty, E. 2011. ea-utils: Command-line tools for processing biological sequencing data.

Bentley, D. R., S. Balasubramanian, H. P. Swerdlow, G. P. Smith, J. Milton, C. G. Brown, K. P. Hall, D. J. Evers, C. L. Barnes, H. R. Bignell, J. M. Boutell, J. Bryant, R. J. Carter, R. K. Cheetham, A. J. Cox, D. J. Ellis, M. R. Flatbush, N. A. Gormley, S. J. Humphray, L. J. Irving, M. S. Karbelashvili, S. M. Kirk, H. Li, X. Liu, K. S. Maisinger, L. J. Murray, B. Obradovic, T. Ost, M. L. Parkinson, M. R. Pratt, I. M. J. Rasolonjatovo, M. T. Reed, R. Rigatti, C. Rodighiero, M. T. Ross, A. Sabot, S. V. Sankar, A. Scally, G. P. Schroth, M. E. Smith, V. P. Smith, A. Spiridou, P. E. Torrance, S. S. Tzonev, E. H. Vermaas, K. Walter, X. Wu, L. Zhang, M. D. Alam, C. Anastasi, I. C. Aniebo, D. M. D. Bailey, I. R. Bancarz, S. Banerjee, S. G. Barbour, P. A. Baybayan, V. A. Benoit, K. F. Benson, C. Bevis, P. J. Black, A. Boodhun, J. S. Brennan, J. A. Bridgham, R. C. Brown, A. A. Brown, D. H. Buermann, A. A. Bundu, J. C. Burrows, N. P. Carter, N. Castillo, M. C. E. Catenazzi, S. Chang, R. N. Cooley, N. R. Crake, O. O. Dada, K. D. Diakoumakos, B. Dominguez-Fernandez, D. J. Earnshaw, U. C. Egbujor, D. W. Elmore, S. S. Etchin, M. R. Ewan, M. Fedurco, L. J. Fraser, K. V. F. Fajardo, W. S. Furey, D. George, K. J. Gietzen, C. P. Goddard, G. S. Golda, P. A. Granieri, D. E. Green, D. L. Gustafson, N. F. Hansen, K. Harnish, C. D. Haudenschild, N. I. Heyer, M. M. Hims, J. T. Ho, A. M. Horgan, K. Hoschler, S. Hurwitz, D. V. Ivanov, M. Q. Johnson, T. James, T. A. H. Jones, G.-D. Kang, T. H. Kerelska, A. D. Kersey, I. Khrebtukova, A. P. Kindwall, Z. Kingsbury, P. I. Kokko-Gonzales, A. Kumar, M. A. Laurent, C. T. Lawley, S. E. Lee, X. Lee, A. K. Liao, J. A. Loch, M. Lok, S. Luo, R. M. Mammen, J. W. Martin, P. G. McCauley, P. McNitt, P. Mehta, K. W. Moon, J. W. Mullens, T. Newington, Z. Ning, B. L. Ng, S. M. Novo, M. J. O'Neill, M. A. Osborne, A. Osnowski, O. Ostadan, L. L. Paraschos, L. Pickering, A. C. Pike, A. C. Pike, D. C. Pinkard, D. P. Pliskin, J. Podhasky, V. J. Quijano, C. Raczy, V. H. Rae, S. R. Rawlings, A. C. Rodriguez, P. M. Roe, J. Rogers, M. C. R. Bacigalupo, N. Romanov, A. Romieu, R. K. Roth, N. J. Rourke, S. T. Ruediger, E. Rusman, R. M. Sanches-Kuiper, M. R. Schenker, J. M. Seoane, R. J. Shaw, M. K. Shiver, S. W. Short, N. L. Sizto, J. P. Sluis, M. A. Smith, J. E. S. Sohna, E. J. Spence, K. Stevens, N. Sutton, L. Szajkowski, C. L. Tregidgo, G. Turcatti, S. vandeVondele, Y. Verhovsky, S. M. Virk, S. Wakelin, G. C. Walcott, J. Wang, G. J. Worsley, J. Yan, L. Yau, M. Zuerlein, J. Rogers, J. C. Mullikin, M. E. Hurles, N. J. McCooke, J. S. West, F. L. Oaks, P. L. Lundberg, D. Klenerman, R. Durbin, and A. J. Smith. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. Nature **456**: 53–59.

Berry, D., K. Ben Mahfoudh, M. Wagner, and A. Loy. 2011. Barcoded Primers Used in Multiplex Amplicon Pyrosequencing Bias Amplification. Applied and Environmental Microbiology **77**: 7846–7849.

Bokulich, N. A., S. Subramanian, J. J. Faith, D. Gevers, J. I. Gordon, R. Knight, D. A. Mills, and J. G. Caporaso. 2013. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. Nature Methods **10**: 57–59.

Brandariz-Fontes, C., M. Camacho-Sanchez, C. Vilà, J. L. Vega-Pla, C. Rico, and J. A. Leonard. 2015. Effect of the enzyme and PCR conditions on the quality of high-throughput DNA sequencing results. Scientific Reports **5**: 8056.

Bray, J. R., and J. T. Curtis. 1957. An Ordination of the Upland Forest Communities of Southern Wisconsin. Ecological Monographs **27**: 325–349.

Caporaso, J. G., J. Kuczynski, J. Stombaugh, K. Bittinger, F. D. Bushman, E. K. Costello, N. Fierer, A. G. Peña, J. K. Goodrich, J. I. Gordon, G. A. Huttley, S. T. Kelley, D. Knights, J. E. Koenig, R. E. Ley, C. A. Lozupone, D. McDonald, B. D. Muegge, M. Pirrung, J. Reeder, J. R. Sevinsky, P. J. Turnbaugh, W. A. Walters, J. Widmann, T. Yatsunenko, J. Zaneveld, and R. Knight. 2010. QIIME allows analysis of high-throughput community sequencing data. Nature Methods **7**: 335–336.

Case, D. H., A. L. Pasulka, J. J. Marlow, B. M. Grupe, L. A. Levin, and V. J. Orphan. 2015. Methane Seep Carbonates Host Distinct, Diverse, and Dynamic Microbial Assemblages. mBio **6**: e01348–15.

Chandler, D. P., J. K. Fredrickson, and F. J. Brockman. 1997. Effect of PCR template concentration on the composition and distribution of total community 16S rDNA clone libraries. Mol Ecol **6**: 475–482.

Clarke, K. R., and R. M. Warwick. 2001. Change in Marine Communities, 2nd ed. PRIMER-E Ltd.

Dominguez-Bello, M. G., K. M. De Jesus-Laboy, N. Shen, L. M. Cox, A. Amir, A. González, N. A. Bokulich, S. J. Song, M. Hoashi, J. I. Rivera-Vinas, K. Mendez, R. Knight, and J. C. Clemente. 2016. Partial restoration of the microbiota of cesarean-born infants via vaginal microbial transfer. Nature Medicine **22**: 250–253.

Edgar, R. C., B. J. Haas, J. C. Clemente, C. Quince, and R. Knight. 2011. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics **27**: 2194–2200.

Edgar, R. C. 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics **26**: 2460-2461.

Faith, J. J., J. L. Guruge, M. Charbonneau, S. Subramanian, H. Seedorf, A. L. Goodman, J. C. Clemente, R. Knight, A. C. Heath, R. L. Leibel, M. Rosenbaum, and J. I. Gordon. 2013. The Long-Term Stability of the Human Gut Microbiota. Science **341**: 1237439.

Gilbert, J. A., F. Meyer, J. Jansson, J. Gordon, N. R. Pace, J. M. Tiedje, R. E. Ley, N. Fierer, D. Field, N. C. Kyrpides, F. O. Gloeckner, H. P. Klenk, K. E. Wommack, E. Glass, K. Docherty, R. Gallery, R. Stevens, and R. Knight. 2011. The Earth Microbiome Project: Meeting report of the "1st EMP meeting on sample selection and acquisition" at Argonne National Laboratory October 6th 2010. 1–5.

Kleppe, K., E. Ohtsuka, R. Kleppe, I. Molineux, and H. G. Khorana. 1971. Studies on polynucleotides: Repair Replication of Short Synthetic DNAs as catalyzed by DNA Polymerases. Journal of Molecular Biology **56**: 341–361.

Kozich, J. J., S. L. Westcott, N. T. Baxter, S. K. Highlander, and P. D. Schloss. 2013. Development of a Dual-Index Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence Data on the MiSeq Illumina Sequencing Platform. Applied and Environmental Microbiology **79**: 5112–5120.

Kruskal, J. B. 1964a. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika **29**: 1–27.

Kruskal, J. B. 1964b. Nonmetric multidimensional scaling: A numerical method. Psychometrika **29**: 115–129.

Legendre, P., and L. Legendre. 2012. Numerical ecology.

Levin, L. A., G. F. Mendoza, B. M. Grupe, J. P. Gonzalez, B. Jellison, G. W. Rouse, A. R. Thurber, and A. Waren. 2015. Biodiversity on the Rocks: Macrofauna Inhabiting Authigenic Carbonate at Costa Rica Methane Seeps. PLoS ONE 1–31.

Margulies, M., M. Egholm, W. E. Altman, S. Attiya, J. S. Bader, L. A. Bemben, J. Berka, M. S. Braverman, Y.-J. Chen, Z. Chen, S. B. Dewell, L. Du, J. M. Fierro, X. V. Gomes, B. C. Godwin, W. He, S. Helgesen, C. H. Ho, G. P. Irzyk, S. C. Jando, M. L. I. Alenquer, T. P. Jarvie, K. B. Jirage, J.-B. Kim, J. R. Knight, J. R. Lanza, J. H. Leamon, S. M. Lefkowitz, M. Lei, J. Li, K. L. Lohman, H. Lu, V. B. Makhijani, K. E. McDade, M. P. McKenna, E. W. Myers, E. Nickerson, J. R. Nobile, R. Plant, B. P. Puc, M. T. Ronan, G. T. Roth, G. J. Sarkis, J. F. Simons, J. W. Simpson, M. Srinivasan, K. R. Tartaro, A. Tomasz, K. A. Vogt, G. A. Volkmer, S. H. Wang, Y. Wang, M. P. Weiner, P. Yu, R. F. Begley, and J. M. Rothberg. 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature **437**: 376–380.

Mason, O. U., D. H. Case, T. H. Naehr, R. W. Lee, R. B. Thomas, J. V. Bailey, and V. J. Orphan. 2015. Comparison of Archaeal and Bacterial Diversity in Methane Seep Carbonate Nodules and Host Sediments, Eel River Basin and Hydrate Ridge, USA. Microbial Ecology **70**: 776–784.

Metcalf, J. L., Z. Z. Xu, S. Weiss, S. Lax, W. Van Treuren, E. R. Hyde, S. J. Song, A. Amir, P. Larsen, N. Sangwan, D. Haarmann, G. C. Humphrey, G. Ackermann, L. R. Thompson, C. Lauber, A. Bibat, C. Nicholas, M. J. Gebert, J. F. Petrosino, S. C. Reed, J. A. Gilbert, A. M. Lynne, S. R. Bucheli, D. O. Carter, and R. Knight. 2016. Microbial community assembly and metabolic function during mammalian corpse decomposition. Science **351**: 158–162.

Morono, Y., T. Terada, T. Hoshino, and F. Inagaki. 2014. Hot-Alkaline DNA Extraction Method for

Deep-Subseafloor Archaeal Communities. Applied and Environmental Microbiology **80**: 1985–1994.

Mullis, K. B., and F. A. Faloona. 1987. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction, p. 335–350. *In* Recombinant DNA Part F. Elsevier.

Nelson, M. C., H. G. Morrison, J. Benjamino, S. L. Grim, and J. Graf. 2014. Analysis, Optimization and Verification of Illumina-Generated 16S rRNA Gene Amplicon Surveys M.M. Heimesaat [ed.]. PLoS ONE **9**: e94249.

Oksanen, J., F. G. Blanchet, R. Kindt, P. Legendre, P. Minchin, R. B. OHara, G. Simpson, P. Solymos, M. H. H. Stevens, and H. Wagner. 2013. vegan: Community Ecology Package.

Parada, A. E., D. M. Needham, and J. A. Fuhrman. 2015. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. Environmental microbiology, doi:10.1111/1462-2920.13023

Pasulka, A. L., L. A. Levin, J. A. Steele, D. H. Case, M. R. Landry, and V. J. Orphan. 2015. Microbial eukaryotic distributions and diversity patterns in a deep-sea methane seep ecosystem. Environmental microbiology doi:10.1111/1462-2920.13185.

R Core Team. 2014. R: A language and environment for statistical computing.

Ramette, A. 2007. Multivariate analyses in microbial ecology. FEMS Microbiology Ecology **62**: 142–160.

Rosenbaum, M., R. Knight, and R. L. Leibel. 2015. The gut microbiota in human energy homeostasis and obesity. Trends in Endocrinology & Metabolism **26**: 493–501.

Ruiz-Calderon, J. F., H. Cavallin, S. J. Song, A. Novoselac, L. R. Pericchi, J. N. Hernandez, R. Rios, O. H. Branch, H. Pereira, L. C. Paulino, M. J. Blaser, R. Knight, and M. G. Dominguez-Bello. 2016. Walls talk: Microbial biogeography of homes spanning urbanization. Science Advances **2**: e1501061–e1501061.

Saiki, R. K., D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis, and H. A. Erlich. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science **239**: 487–491.

Sanger, F., and A. R. Coulson. 1975. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. Journal of Molecular Biology **94**: 441–448.

Sanger, F., G. M. Air, B. G. Barrell, N. L. Brown, A. R. Couson, J. C. Fiddes, C. A. Hutchinson III, P. M. Slocombe, and M. Smith. 1977a. Nucleotide sequence of bacteriophage ΦX174 DNA. Nature **265**: 687–695.

Sanger, F., S. Nicklen, and A. R. Coulson. 1977b. DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences **74**: 5463–5467.

Schadt, E. E., S. Turner, and A. Kasarskis. 2010. A window into third-generation sequencing. Human Molecular Genetics **19**: 227-240.

Shepard, R. N. 1962. The analysis of proximities: Multidimensional scaling with an unknown distance function. I. Psychometrika **27**: 125–140.

Trembath-Reichert, E., A. Green-Saxena, and V. J. Orphan. 2013. Whole Cell Immunomagnetic Enrichment of Environmental Microbial Consortia Using rRNA-Targeted Magneto-FISH, 1st ed. Elsevier Inc.

Trembath-Reichert, E., D. H. Case, and V. J. Orphan. 2016. Characterization of microbial associations with methanotrophic archaea and sulfate-reducing bacteria through statistical comparison of nested Magneto-FISH enrichments. PeerJ **4**: e1913–31.

Wang, Q., G. M. Garrity, J. M. Tiedje, and J. R. Cole. 2007. Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. Applied and Environmental Microbiology **73**: 5261-5267.

Woese, C. R., and G. E. Fox. 1977. Phylogenetic structure of the prokaryotic domain: The primary kingdoms. Proceedings of the National Academy of Sciences **74**: 5088–5090.

REMAINING QUESTIONS AND NEXT STEPS

Biomarker data in this thesis unequivocally shows that seep carbonate-associated microbial communities are sensitive to changes in seepage flux. Longevity of biomarkers appears to increase from 16S rRNA genes to IPLs to core lipids, with IPLs demonstrating turnover times more similar to 16S rRNA genes than core lipids. The work presented in this thesis raises multiple lines of inquiry to further understand the complex relationship between microorganisms, their biomarker taphonomy, carbonate mineralogy, and time- and splace-variant methane seepage:

• Do microbial biomarker profiles differ if a carbonate is sampled with spatially-resolved resolution, rather than in bulk? If so, are differences in microbial biomarker profiles associated with mineralogy, age, other factors, or a combination of multiple factors?

• How sensitive are seep microorganisms to fine-scale differences in magnitude of seepage flux, rather than binary "active" and "low-activity" designations? Is this the same for sediment- and carbonate-hosted seep microorganisms?

• What is the radial distance-decay rate between microbial assemblage similarity and methane seepage? Does it differ for sediment- and carbonate-hosted microorganisms?

• How do carbonate-specific microorganisms spread from seep to seep? Is endemism higher in seep carbonates than in seep sediments?

• Some bacterial IPLs appear to be preferentially associated with dolomite mineralogy – what species are these associated with, and can they inform the precipitation of dolomite in subsurface sediments?

• Do new carbonates precipitate at low-activity sites even after "high" levels of seepage disappear? What role does secondary alteration of mineralogy play in seep carbonates, and on what timescale?

• Is the hypothesis of log-linear changes in microbial biomarkers as a function of time since quiescence supported by experiments with finer time resolution?

PERSPECTIVES

My choice to study science was the result of a chain of inspirational mentors. In 7th grade, at Glacier Creek Middle School, Andrew Harris infused science with merriment and inventiveness. Three years later, in 10th grade Honors Chemistry, Julie Jensen appealed to my sense of organization and structure. In 12th grade, Kathryn Eilert's Biotechnology class drew intimate connections between science and daily life. Entering college, I knew I wanted to major in Chemistry, and was inspired over four years by various professors, including Bill Buhro and Dewey Holton. I was also introduced to laboratory research through the advising of Dan Giammar at Washington University and Laura Robinson at Woods Hole Oceanographic Institution. Their leadership helped me develop a keen interest in applying my Chemistry background to Earth science, which, with their support, I decided to pursue in graduate school. At Caltech a multitude of advisors, mentors, and colleagues further shaped my understanding of the natural world – Victoria Orphan, Jess Adkins, and Fumio Inagaki, to name a few.

I mention these to highlight the human component of one's intellectual development. Science can be, and often is, portrayed as an emotionless discipline, immune to sentiment and driven by dispassionate analysis in white lab coats and pristine laboratories. While scientists of course strive for objective analysis in their experiments, the actual day-to-day life of a scientist is very much driven by human interactions. Indeed, the inevitable challenges of scientific pursuit – failed experiments, unsupported hypotheses, conflicting results – are offset by the commiseration and positive encouragement of colleagues. It would probably not be possible, and certainly would not be pleasurable, to pursue science absent scientist friends. And so as I conclude my time in graduate school, I want to emphasize the importance of fostering not just a laboratory with top-notch equipment, but of seeking, growing, and maintaining an academic environment with friendly and supportive students, staff, and faculty throughout. Caltech and Victoria Orphan have provided this environment, for which I am grateful.

Inseparable from the view of science as a human-driven endeavor is the role of creativity in scientific progress. Despite the often promulgated view of scientists as stodgy Type-A persons who don't have the chops for the arts, I have found the best and brightest scientists to be among the most creative persons I have known. Working at the boundaries of human knowledge, exploring ideas and experiments which no human has ever before encountered, by definition requires an ability to think creatively. I have found, and indeed I did not appreciate this before graduate school, that creativity and a taste for the unknown are indispensable character traits for the successful scientist.

Perhaps the largest difference I can identify between my current understanding of science and that with which I arrived in graduate school is an appreciation for the inter-disciplinary nature of the endeavor. It is now clear to me that the borders between one discipline and another are diffuse – perhaps they do not exist at all. Certainly as a high school student, and even for much or all of my undergraduate education, it was natural to feel that science was divided into discrete units – Chemistry, Physics, and Biology, for example. Graduate school has taught me that not only are the disciplines not unique, but in fact understanding one is essential to understanding the others. In a sense this is an old viewpoint – more that of the well-rounded Renaissance scholar than the 20th century specialist. Coming around to this appreciation for the holistic nature of science has felt like the single largest "turning point" of my graduate career and maturation of my intellectual self. Furthermore, it feels entirely consistent with my now-credentialed title of *ecologist*. Ecology, by definition, deals with the relationships between organisms and their surroundings. The principles of ecology promote an appreciation for not merely the description of individuals, but the overarching principles that relate individuals to one another. Ecology is, by nature, a study of the in-betweens, the derivatives, the connectedness between units. I would argue, then, that if a PhD fundamentally involves an appreciation for the connectedness of all disciplines, then a PhD is, at its base level, an understanding of all science in an ecological manner. I believe that, whether they see it or not, all scientists are in some manner really ecologists, and that the title *Ecologist* is the highest honor to be afforded to a student of the natural world.

A p p e n d i x O n e

IN SITU DEPLOYABLE REACTOR ENABLES EXAMINATION OF MICROBIAL COMMUNITIES FROM HIGH PRESSURE ENVIRONMENTS

David H. Case[1]

in collaboration with,

Yuki Morono[2], Akira Ijiri[2], Patricia Tavormina[1], Makoto Nagasawa[3], Masamitsu Matsumoto[3], Victoria Orphan[1], and Fumio Inagaki[2]

[1]Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA, 91125, USA

[2]Geomicrobiology Group, Kochi Institute for Core Sample Research, Japan Agency for Marine-Earth Science and Technology, Monobe, Nankoku, Kochi, Japan

[3]Syn Corporation, Inc., Kyotanabe City, Japan

This chapter is in preparation for publication.

A1.0 ABSTRACT

High-pressure environments represent the volumetric majority of habitat space for microorganisms on the planet, including the pelagic deep-sea, deep-sea sediments, and both the terrestrial and marine deep subsurface biospheres. However, the importance of pressure as an environmental variable affecting, and possibly constraining, microbial life remains poorly constrained. This is due in part to the difficulty of accessing, sampling, and transporting samples from high-pressure environments back to the laboratory. In order to address this obstacle, we designed a new high-pressure corer which is deployable on the payload of a piloted or remotely operated deep-sea vehicle, can retrieve sediment samples *in situ*, can be recovered shipboard and transported back to onshore laboratories, and which can maintain high pressure conditions throughout multi-month incubations including daily amendments with liquid media and gases and daily effluent sampling for geochemical or microbiological analysis.

A 45-day incubation at 10 MPa and 4˚C of sediments from the seafloor of the Joetsu Knoll, Japan, indicated periods of both aerobic and anaerobic methanotrophy. These rates were generally in agreement with previously reported rates of methane oxidation in aqueous environments, with the exception that our calculated aerobic rates outpaced anaerobic rates. Whether, how, and to what extent pressure impacts the physiology of microorganisms in the deep-sea remains an open area of research, one hopefully made more accessible to researchers through the emergence of new technologies for high-pressure sampling and incubations.

A1.1 INTRODUCTION

High-pressure (HP) environments, including the deep sea, seafloor, and deep subsurface, represent the most volumetrically abundant habitats on the planet for microorganisms. However, our understanding of the diversity, physiology, and adaptability of HP-tolerant ("piezotolerant"), HP-preferring ("piezophilic"), and HP-requiring ("hyperpiezophilic" or "obligately piezophilic") microorganisms remains in early stages of research. The first active microbial communities from deep-sea sediments were described in 1957 from >10,000 meters below sea level (mbsl) in the Philippine Trench (Zobell and Morita 1957). The first isolation of an obligate microbial piezophile species from deep-sea sediments did not occur until over 20 years later, of a gammaproteobacterial *Colwellia* species (Yayanos et al. 1981). Since then, driven by the widespread use of molecular techniques, the diversity of piezophilic and hyperpiezophilic microorganisms has been extended to many clades of *Bacteria* (Yanagibayashi et al. 1999; Kato et al. 2008; Nagata et al. 2010; Zhang et al. 2015) and *Archaea* (Zeng et al. 2009; Birrien et al. 2011; Zhang et al. 2015), and probably varies significantly depending on particular environment.

Preliminary experiments have shown sediment-hosted microbial communities from the deep sea to be sensitive to changes in pressure. Diversity, as measured by 16S rRNA genes, diverges over time if sediments from the deep sea are maintained at atmospheric vs at representative deep-sea pressures (Yanagibayashi et al. 1999). Unsurprisingly, then, metabolic activity also differs whether experiments on deep-sea sediments are conducted at low or high pressures (Picard and Ferdelman 2012). Similar tests of diversity and metabolic activity as a function of pressure from deep biosphere and hydrothermal vent samples would increase our understanding of microbial life in extreme environments.

A significant obstacle in the study of piezophilic microorganisms has been sample recovery. High-pressure (e.g., deep-sea) environments are generally difficult and expensive to access, and even once accessed it is challenging to maintain samples at HP during transport back to a

research vessel and/or the home laboratory. Because of these logistical challenges, development of new sampling technology has been identified as a top priority in the field of HP microbiology (Kim and Kato 2010; Kato 2011; Zhang et al. 2015). Recent development of deep-subsurface coring technology which can maintain *in situ* pressure through recovery represents good progress (Kubo et al. 2014), but can only be deployed on large drill ships such as the D/V *Chikyu* at great expense and time commitment. Development of an affordable HP sampling device that could be deployed on the payload of a piloted or remotely operated vehicle (e.g., DSV *Shinkai 6500* or ROV *Hyperdolphin*) and which could retain *in situ* pressure through sample retrieval and shipment back to onshore laboratories would enable many members of the scientific community to pursue environmental microbiology research at high pressures. A device meeting many of these criteria was developed at the Japan Agency for Marine-Earth Science and Technology (JAMSTEC) in the 1990s, and successfully deployed, but was limited to small volumes of surface sediments and was specifically designed to perform dilution-to-extinction experiments rather than stable isotope or amendment incubations (Kyo et al. 1991; Kato et al. 2008).

In order to address the technical considerations of working in deep-sea HP environments, we (specifically co-authors Inagaki, Morono, and Ijiri) worked with engineers at Syn Corporation Ltd. (co-authors Nagasawa and Matsumoto) to develop and test a new HP corer. The goal for this device was to be (i) deployable on the payload of an ROV, (ii) to have a "push core-like" structure enabling sampling down to >10 cm below seafloor, (iii) to maintain HP through recovery onboard ship and shipment to onshore laboratories, (iv) to have inlet ports for adding liquid media and/or gas phase (including stable isotope) amendments to the incubation chamber, and (v) to have an outlet port to enable time-course tracking of an experiment without sacrificing pressure on the entire vessel. After fabrication, deployment of the device was tested on deep-sea sediments at the Joetsu Knoll, Japan, during R/V *Natsushima* cruise NT13-15 in July 2013 at a depth of 985 meters below sea level (mbsl; 9.9 MPa). Subsequent onshore incubation of the deep-sea sediments within the HP corer chamber, including liquid media and gaseous amendments, was performed for 45

days at the Kochi Core Center branch of JAMSTEC in Kochi, Japan. This thesis chapter reports microbiological and geochemical results indicating a successful deployment and onshore use of the HP corer. Technical details of the HP corer are in preparation for publication at a later date.

The well-characterized geology, geochemistry, and microbiology of the Joetsu Knoll made it a favorable control site for testing deployment of the HP corer. Massive gas hydrates outcrop on the seafloor at the Joetsu Knoll, sourced from thermogenically produced methane (Matsumoto et al. 2005). In addition to this rich source of reduced carbon, the Joetsu Knoll is bathed in oxygen-rich bottom water (>210 μmol/kg; Gamo and Horibe 1983), fueling diverse chemosynthetic microbial consortia. Previous 16S rRNA clone libraries from sediments at the Joetsu Knoll have revealed the presence of anaerobic methane oxidizing archaea (e.g., ANME-1 and ANME-2) in addition to a diversity of other *Archaea* and *Bacteria* (Yanagawa et al. 2011). Despite the high concentration of oxygen in overlying bottom waters, the presence, distribution, and/or activity of aerobic methanotrophs has not been specifically investigated at the Joetsu Knoll.

A1.2 METHODS

A1.2.1 EXPERIMENTAL SETUP

Sample collection was performed during cruise NT13-15 aboard the R/V *Natsushima* during July 2013. The study site was the Joetsu Knoll, a well-characterized location of methane seepage offshore Joetsu, Japan (Fig. 1; 37˚31.1'N, 137˚58.0'E, 985 meters below sea level, 9.9 MPa pressure, Yanagawa et al. 2011). Two sediment cores were collected during Dive 1555 of the ROV *Hyperdolphin*: firstly, sediment was collected into the HP corer (hereafter, "HP Core") by abrading the internal core cylinder against an exposed vertical wall of sediment interlaced with white methane hydrates and bacterial films (Fig. 1). The HP Core's internal cylinder was then immediately placed into the external cylinder and secured by tightening. In this manner, sediment

was collected at environmentally relevant pressure and sealed into the HP Core *in situ*, in order to maintain pressure throughout core recovery and onshore experimentation. Unfortunately the Teflon seal on the HP Core was compromised, most likely by sediment grains lodged against the Teflon core liner, resulting in a loss of pressure during transit of the ROV *Hyperdolphin* from seafloor to the R/V *Natsushima*. The core was quickly re-pressurized to 10 MPa onboard ship by injection of filtered artificial seawater; in total, the HP Core experienced a loss of pressure for <3.5 hours. The HP Core was stored at 4˚C and 10 MPa onboard, during shipment, and upon arrival at the Kochi Core Center. Besides the HP Core, a second core was collected from adjacent sediment into a traditional M-type corer (hereafter, "M core"). The material collected into the M core contained a mixture of sediment and bottom water, which by the time of recovery onboard ship had separated by density. Immediately onboard ship, subsamples of the "M core water" and "M core sediment" were frozen at -80˚C for later DNA extraction and sequencing.

The HP Core was kept for the duration of experimentation (total 45 days) in a walk-in 4˚C refrigerator in the laboratory. Twelve days after collection from the seafloor, the HP Core was amended with $^{13}CH_4$ (50 mL of 50% $^{13}CH_4$) and $^{15}N_2$ (50 mL of 50% $^{15}N_2$) and daily tracking of pressure, temperature, dissolved inorganic carbon concentration (DIC), and $\delta^{13}C_{DIC}$ began for the course of a 45-day experiment in high pressure incubation of seafloor microbial assemblages (Fig. 3-4). Temperature and pressure were continuously monitored (Δt=1 sec), with daily samples taken for $\delta^{13}C_{DIC}$. During daily sampling, pressure was reset to 10 MPa by injection of sterile artificial seawater that contained no carbon sources (Supplemental Text). Samples for DNA extraction were taken at 11, 25, and 45 days; the T11 days and T25 time points were by necessity of design captured from the effluent outflow (the only outflow port) at the top of the HP Core. This involved bleeding 6 mL of effluent from the top port, following by filtration onto a polycarbonate membrane and freezing. The T45 time point, however, was taken from the sediment at the bottom of the HP Core, which was only possible because the vessel could exclusively be opened to the atmosphere during takedown of the experiment. Besides the addition

of $^{13}CH_4$ and $^{15}N_2$ at the beginning of the experiment, the only other injections of exogenous gas were addition of 100% O_2 on days 29, 30, 35, 37, 39, and 44. In all cases 10 mL of O_2 was injected, excepted for the first injection on day 29, which was 5 mL of O_2.

A1.2.2 DISSOLVED INORGANIC CARBON (DIC) AND $\delta^{13}C$ MEASUREMENT

Carbon concentration and isotopic measurements were conducted on 0.2 μm-filtered effluent water samples <24 hours after collection. Measurements were performed on an isotope-monitoring gas chromatography/mass spectrometry (irm-GC/MS) ThermoFinnigan Delta Plus XP isotope-ratio mass spectrometer connected to TRACE GC as previously described (Ijiri et al. 2012).

A1.2.3 DNA EXTRACTION AND SEQUENCING

The M-core-sediment and M-core-water samples were extracted with the MoBio PowerMax soil DNA isolation kit according to manufacturer protocols (~5 g/extraction). The T11, T25, and T45 (duplicate samples of T45 were extracted and sequenced) time points were extracted with the MoBio PowerSoil DNA isolation kit according to manufacturer protocols (~0.5 g/extraction). In addition, duplicate T45 sediments were separately subjected to a simplified hot alkaline extraction (Morono et al. 2014), in which sequential cell lysis is performed in heated 1M sodium hydroxide solution (Supplemental Text).

Samples were prepared for deep sequencing of the V4 region of the 16S rRNA gene according to a slightly modified version of the Earth Microbiome Project's recommended protocol (Mason et al. 2015). New England Biolabs Q5 polymerase enzyme was substituted for 5-PRIME HotMasterMix. Sequencing was performed on an Illumina MiSeq platform at Laragen, Inc., and data processing (joining paired ends, trimming sequences, chimera checking, 97% OTU

picking, and taxonomic assignment) were performed as previously described (Case et al. 2015). Nonmetric multidimensional scaling (NMDS) analyses were performed in the R environment using the 'vegan' package on square-root-transformed tables of relative sequence abundance (Oksanen et al. 2013; R Core Team 2014).

In addition to sequencing of the 16S rRNA gene, an assay of the monooxygenase intergenic spacer region ("MISA") between pmoC and pmoA was performed following previously described protocols (Tavormina et al. 2010; see Supplemental Text for primer sequences, which have been modified since the initial publication) on two samples: the M-core-sediment and T45 sediment. Transformation of the MISA fragment into E. coli was performed with the 10G Elite Solo kit from Lucigen Corporation. Inserts were amplified with the Lucigen Corporation GC Vector Amplification pSMART kit and separately digested with HaeIII and RsaI restriction enzymes in order to generate restriction fragment length polymorphism (RFLP) patterns. Unique inserts were sequenced at Laragen, Inc. The resulting traces were manually checked for quality, translated to amino acid sequences, aligned against pure culture and previously published pmoA fragments in MUSCLE (Edgar 2004), and trimmed to the pmoA amino acids positions 5-49 of *M. capsulatas* Bath (an approach employed in Tavormina et al. 2010). These pmoA fragments, both experimental and from known organisms, were used to generate a 100-bootstrap, maximum likelihood tree in RAxML (Stamatakis 2014).

A1.3 RESULTS

The HP-Core was successfully deployed on the payload of the ROV *Hyperdolphin* during Dive 1555, capturing sediment in a challenging deep-sea environment and retaining sediment within the reaction chamber through recovery onboard ship. The loss of pressure during recovery was unfortunate, but with increased deployment experience and technical improvements, the HP-Core has been successfully recovered without pressure loss (F. Inagaki, personal communication).

Further, the HP-Core successfully maintained high pressure through shipping and during the course of 45 days of experimentation. There is no indication that the HP-Core would not have continued to retain pressure for a significantly longer experimental duration.

Over the course of 45 days, $\delta^{13}C_{DIC}$ was observed to increase, albeit relatively slowly for the first ~30 days (Fig. 4). A model of exponential increase in $\delta^{13}C_{DIC}$, fit from the data between T29 and T32, fits the data from T33-T45 very well (R^2=0.97; Fig. 4); this increase appears to be linked to the addition of O_2 beginning at T29. Stepwise rates of methane oxidation were calculated by subtracting the moles of ^{13}C observed between time points ($t_n - t_{n-1}$), on the assumption that new ^{13}C in the DIC pool represented newly oxidized $^{13}CH_4$ (Eq. 1; a factor of 2 was added because the methane amendment was only 50% $^{13}CH_4$). The trend in methane oxidation rate, by definition, mirrors the increase in $\delta^{13}C_{DIC}$ and shows increasing rates of methane oxidation late in the experiment.

[Eq.1] CH_4 oxidation rate (nM/day) =
$$2 \cdot 10^6 \cdot [([DIC]_n \cdot V_{HP\text{-}Core} - ([DIC]_n/(1+R_{std}((\delta^{13}C_n/1000)+1)) - ([DIC]_{n-1}/(1+R_{std}((\delta^{13}C_{n-1}/1000)+1))] \; / \; (n - n_{-1})$$

Within the period of the experiment prior to rapid increase in methane oxidation rate (T0-T29), slow but measurable methane oxidation is observed (c.f. inset of Fig. 4). Within the period T0-T29, more significant increase in $\delta^{13}C$ is apparent from T0-10 than T11-29.

iTag sequencing reveals three categories of microbial communities with our HP-Core and M-core dataset: the M-core water and sediment samples, the HP-Core effluent samples from T11 and T25, and the HP-Core sediment samples from T45 (Fig. 5-6). The M-core samples, both sediment and water, are characterized by high relative abundances of OTUs associated with Candidate Division JS1 bacteria (20-30%), Desulfobacteraceae (6-10%), Methylococcales (2-12%), and various ANME archaea (1-4%; Fig. 5; Table S1). The effluent samples from T11 and T25 lack OTUs associated with AMNE or Methylococcales and instead demonstrate high relative abundances of OTUs associated with Bacteroidetes, delta-, epsilon-, and gamma-proteobacteria;

they are the most abundant in the a Helicobacteraceae-associated OTU (9-17%), which is observed in moderate relative abundance throughout the iTag dataset (Fig. 5). In contrast the sediment samples from T45, similar to the M-core samples, are rich in OTUs associated with Methylococcales (31-30%) and moderately rich in Candidate Division JS1 OTUs (2-4%). However, the T45 sediment samples also share much of the diversity of delta-, epsilon-, and gamma-proteobacteria observed in the HP-Core effluent samples. Furthermore, the T45 sediment samples appear to be characteristically rich in a Pisirickettsiaceae-associated OTU (16-20%) which is relatively poor in the M-core and HP-core effluent samples (Fig. 5). The extraction method (MoBio vs Hot Alkaline) appears to make a measurable but overall small difference in the overall microbial 16S rRNA signature recovered from T45 sediments. When extracted with the MoBio kit, a Methylococcales-associated OTU is recovered at about two-thirds the relative abundance as recovered in samples extracted with the Hot Alkaline method. In contrast, a BD1-5-associated OTU is twice as abundant in MoBio-extracted samples as compared to Hot Alkaline-extracted samples (Fig. 4). These differences are apparent in multidimensional ordination, where the T45 samples overall plot closely together, but are distinctly separate according to extraction method (Fig. 5).

The MISA assay, applied to our M-core-water and T45.1-MoBio sediment from the HP-Core, is intentionally broad enough to recover sequences from both pmoA and amoA genes (Tavormina et al. 2010). Our dataset revealed both gene types, with pmoA more abundance than amoA. Furthermore, the pmoA genes recovered indicate both a diversity of pmoA genes present, as well as a difference in the pmoA composition between the M-core-water and T45.1-MoBio HP-Core sediment. Each of the two samples was particularly rich in one pmoA gene (Patterns 1 and 16 in Fig. 7), although overall these pmoA contigs were more similar to one another than to other pmoA gene from cultured or genome-sequenced organisms in the dataset (Fig. 7).

A1.4 DISCUSSION

A1.4.1 EVALUATION OF SUCCESS OF TECHNICAL GOALS

Of the five goals described for the first iteration of the HP corer, three were successfully achieved, one was partially achieved, and one was unachieved. The first goal, to develop a HP corer deployable on the payload of a piloted or remotely operated deep-sea vehicle, was achieved by deployment with the ROV *Hyperdolphin*. Furthermore, the HP corer was accompanied on the dive by a variety of other sampling devices: 6 push cores, a plastic tote for recovery of push cores previously deployed on the seafloor, a temperature probe and deep-sea high-pressure CO_2 injection system (Ohtomo et al. 2015), and an "M-type" corer. These other instruments and sampling devices were successfully employed on the same dive as the HP corer, lasting exactly two hours on the seafloor. Therefore, deployment of the HP corer does not impede other research aims during valuable at-sea time, and individual dives do not need to be dedicated to exclusive deployment of the HP corer.

The second goal, to sample >10 cm beneath the seafloor, was partially achieved. Due to the nature of the environment on the seafloor at the Joetsu Knoll, rather than choose to employ the HP corer like a traditional push core (straight down into sediment), we instead abraded the HP corer against a wall of mixed sediment, methane hydrate, and bacterial mats. This sampling action, while not strictly the equivalent to a push-core-like method, nonetheless captured a significant amount of sediment that might have been similar to the amount captured in a push-core-like maneuver. Abrading the HP corer against a wall of sediment allowed us to visualize the corer at all times, preferable for a first deployment of the new technology. There was no indication that pushing the HP corer vertically into a flat sediment bed would not have produced good sampling results as well.

Unfortunately, the third technical goal, to maintain HP conditions through recovery onboard ship, was not achieved. The HP corer arrived at the sea surface having lost pressure, most likely as

a result of sediment grains compromising the Teflon seal where the core interfaces with the core liner. Pressure was quickly restored with filtered seawater, and held stably over the next 24 hours, which encompassed the remainder of the time at sea. It is possible that the sediment grains compromising the vessel's seal were washed out during repressurization, enabling the maintenance of HP onboard. The HP corer successfully maintained pressure during shipment to the Kochi Core Center, demonstrating its resilience to shipping and handling. In subsequent deployments, the HP corer has been successfully employed and recovered onboard without pressure loss. Therefore, the system is capable of achieving the third goal of maintaining pressure from deep-sea sample collection to onshore incubation, but caution must be taken when choosing a sampling location and type.

Goals four and five, to be able to add liquid and gas amendments to the incubator during an experiment, and to be capable of extracting time-resolved output samples from the incubator, were successful. Sampling from the outflow port was performed daily. Sampling generated minor loss of internal pressure (generally less than 1 MPa, depending on user technique; Fig. 3). Pressure was restored daily by pumping in of fresh, sterile liquid media. Additionally, gas-phase amendments were injected in-line with the liquid media throughout the incubation.

Based on our experience with this first deployment, when deploying this first generation of the HP corer we recommend choosing carefully the sampling site and being extra cautious to avoid unnecessary sediment disturbance – excess sediment clouding bottom waters increase the likelihood of a compromised Teflon seal. Depending on time constraints, it is advisable to choose a seafloor location, hold the vehicle steady for enough time to let particles settle out, and only then to perform sampling. Additionally, it is good practice to perform HP corer sampling as the last function of a deep-sea dive. This minimizes both the amount of jostling on the HP corer and the time duration between sampling and recovery onboard ship.

Some recommendations can be made for future iterations of the technical design for the HP corer. The first generation had one outflow sampling port, located at the top of the vessel. This

port worked well, but daily outflow samples only represented the pelagic microbial community at the top of the incubation. Based on our 16S rRNA sequencing data (Fig. 5-6), we suspect this resulted in observing a different microbial community from that which resided deeper in the incubation column, in particular in the sediments settled at the bottom. Future iterations of the HP corer would be improved by having multiple outflow ports located at various heights along the incubation column. Similarly, the inflow port for adding liquid media and gas amendments only existed at the bottom of the incubation column. Although this worked for our experimental design, it is conceivable that future experiments would benefit from an ability to add amendments from either the top or bottom of the chamber – requiring engineering of additional inflow ports in future designs.

A1.4.2 MICROBIALLY MEDIATED METHANE OXIDATION DURING HP INCUBATION WITH $^{13}CH_4$

During the 45-day incubation, DIC and $\delta^{13}C_{DIC}$ data (collected and analyzed daily) suggested methane oxidation (Fig. 4). In the first 11 days (T0-T11) of the experiment methane oxidation, as determined by incorporation of ^{13}C into the DIC pool, appeared to be accelerating. However, it then plateaued and between T11 and T28 little methane oxidation was observed. We suspect the initial methane oxidation occurred by aerobic processes, and ceased when O_2 was fully consumed. If so, the theoretical amount of O_2 consumed by aerobic methanotrophy can be calculated by stoichiometric conversion using the following equation and ^{13}C data to track the number of moles of methane consumed between T0 and T11:

$$CH_4 + 2O_2 \rightleftharpoons CO_2 + 2H_2O \qquad \text{Eq. 2}$$

With 4.25 μmol of CH_4 consumed between T0 and T11 (calculated from data in Fig. 4), complementary oxidation of 8.50 μmol of O_2 is required. This is a relatively small amount compared to known bottom water O_2 concentrations in the Japan Sea (>220 μmol/kg), but it is likely that through the course of shipment of the HP corer to Kochi and static storage for twelve days at 4˚C prior to stable isotope amendment, a significant amount of aerobic respiration could have consumed most of the available O_2 in the incubation chamber. Respiratory processes could have continued during T0-T11, all independent of ^{13}C label and thus undetected by our geochemical measurements.

Between T11 and T28, the rate of methane oxidation was slower, with 7 μmol of CH_4 oxidized during the seventeen days. If we assume that during this period aerobic methanotrophy was replaced by sulfate-coupled anaerobic methanotrophy (AOM; Eq. 3), then only 7 μmol of sulfate are stoichiometrically required:

$$CH_4 + SO_4^{2-} \rightleftharpoons HS^- + HCO_3^- + H_2O \qquad \text{Eq. 3}$$

This is well within the bounds of seawater chemistry, where sulfate is generally present at ~28 mmol/kg. Anaerobic conditions during this period were supported by oxidation-reduction potential (ORP) measurements. When the first ORP measurement was taken, at T29, it was at the very reduced value of -300 mV. Although we do not have ORP data to help define exactly when anaerobic conditions began, it is clear than by T29 anaerobic conditions prevailed.

Further evidence for aerobic methanotrophy was the nearly immediate onset of an exponential rise in $\delta^{13}C_{DIC}$ upon additions of O2 beginning at T29. Over the course of six O_2 injections between T29 and T45, 55 mL of 100% were added to the incubation chamber, corresponding to 242 μmol of O_2. Conversion of $\delta^{13}C_{DIC}$ into consumption of methane molecules, we find aerobic methanotrophy would have consumed 143 μmol of O_2 between T29 and T45, consistent with our known amount of injected O_2. Regular ORP measurements between T29 and T45 reflected the addition of oxygen but also its rapid consumption: after T30, ORP averaged -33 mV (max=67 mV, min=-120 mV).

The hypothesis of sequential aerobic, anaerobic, and aerobic phases in the HP incubation chamber are additionally supported by sequencing data of the 16S rRNA and pmoA genes (Fig. 5-7). M-core samples, taken from sediments nearby the sampling location for the HP corer at the Joetsu Knoll, revealed the presence of anaerobic methanotrophs spanning the ANME-1, -2, and -3 clades as well as likely aerobic methanotrophs of the gammaproteobacterial *Methylococcales* order (Fig. 5). Thus, the requisite microorganisms for aerobic and anaerobic methane oxidation metabolisms were most likely present, at least at the time of sampling, in the sediments in the HP incubation chamber. The presence of aerobic methanotrophs in the M-core, and therefore also likely the HP corer, is further supported by abundant recovery of pmoA genes related to gammaproteobacterial methane oxidizers (Fig. 7). In addition, the microbial diversity data collected from the HP incubation chamber at T45 provides an excellent portrait of the microorganisms inhabiting the HP-incubated sediments at the end of experimentation. These sediments abundantly contained a *Methylococcales*-associated OTU but lacked ANME OTUs (Fig. 5). Sequences from the pmoA gene also revealed an abundance of gammaproteobacterial methanotrophs in the T45 sediment.

On the bulk scale, the sediments from the HP incubation at T45 differed in 16S rRNA gene diversity from the M-core samples (Fig. 6). This difference was likely influenced by several factors. First, because the M-core and HP core were not sampled at the exact same position on the seafloor, we cannot rule out spatial heterogeneity of microbial diversity and geochemistry. Indeed, short-range spatial diversity has been recorded previously at methane seep locations (Orphan et al. 2004). Second, it is likely that during the course of the 45-day incubation microbial successional processes resulted in a change in microbial diversity. Some of the major differences in 16S rRNA gene diversity between the M-core samples and the HP core T45 samples were the loss of ANME-associated OTUs and the retention of only one of two major *Methylococcales*-associated OTUs (Fig. 5).

In addition, the T45 sediments are richer in other gammaproteobacterial (e.g., *Colwelliaceae*) and epsilonproteobacterial OTUs (e.g., *Campylobacteraceae*) than the M-core sediments. These OTUs were observed at high relative abundance in the T11 and T25 effluent samples, suggesting they grew up during the course of incubation (Fig. 5). A *Colwelliaceae*-associated OTUs was observed to colonize sterile carbonate substrates in 13-month *in situ* methane seep incubations, suggesting the clade is adaptable to changing environmental parameters (Case et al. 2015). Similarly, many members of the *Epsilonproteobacteria*, including *Campylobacteraceae*, are sulfur-oxidizers (Campbell et al. 2006). As oxygen was depleted in the HP incubation chamber, sulfate would have been used as an electron acceptor, producing sulfide (both in AOM and anaerobic respiratory processes). This might explain the increased abundance of epsilonproteobacterial-associated OTUs at T11 onward.

The T45 sediments demonstrated slightly different 16S rRNA gene diversity whether extracted using a MoBio kit or with the hot alkaline lysis method (Fig. 6). This is likely due to differential lysis of cell walls (Morono et al. 2014), with the hot alkaline method yielding significantly higher relative abundances of *Methylococcales*-associated OTUs than the MoBio kit. Overall, however, the similarity among all T45 sediment samples (regardless of extraction method) is higher than between T45 samples and either T11 and T25 effluent samples or the M-core samples (Fig. 6).

A1.4.3 COMPARISON OF METHANE OXIDATION RATES MEASURED IN THE HP INCUBATION TO PREVIOUSLY PUBLISHED RATES

The rates of methane oxidation calculated from our high pressure experiment are similar to, but not higher than, previously published rates of aerobic and anaerobic methane oxidation (Fig. 8). Methane oxidation rates are often observed to vary by many orders of magnitude, depending on both methane and electron acceptor concentrations (Rudd and Hamilton 1974;

Harrits and Hanson 1980; Devol 1983; Iversen et al. 1987; Reeburgh et al. 1991; De Angelis and Lilley 1993; Ward and Kilpatrick 1993; Hoehler et al. 1994; Joye et al. 1999; Valentine et al. 2001; Nauhaus et al. 2002; Girguis et al. 2003; Carini et al. 2005). Our rates of methane oxidation between T11 and T29 (presumed to be AOM) are most similar to rates observed in Hoehler et al. (1994), in which sediment samples from shallow (10 mbsl) waters were incubated with $^{14}CH_4$. However, our relative rates of aerobic vs anaerobic methanotrophy are in contrast to multiple studies which examined both aerobic and anaerobic methane oxidation rates (Reeburgh et al. 1991; Hoehler et al. 1994; Joye et al. 1999). We note that of these studies, ours is the only one to be performed at high pressure. Our observation of higher aerobic than anaerobic methane oxidation rates may reveal a difference in sensitivity of aerobic vs anaerobic microorganisms to pressure, although we cannot definitely conclude this without more detailed physiological studies of pure cultures and environmental samples at a variety of pressure conditions.

A1.5 CONCLUSIONS

We have presented microbiological and geochemical data representing the first successful *in situ* deployment of a new high pressure chamber. The chamber fulfills most of the engineering goals, including: (i) deployable on the payload of a piloted or remotely operated deep-sea vehicle, (iv) containing input ports for liquid media and gas phase amendments, and (v) including an outlet port for time-resolved sampling without loss of pressure. Other goals, including (ii) the ability to sample >10 cm below the seafloor and (iii) ability to maintain HP conditions through recovery onboard and shipment to onshore laboratories, were partially successful or have been successful subsequent to our experiments. Future iterations of the HP corer design will incorporate improvements for sampling (e.g., multiple effluent outflow ports), and *in situ* deployment of the HP corer will hopefully increase over time as multiple laboratory groups gain access to the technology.

We demonstrated the usefulness of the HP corer by recovering sediments from the Joetsu Knoll, a methane-rich outcrop at ~1,000 mbsl in the Japan Sea. Incubation of these sediments at 10 MPa for 45 days in the presence of $^{13}CH_4$ demonstrated likely periods of aerobic and anaerobic methane oxidation, supported by multiple geochemical and microbiological measurements. The high rates of aerobic methane oxidation as compared to anaerobic methane oxidation differed from previous studies, further demonstrating the usefulness of the HP corer and importance of including pressure as a variable in future studies.

A1.6 ACKNOWLEDGEMENTS

We thank the crew of the R/V *Natsushima*, cruise 13-15, for their help at sea, without which this work would not have been possible. Masazumi Tsutsumi, Yoko Ohtomo, Go-Ichiro Uramoto, Tatsuhiko Hoshino, Takeshi Terada, Nan Xiao, and other members of the Kochi Core Center's Geomicrobiology Group helped with field and laboratory work. David Case was supported by a National Science Foundation (NSF) Graduate Research Fellowship as well as the East Asian and Pacific Summer Institutes (EAPSI) Summer Fellowship, co-funded by the NSF and the Japan Society for the Promotion of Science (JSPS).

A1.7 FIGURES

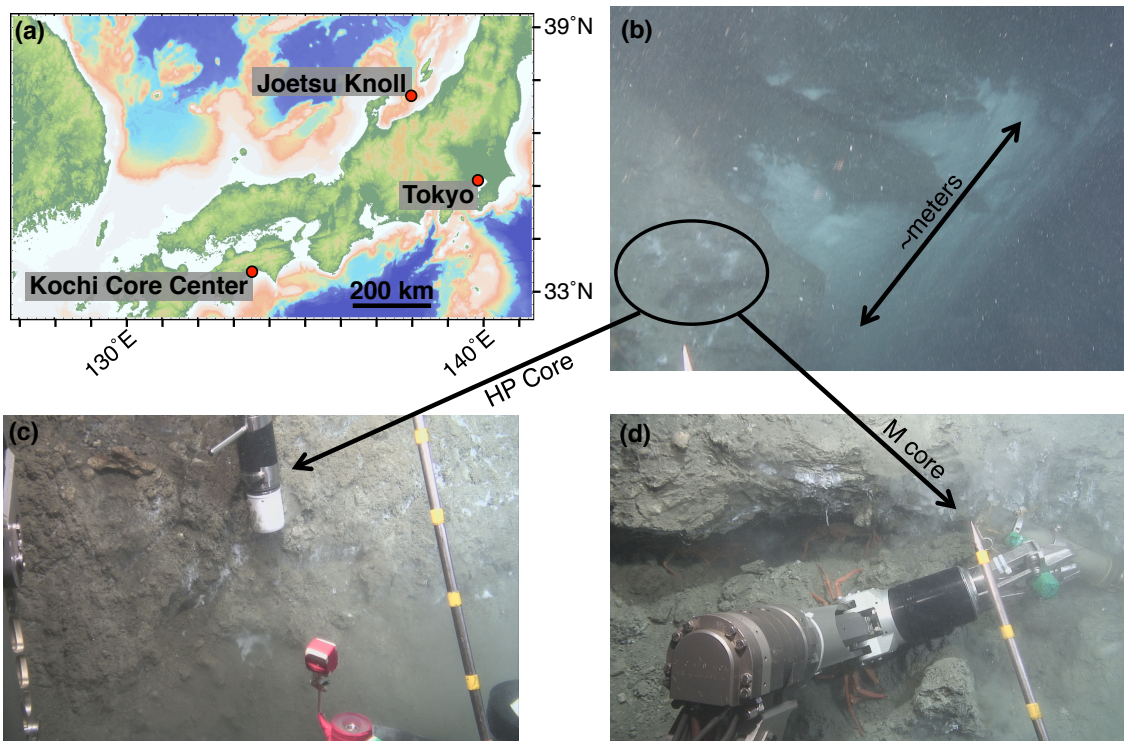


Figure 1. Contextualization of study site. (a) Map of central Japan, including the Joetsu Knoll study site. (b) Image capture from the ROV *Hyperdolphin* dive 1555, demonstrating the sampling location along a several meter-sized wall of methane clathrate. Samples were taken from roughly in the area of the black circle, near methane clathrates but with clear access to sediments for capture. (c) Sediment captured using the HP Core. (d) Sediment captured using the M-core.

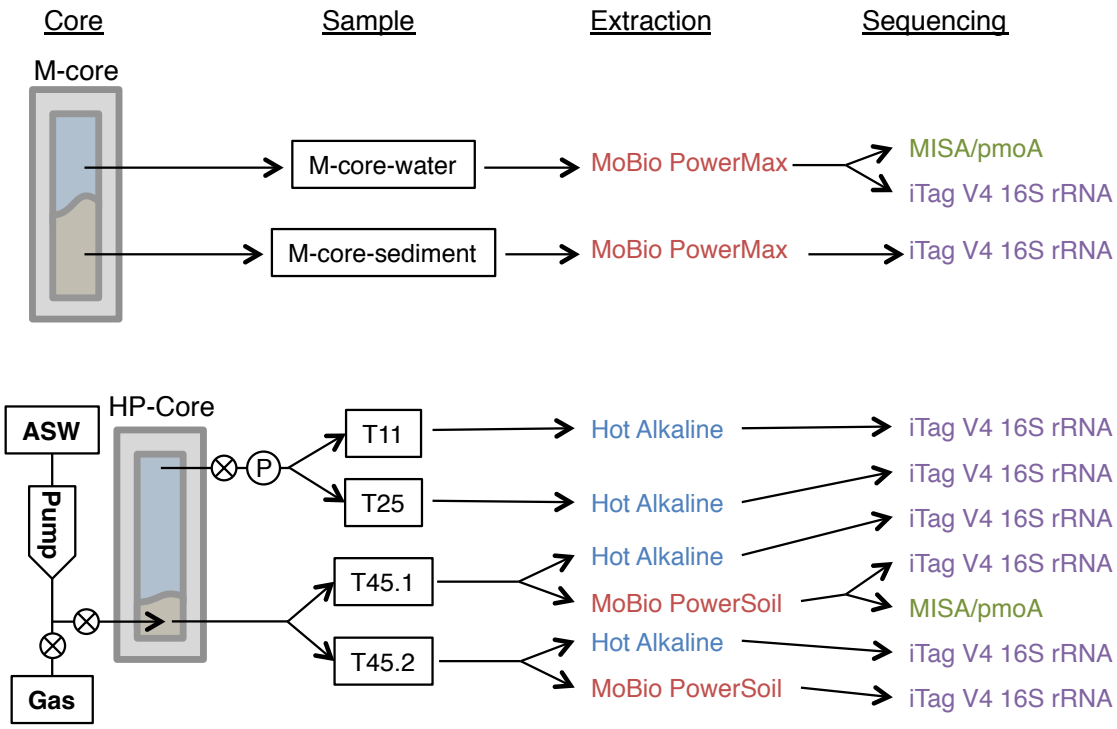


Figure 2. Schematic of samples and analyses. The M-core provided contextual evidence for the microbial diversity present in sediments at the time of sampling. The HP-Core, due to its technical aspects, was only able to be sampled from the effluent port during the course of experimentation, and from the sediment settled at the bottom during experimental takedown. All samples were subjected to 16S rRNA iTag sequencing. Two sediment samples, one each from the M-core and HP-Core T45, were subjected to the MISA/pmoA assay which specifically targets methane- and ammonia-oxidizing microorganisms.

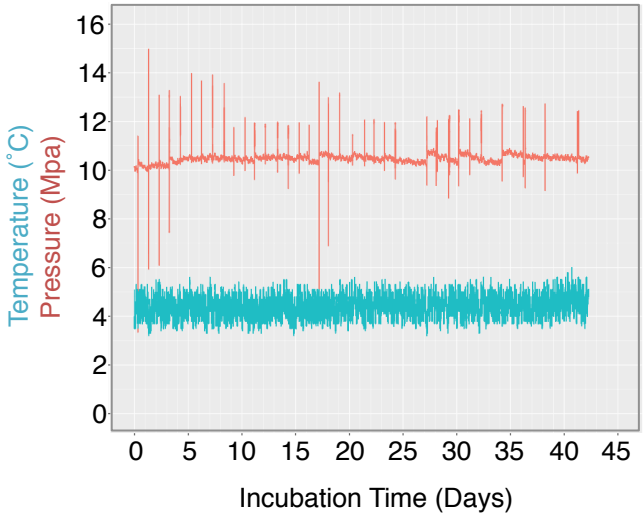


Figure 3. Log of HP-Core temperature and pressure over the duration of incubation. Temperature was maintained at ~4.5˚C by storing the HP-Core in a walk-in refridgerator throughout the experiment. Pressure was maintained at ~10 MPa (chosen to match the environmental pressure at the sampling depth of 985 mbsl) by injection of sterile artificial seawater via modified HPLC pump (c.f. Methods). Spikes in the pressure log record the daily effluent sampling for $\delta^{13}C_{DIC}$, during which time pressure fluctuated as the effluent port was opened. Over the course of >40 days, user technique improved and the fluctuations in pressure decreased in frequency and magnitude.

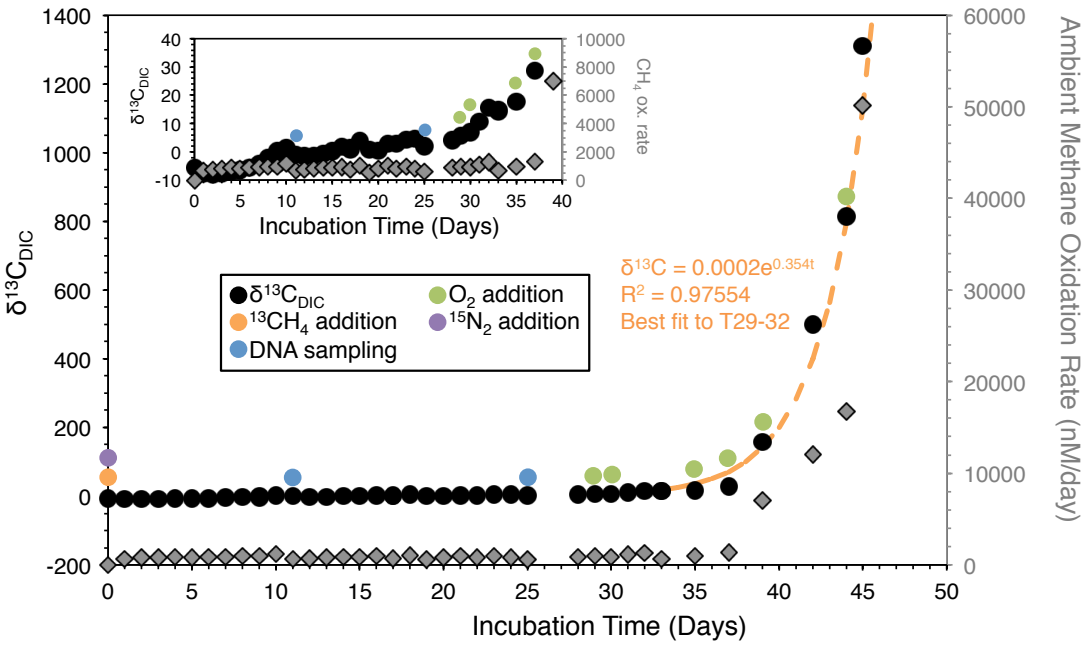


Figure 4. Time-resolved record of HP-Core incubation. Daily $\delta^{13}C_{DIC}$ measurements are given in black circles. Colored circles represent sampling or amendments (see legend). Gray diamonds are the calculated methane oxidation rate between each day and the day prior. Inset shows the same data on a smaller y-axis in order to better resolve trends within the first 40 days of the experiment.

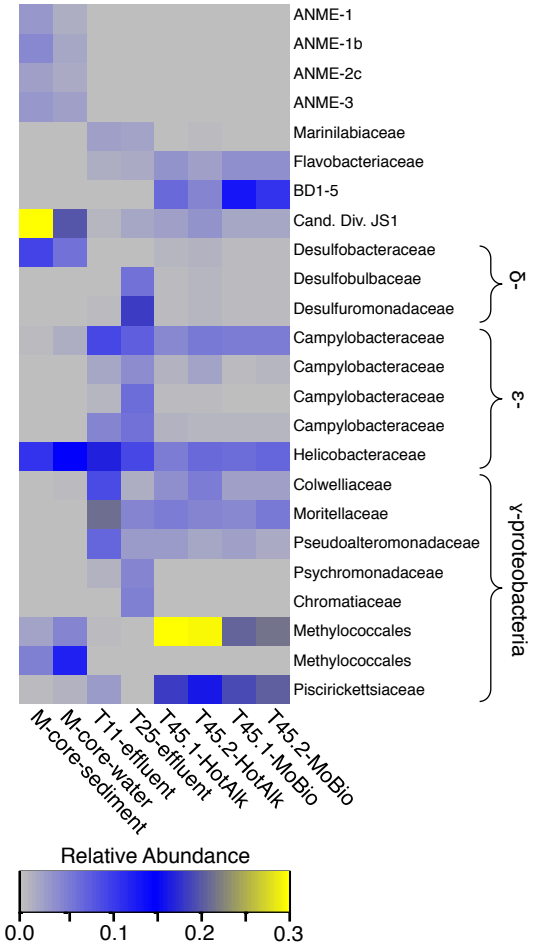


Figure 5. Heat map of major OTUs identified in the 16S rRNA iTag dataset. OTUs were only selected for presentation if they were present at >2% relative abundance in the M-core, HP-Core-effluent (T11 & T25), or HP-Core-sediment (T45) samples. M-core samples are characterized by their richness in Candidate Division JS1 bacteria. T11 and T25 effluent samples host a wide diversity of δ-, ε-, and ɣ-proteobacteria, but notably differ from the T45 samples which are rich in a Methylococcales-associated OTU. The full table of 16S rRNA data is provided in the Supplementary Information.

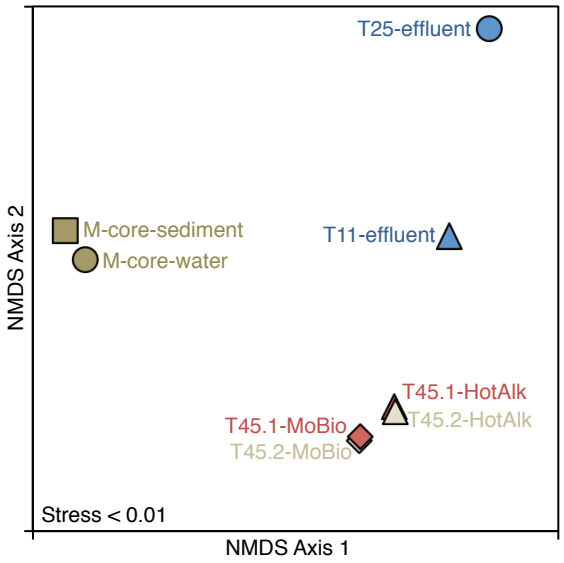


Figure 6. Nonmetric multidimensional scaling (NMDS) plot of 16S rRNA iTag data from this study. The microbial communities in the samples naturally break into three categories: M-core sediments, HP-Core effluent, and HP-Core sediments. Among the HP-Core sediments, DNA extraction method accounts for a measurable but small difference in recovered microbial community composition.

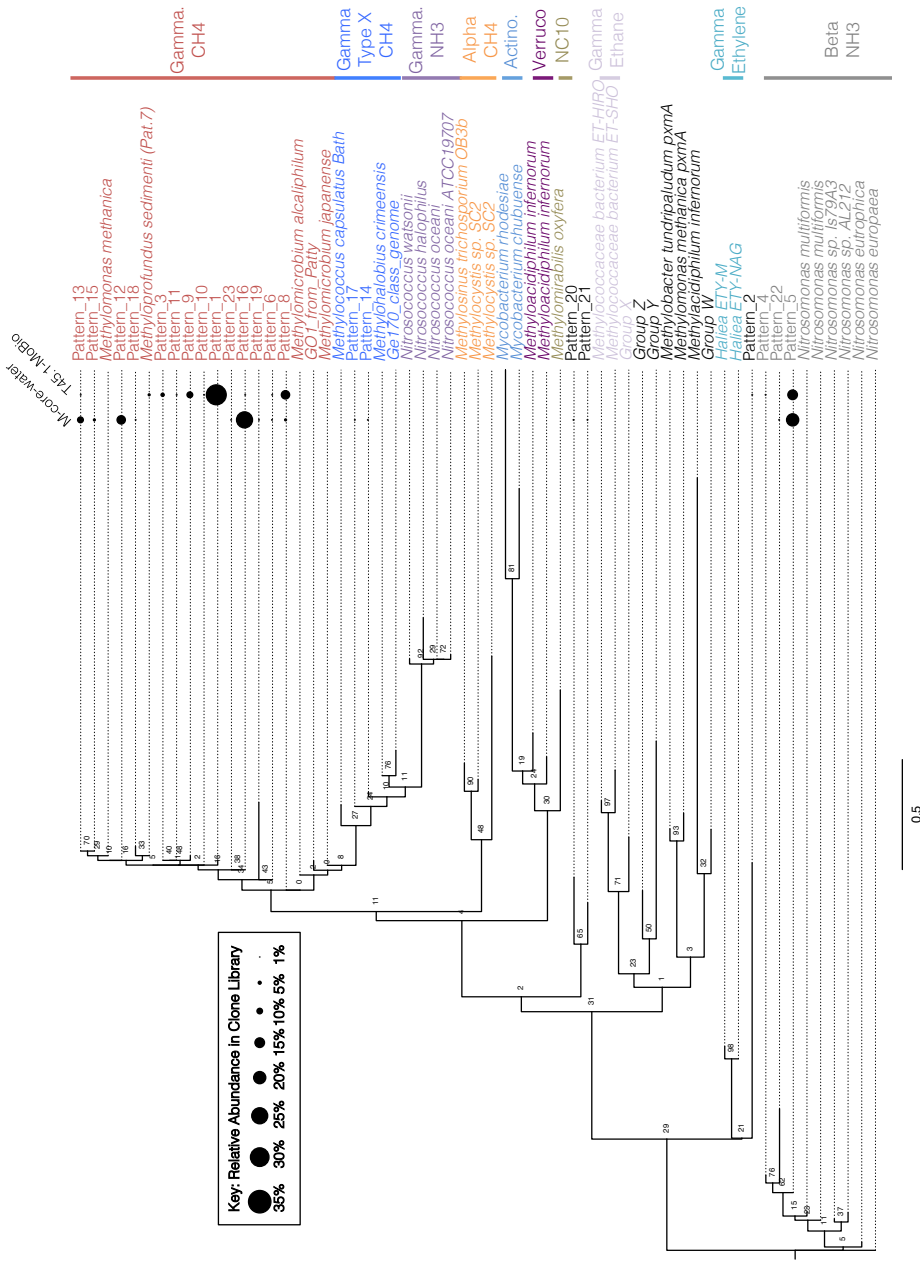


Figure 7. Maximum likelihood tree of pmoA (trimmed to amino acids 5-49 of *M. capsulatas Bath*) sequences generated in the MISA assay. Sequences from this study are defined as "Pattern_*x*" according to unique HaeIII and RsaI RFLP profiles. Sequences from cultured organisms and sequenced genomes of methane- and ammonia-oxidizers are given with their species name as appropriate. Multiple sequences alignments were generated in MUSCLE and the tree was generated in RAxML with 100 bootstraps. Black circles represent relative abundance of pmoA sequences in the M-core-water and T45.1-MoBio samples. The largest contrast between the two samples is seen in the abundance of different methane-oxidizing organisms.

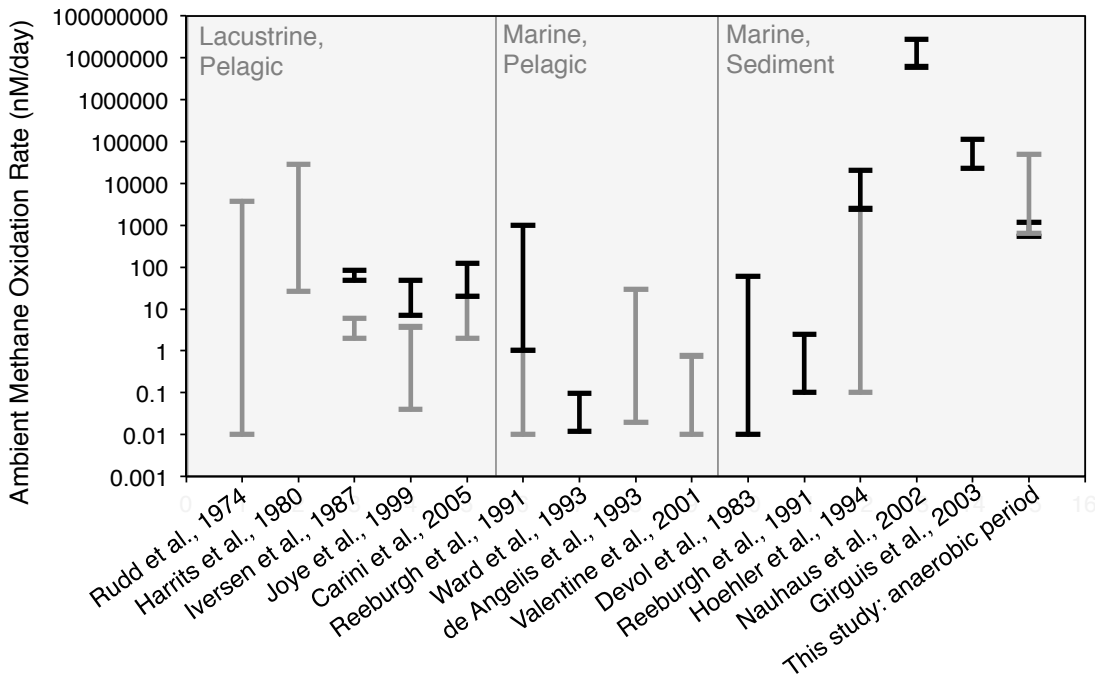


Figure 8. Comparison of ambient methane oxidation rate measurements between this study and previous studies. Two values are given for this study: calculated methane oxidation rates for the period of putative aerobic conditions (T0-T10 & T29-T45) and putative anaerobic conditions (T11-T28). Methane oxidation rates derived from aerobic methanotrophy are given in gray, whereas rates from anaerobic methane oxidation are given in black.

A1.8 SUPPLEMENTAL MATERIAL: TEXT

Composition of artificial seawater used to pressurize HP Core daily.
The artificial seawater (ASW) was made by reference to composition of methanosarchina sp. Strain BT-MS1, without the addition of yeast extract or any carbon source. Trace element solution used was not for methanosarchina but methanothermococcus okinawensis, because it was hypothesized that a trace element solution including $CuSO_4$ may be better to stimulate methanotrophy in our incubation. After all components were dissolved, the ASW was autoclaved.

Specific protocol for hot alkaline DNA extraction.
The protocol is modified after (Morono et al. 2014). Recipes for lysis and neutralization buffer can be found in the original publication.

1. Prewarm water baths to 50˚C and 70˚C.
2. Add 50 µL of lysis solution to ~50 mg sediment in PCR tube.
3. Heat for 20 min at 50˚C.
4. Centrifuge in mini-fuge for 30 sec at 25˚C.
5. Transfer supernatant to new PCR tube pre-loaded with 37.5 µL of neutralization buffer.
6. Wash remaining sample with 50 µL of 50˚C water.
7. Centrifuge in mini-fuge for 30 sec at 25˚C.
8. Transfer supernatant to the PCR tube from step (5).
9. Add 50 µL of lysis solution to remaining sediment in PCR tube.
10. Heat for 20 min at 70˚C.
11. Centrifuge in mini-fuge for 30 sec at 25˚C.
12. Transfer supernatant to new PCR tube pre-loaded with 37.5 µL of neutralization buffer.
13. Wash remaining sample with 50 µL of 70˚C water.
14. Centrifuge in mini-fuge for 30 sec at 25˚C.
15. Transfer supernatant to the PCR tube from step (12).
16. Combine supernatant from (8) and (15) into one tube.

Primers used for MISA assay.
The MISA assay involves two PCR steps, both targeting the pmoC-pmoA intergenic spacer region. Since the original MISA publication in 2010 (Tavormina et al. 2010), the primers have been further modified. The primers employed in this study were:

PCR#1, Forward primer, spacer_pmoC_599f:	`AAY GAR TGG GGH CAY RCB TTC`
PCR#1, Reverse primer, spacer_pmoA_192r:	`TCD GMC CAR AAR TCC CAR TC`
PCR#2, Forward primer, spacer_pmoC626_mod_f:	`RCB TTC TGG HTB ATG GAA GA`
PCR#2, Reverse primer, spacer_pmoA_189r:	`CCA RAA RTC CCA RTC NCC`

A1.9 SUPPLEMENTAL MATERIAL: TABLES

Table S1. Relative abundance of 16S rRNA gene iTag data for all samples in the study.

Supplementary Table 1 can be found in .xlsx format in the Caltech repository along with this thesis.

A1.10 REFERENCES

Birrien, J.-L., X. Zeng, M. Jebbar, M. A. Cambon Bonavita, J. Quérellou, P. Oger, N. Bienvenu, X. Xiao, and D. Prieur. 2011. Pyrococcus yayanosii sp. nov., an obligate piezophilic hyperthermophilic archaeon isolated from a deep-sea hydrothermal vent. Int J Syst Evol Microbiol **61**: 2827–2881.

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

Carini, S., N. Bano, and G. LeCleir. 2005. Aerobic methane oxidation and methanotroph community composition during seasonal stratification in Mono Lake, California (USA). Environmental Microbiology **7**: 1127-1138.

Case, D. H., A. L. Pasulka, J. J. Marlow, B. M. Grupe, L. A. Levin, and V. J. Orphan. 2015. Methane Seep Carbonates Host Distinct, Diverse, and Dynamic Microbial Assemblages. mBio **6**: e01348–15.

De Angelis, M., and M. Lilley. 1993. Methane oxidation in deep-sea hydrothermal plumes of the Endeavour Segment of the Juan de Fuca Ridge. Deep Sea Research I **40**: 1169-1186.

Devol, A. H. 1983. Methane oxidation rates in the anaerobic sediments of Saanich Inlet1. Limnology and Oceanography **28**: 738–742.

Edgar, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucl. Acids Res. **32**: 1792–1797.

Gamo, T., and Y. Horibe. 1983. Abyssal Circulation in the Japan Sea. Journal of the Oceanographical Society of Japan **39**: 220–230.

Girguis, P., V. Orphan, S. Hallam, and E. F. DeLong. 2003. Growth and methane oxidation rates of anaerobic methanotrophic archaea in a continuous-flow bioreactor. Applied and Environmental Microbiology **69**: 5472–5482.

Harrits, S. M., and R. S. Hanson. 1980. Stratification of aerobic methane-oxidizing organisms in Lake Mendota, Madison, Wisconsin1. Limnology and Oceanography **25**: 412–421.

Hoehler, T. M., M. J. Alperin, D. B. Albert, and C. S. Martens. 1994. Field and laboratory studies of methane oxidation in an anoxic marine sediment: Evidence for a methanogen-sulfate reducer consortium. Global Biogeochemical Cycles **8**: 451.

Ijiri, A., N. Harada, A. Hirota, U. Tsunogai, N. O. Ogawa, T. Itaki, B.-K. Khim, and M. Uchida. 2012. Biogeochemical processes involving acetate in sub-seafloor sediments from the Bering Sea shelf break. Organic geochemistry **48**: 47–55.

Iversen, N., R. S. Oremland, and M. J. Klug. 1987. Big Soda Lake (Nevada). 3. Pelagic methanogenesis and anaerobic methane oxidation. Limnology and Oceanography **32**: 804-814.

Joye, S. B., T. L. Connell, L. G. Miller, R. S. Oremland, and R. S. Jellison. 1999. Oxidation of ammonia and methane in an alkaline, saline lake. Limnology and Oceanography **44**: 178–188.

Kato, C., Y. Nogi, and S. Arakawa. 2008. Isolation, Cultivation, and Diversity of Deep-Sea Piezophiles, p. 203–217. *In* C. Michiels, D.H. Bartlett, and A. Aertsen [eds.], High-Pressure Microbiology. ASM Press.

Kato, D. C. 2011. Cultivation Methods for Piezophiles, p. 719–726. *In* Extremophiles Handbook. Springer Japan.

Kim, S. J., and C. Kato. 2010. Sampling, Isolation, Cultivation, and Characterization of Piezophilic Microbes, p. 3869–3881. *In* K.N. Timmis [ed.], Handbook of Hydrocarbon and Lipid Microbiology. Springer Berlin Heidelberg.

Kubo, Y., Y. Mizuguchi, F. Inagaki, and K. Yamamoto. 2014. A new hybrid pressure-coring system for the drilling vessel *Chikyu*. Sci. Dril. **2**: 1–7.

Kyo, M., Y. Tuji, H. Usui, and T. Itoh. 1991. Collection, Isolation And Cultivation System For Deep-sea Microbes Study. Proceedings of the Ocean Technologies and Opportunities in the Pacific. IEEE. 419–423.

Mason, O. U., D. H. Case, T. H. Naehr, R. W. Lee, R. B. Thomas, J. V. Bailey, and V. J. Orphan. 2015. Comparison of Archaeal and Bacterial Diversity in Methane Seep Carbonate Nodules and Host Sediments, Eel River Basin and Hydrate Ridge, USA. Microbial Ecology **70**: 776–784.

Matsumoto, R., Y. Okuda, C. Aoyama, A. Hiruta, Y. Ishida, M. Sunamura, H. Numanami, H. Tomaru, G. T. Snyder, J. Komatsubara, R. Takeuchi, and M. Hiromatsu. 2005. Methane plumes over a marine gas hydrate system in the eastern margin of Japan Sea: A possible mechanism for the transportation of sub-surface methane to shallow waters. 749–754.

Morono, Y., T. Terada, T. Hoshino, and F. Inagaki. 2014. Hot-Alkaline DNA Extraction Method for Deep-Subseafloor Archaeal Communities. Applied and Environmental Microbiology **80**: 1985–1994.

Nagata, T., C. Tamburini, J. Arístegui, F. Baltar, A. B. Bochdansky, S. Fonda-Umani, H. Fukuda, A. Gogou, D. A. Hansell, R. L. Hansman, G. J. Herndl, C. Panagiotopoulos, T. Reinthaler, R. Sohrin, P. Verdugo, N. Yamada, Y. Yamashita, T. Yokokawa, and D. H. Bartlett. 2010. Emerging concepts on microbial processes in the bathypelagic ocean – ecology, biogeochemistry, and genomics. Deep Sea Research Part II: Topical Studies in Oceanography **57**: 1519–1536.

Nauhaus, K., A. Boetius, and M. Kruger. 2002. In vitro demonstration of anaerobic oxidation of methane coupled to sulphate reduction in sediment from a marine gas hydrate area. Environmental Microbiology **4**: 296-305.

Ohtomo, Y., A. Ijiri, Y. Morono, Y. Ikegawa, H. Suenaga, D. Case, H. Machiyama, F. Yamamoto, S. Goto, T. Fukuba, M. Nagasawa, M. Matsumoto, and F. Inagaki. 2015. CO_2 emission and shallow-type methane hydrate decomposition experiment on deep-sea floor. JAMSTEC-R **20**: 61–71.

Oksanen, J., F. G. Blanchet, R. Kindt, P. Legendre, P. Minchin, R. B. OHara, G. Simpson, P. Solymos, M. H. H. Stevens, and H. Wagner. 2013. vegan: Community Ecology Package.

Orphan, V., W. Ussler, T. H. Naehr, C. H. House, Hinrichs, and C. K. Paull. 2004. Geological, geochemical, and microbiological heterogeneity of the seafloor around methane vents in the Eel River Basin, offshore California. Chemical Geology **205**: 265–289.

Picard, A., and T. G. Ferdelman. 2012. Microbial activity in deep marine sediments: does pressure make the difference? J. Phys.: Conf. Ser. **377**: 1–5.

R Core Team. 2014. R: A language and environment for statistical computing.

Reeburgh, W., B. Ward, and S. Whalen. 1991. Black Sea methane geochemistry. Deep Sea Research **38**: 1189-1210.

Rudd, J., and R. Hamilton. 1974. Measurement of microbial oxidation of methane in lake water. Limnology and Oceanography **19**: 519-524.

Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 1-2.

Tavormina, P., W. Ussler, S. Joye, B. Harrison, and V. Orphan. 2010. Distributions of putative aerobic methanotrophs in diverse pelagic marine environments. The ISME Journal **4**: 700–710.

Valentine, D., D. Blanton, and W. Reeburgh. 2001. Water column methane oxidation adjacent to an area of active hydrate dissociation, Eel River Basin. Geochimica et Cosmochimica Acta **65**: 2633-2640.

Ward, B., and K. Kilpatrick. 1993. Methane oxidation associated with mid-depth methane maxima in the Southern California Bight. Continental Shelf Research **13**: 1111–1122.

Yanagawa, K., M. Sunamura, M. A. Lever, Y. Morono, A. Hiruta, O. Ishizaki, R. Matsumoto, T. Urabe, and F. Inagaki. 2011. Niche Separation of Methanotrophic Archaea (ANME-1

and -2) in Methane-Seep Sediments of the Eastern Japan Sea Offshore Joetsu. Geomicrobiology Journal **28**: 118–129.

Yanagibayashi, M., Y. Nogi, L. Li, and C. Kato. 1999. Changes in the microbial community in Japan Trench sediment from a depth of 6292 m during cultivation without decompression. FEMS Microbiology Letters **170**: 271–279.

Yayanos, A. A., A. S. Dietz, and R. Van Boxtel. 1981. Obligately barophilic bacterium from the Mariana trench. Proceedings of the National Academy of Sciences **78**: 5212–5215.

Zeng, X., J.-L. Birrien, Y. Fouquet, G. Cherkashov, M. Jebbar, J. Quérellou, P. Oger, M. A. Cambon Bonavita, X. Xiao, and D. Prieur. 2009. Pyrococcus CH1, an obligate piezophilic hyperthermophile: extending the upper pressure-temperature limits for life. The ISME Journal **3**: 873–876.

Zhang, Y., X. Li, D. H. Bartlett, and X. Xiao. 2015. Current developments in marine microbiology: high-pressure biotechnology and the genetic engineering of piezophiles. Current Opinion in Biotechnology **33**: 157–164.

Zobell, C. E., and R. Y. Morita. 1957. Barophilic Bacteria in Some Deep Sea Sediments. Journal of Bacteriology **73**: 563.